Math 112: Introduction to Analysis
Lecture Notes

David Perkinson

Spring 2024

Contents

Week 1, Monday: Course Information 4

Week 1, Wednesday: Induction 10

Week 1, Friday: Sets 16

Week 2, Monday: More sets, Cartesian products 20

Week 2, Wednesday: Relations and equivalence relations 23

Week 2, Friday: Equivalence classes 26

Week 3, Monday: Functions 28

Week 3, Wednesday: More functions 32

Week 3, Friday: Modular arithmetic 37

Week 4, Friday: Field axioms 41

Week 4, Friday: Field axioms addendum 46

Week 5, Monday: Order axioms 48

Week 5, Wednesday: Completeness 52

Week 5, Friday: Extrema 55

Week 6, Monday: Complex numbers I 59

Week 6, Wednesday: Complex numbers II 64

Week 6, Friday: Complex numbers III 69

Week 7, Monday: Topology **74**

Week 7, Wednesday: Sequences I **79**

Week 7, Friday: Sequences II **84**

Week 8, Monday: Sequences II **88**

Week 9, Monday: New-from-old limit theorem **92**

Week 9, Wednesday: Monotone Convergence Theorem **97**

Week 9, Friday: Misc. limit theorems, infinite limits, and Cauchy sequences **101**

Week 10, Monday: Series **107**

Week 10, Wednesday: Series tests I **113**

Week 10, Friday: Series tests II **118**

Week 11, Monday: Series tests III **124**

Week 11, Wednesday: Limits of functions **129**

Week 11, Friday: Continuity and derivatives **135**

Week 12, Monday: Power series I **141**

Week 12, Wednesday: Power series II **146**

Week 12, Friday: Taylor series I **150**

Week 13, Monday: Taylor series II **155**

Week 13, Wednesday: Two theorems from calculus **163**

Week 13, Friday: Three theorems by Euler **169**

HANDOUTS **173**

Mathematical writing 174

Proof templates 178

Axioms for the real numbers 181

Week 1, Monday: Course Information

Place & Time:	Eliot 314, MWF 12:00–12:50 P.M.
Instructor:	David Perkinson (davidp@reed.edu)
Office Hours:	See our course Moodle page.
Lecture notes:	See our course homepage.
Supplemental reference:	Introduction to Analysis, by Irena Swanson
Course homepage:	https://people.reed.edu/~davidp/112/

Course description

This course covers the field axioms, the real and complex fields, real and complete sequences and series, an introduction to complex functions, continuity and differentiation; power series and the complex exponential. Prerequisite: Mathematics 111 (Calculus) or equivalent.

Learning outcomes

After taking this course, students will be able to:

- understand the field axioms and use them to derive properties of the real numbers;
- understand the basic properties of complex numbers;
- understand limits of sequences, series, and real and complex functions;
- better understand the underlying theory of calculus;
- read and write rigorous mathematical proofs in the context of analysis;
- work as part of a small group to solve mathematical problems; and
- communicate mathematical ideas verbally and in writing.

Distribution requirements

This course can be used towards your Group III, "Natural, Mathematical, and Psychological Science," requirement. It accomplishes the following goals for the group:

- Use and evaluate quantitative data or modeling, or use logical/mathematical reasoning to evaluate, test, or prove statements.
- Given a problem or question, formulate a hypothesis or conjecture, and design an experiment, collect data or use mathematical reasoning to test or validate it.

This course **does not** satisfy the "primary data collection and analysis" requirement.

Course design

Nearly all of our meetings will break down into four components:

- *Reading.* Every class will have an assigned reading which you must complete and engage with before we meet.

- *Lecture.* A short online lecture and quiz accompanies each reading assignment and must be completed before the start of the corresponding class. Questions that arise during the lecture/quiz will be addressed in class or during office hours.

- *Active class sessions.* Our 50-minute meetings will focus on group work with your peers. Collaborative problem-solving will allow you to interact with and grow your understanding of the material.

- *Homework.* Based on each day's work, I will assign two or three harder homework problems for you to complete after class. These will be due via Gradescope on the Friday of the week after the material is covered.

The purpose of this structure is to scaffold your learning so that you will first engage with easy quiz problems based on your reading and the recorded mini-lecture, then bolster skills through collaborative problem-solving, and finally gain mastery over content by engaging with homework problems.

Expectation: Before class starts, you'll complete the reading, lecture, and online quiz related to that day's content.

Texts

The course will use our *Math 112 Lecture Notes* as its primary text. This is a free PDF file available on the course homepage which will be updated as the semester proceeds. (If you find typos or have suggestions for improving the notes, please let me know!) We will use Irena Swanson's *Introduction to Analysis*, also freely available at our course homepage, as a supplement to support the reading in the lecture notes.

Reading assignments and mini-lectures

The required reading and recorded lectures are essential to the course and provide a leaping off point for each of our class meetings. The associated quizzes are intended to ensure that you are following the text and lecture at an appropriate level; they should not be particularly hard, though some of the problems will be nontrivial. The quizzes are embedded into the mini-lecture videos (posted to the course homepage) and are due before class. These quizzes will be assessed on the basis of completion, not on score.

Group work

Most of class time will be spent working in small groups with classmates. The participants in each group will vary from class-to-class. I will rotate among the groups to see how everyone is doing. It is also possible to request my assistance if your group needs immediate help. You should work through the problems for the day *in order*. At times, your group will not be able to make it through every problem—that's expected and is OK, in general. You can work on these problems on your own, if interested, and in any case, solutions will be provided. Class will end with everyone together for a brief discussion of the day's problems.

The ability to work collaboratively and to communicate mathematics verbally is a major goal of the course. Members of each group should work together to make sure everyone is supported, is comfortable, and participates—it's not just about finding a solution to each problem. So please look out for the other members of your group.

Homework

Homework is due via Gradescope[1] before class each Friday. Excellent solutions take many forms, but they all have the following characteristics:

- they use complete sentences, even when formulas or symbols are involved;
- they are written as explanations for other students in the course; in particular, they fully explain all of their reasoning and do not assume that the reader will fill in details;
- when graphical reasoning is called for, they include large, carefully drawn and labeled diagrams;
- they are neatly typeset using the LaTeX document preparation system. A guide to LaTeX resources is available on the course homepage.

I reserve the right to not accept late homework. If health or family matters might impede the timely completion of your homework, please contact me as early as possible.

Collaboration

You are permitted and encouraged to work with your peers on homework problems. You must cite those with whom you worked, and you must write up solutions independently. **Duplicated solutions will not be accepted and constitute a violation of the Honor Principle.**

[1] Gradescope is an online homework submission and evaluation platform. You will receive a link to register for our class's Gradescope page during the first week of classes.

Feedback

You will receive timely feedback on your homework via Gradescope, either from me or the course grader (another mathematics undergraduate). Most homework problems will be graded on a five-point scale (5 = perfect; 4 = minor mistake; 3 = major mistake, right idea; 2 = significant idea; 1 = attempted, 0 = none of the above.) The quality of your writing will be taken into account. If your answer is incorrect, this will be reflected in the score, and there will also be a comment indicating where things went wrong with your solution. You are strongly encouraged to engage with this comment, understand your error, and try to come up with a correct solution. You are welcome to talk about problems with me during office hours (see the Help section).

Exams

We will have two midterm exams and a final exam. Calculators, computers, phones, collaboration, books, and the Internet are prohibited during exams. Depending on how the semester goes, we may need to revise the following times:

- Exam 1: Monday, February 12; due via Gradescope noon, Wednesday, February 14.
- Exam 2: distributed via email Wednesday, March 20; due via Gradescope noon, Friday, March 22.
- Final Exam: in-person, as scheduled by the Registrar, May 6–9.

Quizzes

In addition to the online quizzes that accompany the video lectures, we will have short in-class weekly quizzes. Each quiz is designed to summarize important work from the previous week.

Joint expectations

As members of a communal learning environment, we should all expect consideration, fairness, patience, and curiosity from each other. Our aim is to all learn more through cooperation and genuine listening and sharing, not to compete or show off. I expect diligence and academic and intellectual honesty from each of you. I will do my best to focus the course on interesting, pertinent topics, and provide feedback and guidance which will help you excel as a student.

Help

Everyone is welcome and encouraged to attend my **office hours**. They are an opportunity to clarify difficult material and also delve deeper into topics that interest you. The math/stats department also hosts **drop-in tutoring** Sunday, Monday, Tuesday,

Wednesday, and Thursday 7–9 P.M. Tutors will be available to clarify concepts and help you with homework problems. Links for office hours and drop-in tutoring are available at our course Moodle page.

Finally, every Reed student is entitled to one hour of free individual tutoring per week. Use the tutoring app in IRIS to arrange to work with a student tutor.

The Internet

You are welcome to use Internet resources to supplement content we cover in this course, with the exception of solutions to homework problems. **Copying solutions from the Internet is an Honor Principle violation and will result in an academic misconduct report.**

Academic accommodations

If you have a documented disability requiring academic accommodation, I will be notified by Disability & Accessibility Resources (DAR). I am happy to communicate with you in person or via email about accommodations. If you believe you have an undocumented disability and that accommodations would ensure equal access to your Reed education, I would be happy to help you contact DAR.

Grades

Your grade will reflect a composite assessment of the work you produce for the class, weighted in the following fashion: 35% homework, 25% final exam, 20% exam 2, 10% exam 1, 5% completion of quizzes, 5% class participation.

Week 1, Wednesday: Induction

(Supplemental reading: Sections 1.4 and 1.5 in Swanson.)

Our first goal is to learn how to write a perfect proof by induction. (This material overlaps with that in Math 113, but it's important enough to go over twice.)

Example (template).

Proposition 1. For all integers $n \geq 1$,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

We'll discuss this theorem a bit before proving it by induction. For the case $n = 3$, it says:

$$1 + 2 + 3 = \frac{3 \cdot 4}{2},$$

and it's easy to see that both sides of this equation are equal to 6. When $n = 4$, we have

$$1 + 2 + 3 + 4 = 10 = \frac{4 \cdot 5}{2}.$$

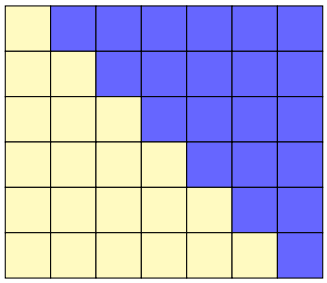
There is a tricky way to prove this theorem without using induction. Consider the following for the case $n = 6$:

$$\begin{array}{rl} & 1+2+3+4+5+6 \\ + & 6+5+4+3+2+1 \\ \hline & 7+7+7+7+7+7 \end{array} = 6 \cdot 7$$

Adding the sum twice gives $6 \cdot 7 = 42$. Divide by two to get the sum:

$$1 + 2 + 3 + 4 + 5 + 6 = \frac{6 \cdot 7}{2} = 21.$$

In a picture:



The 6×7 square contains our sum twice—once in yellow and once in blue. The proof clearly generalizes:

$$\begin{array}{rccccccccc} & 1 & + & 2 & + & \cdots & + & (n-1) & + & n \\ + & n & + & (n-1) & + & \cdots & + & 2 & + & 1 \\ \hline & (n+1) & + & (n+1) & + & \cdots & + & (n+1) & + & (n+1) \end{array} \quad = n \cdot (n+1).$$

Divide by two to get the general sum formula:

$$1 + 2 + \cdots + (n-1) + n = \frac{n(n+1)}{2}.$$

We now give a proof of the proposition *using induction*. Please use it as a template for your own induction proofs.

Proof. We will prove this by induction. For the base case, $n = 1$, the result holds since in that case

$$1 + 2 + \cdots + n = 1 = \frac{1 \cdot (1+1)}{2}.$$

Suppose the result holds for some $n \geq 1$. It follows that

$$\begin{aligned} 1 + 2 + \cdots + (n+1) &= (1 + 2 + \cdots + n) + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) && \text{(by the induction hypothesis)} \\ &= \frac{n(n+1) + 2(n+1)}{2} \\ &= \frac{(n+1)(n+2))}{2} \\ &= \frac{(n+1)((n+1)+1)}{2}. \end{aligned}$$

So the result holds for the case $n + 1$, too. The proposition follows by induction. □

Rules.

1. Always start a proof by induction by telling your reader that you are giving a proof by induction.

2. Next, show that result holds for the smallest value of n in question—in this case, $n = 1$.

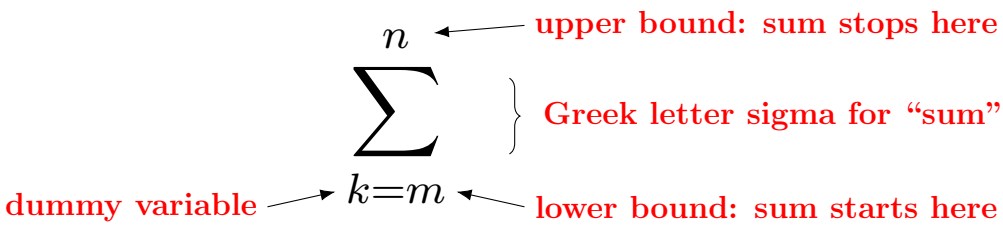
3. Note that we assume the result is true *for some* $n \geq 1$. If we said, instead: "assume the result holds for $n \geq 1$", this would mean we're assuming the result for all $n \geq 1$. But that would be circular: we'd be assuming what we are trying to prove. Instead, at the induction step, we are merely saying that *if* we did know the result for a particular value of n, we could prove that it follows for the next value of n.

4. End the proof with a $\square$. This tells the reader the proof is over.

Summation notation. For integers $m \leq n$, and a function f defined at the integers $m, m+1, \ldots, n$, we use the notation

$$\sum_{k=m}^{n} f(k) := f(m) + f(m+1) + \cdots + f(n).$$

(Note: the notation $A := B$ means A is *defined* to be B.) A closer look:



Example. Suppose that $f(k) = k^2$. Then

$$\begin{aligned}\sum_{k=-1}^{2} f(k) &= f(-1) + f(0) + f(1) + f(2) \\ &= (-1)^2 + 0^2 + 1^2 + 2^2 \\ &= 1 + 0 + 1 + 4 \\ &= 6.\end{aligned}$$

For this same sum we could write $\sum_{k=-1}^{2} k^2$ or $\sum_{t=-1}^{2} t^2$, for example.

Note: If $m > n$, then by convention, we take $\sum_{k=m}^{n} f(k) := 0$. This is called the *empty sum*.

More examples.

$$\sum_{i=2}^{4}(2i+i^2) = (2\cdot 2+2^2)+(2\cdot 3+3^2)+(2\cdot 4+4^2) = 47$$

$$\sum_{k=1}^{5} 2 = 2+2+2+2+2 = 10.$$

In general,

$$\sum_{k=m}^{n}(af(k)+bg(k)) = a\sum_{k=m}^{n} f(k) + b\sum_{k=m}^{n} g(k).$$

Product notation. There is a similar notation for products:

$$\prod_{k=m}^{n} f(k) := f(m)\cdot f(m+1)\cdots f(n).$$

For example,

$$\prod_{k=1}^{n} k = 1\cdot 2\cdots n =: n!.$$

If $m > n$, we define the *empty product* by $\prod_{k=m}^{n} f(k) := 1$. (So, for example, $0! = 1$.)

Back to induction. We now give our induction proof of Proposition 1 using summation notation:

Proposition 1 (using summation notation). For $n \geq 1$

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Proof. We will prove this by induction. The base case holds since

$$\sum_{i=1}^{1} i = 1 = \frac{1\cdot(1+1)}{2}.$$

Suppose the result holds for some $n \geq 1$. Then

$$\begin{aligned}\sum_{i=1}^{n+1} i &= \left(\sum_{i=1}^{n} i\right) + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) \qquad \text{by the induction hypothesis}\end{aligned}$$

$$= \frac{n(n+1)+2(n+1)}{2}$$
$$= \frac{(n+1)(n+2)}{2}$$
$$= \frac{(n+1)((n+1)+1)}{2}.$$

So the result then holds for $n+1$, too. The result holds for all $n \geq 1$, by induction. □

Another example.

Proposition 2. Show

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

for $n \geq 1$.

Proof. We will prove this by induction. The base case holds since

$$\sum_{k=1}^{1} k^2 = 1^2 = 1 = \frac{1 \cdot (1+1)(2 \cdot 1+1)}{6}.$$

Suppose the result holds from some $n \geq 1$:

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.1}$$

Then

$$\sum_{k=1}^{n+1} k^2 = 1^2 + \cdots + n^2 + (n+1)^2$$
$$= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \qquad \text{(by equation (2.1))}$$
$$= \frac{n(n+1)(2n+1)+6(n+1)^2}{6}$$
$$= \frac{(n+1)(n(2n+1)+6(n+1))}{6}$$
$$= \frac{(n+1)(2n^2+7n+6)}{6}$$
$$= \frac{(n+1)(n+2)(2n+3)}{6}$$
$$= \frac{(n+1)((n+1)+1)(2(n+1)+1)}{6}.$$

Thus, the result then holds for $n+1$, too. Our result follows for all $n \geq$ by induction. $\square$

Week 1, Friday: Sets

(Supplemental reading: Section 2.1 in Swanson.)

A *set* is a collection of objects. The objects in a set are the set's *elements* or *members*. If A is a set, we write $m \in A$ if m is an element of A and $m \notin A$ if m is not in A.

Examples.

1. $A = [0, 1)$ is the set of real numbers x such that $0 \leq x < 1$.
2. $\{1, 2, 3\}$ has three elements: 1, 2, and 3.
3. Note that
$$\{0, 1, 0, 0, 2, 2, 1, 1\} = \{0, 1, 2\}.$$
Sets don't contain the same element twice.
4. The elements of sets do not need to be numbers, e.g., $\{☺, ☹, 17\}$.
5. $A = \{\{1, 2, 3\}\}$ has only one element—the set $\{1, 2, 3\}$. Thus, for this set,
$$\{1, 2, 3\} \in A \quad \text{and} \quad 1 \notin A.$$

The empty set. The *empty set* is the set with no elements. It is denoted by $\emptyset$, and we write $\emptyset = \{\}$. For any object x, we have $x \notin \emptyset$. For example, $\emptyset \notin \emptyset$. On the other hand, note that the set $\{\emptyset\}$ is not empty—it has one element, namely, $\emptyset \in \{\emptyset\}$.

Propositional definition of a set. Let A be a set, and let $P(x)$ be a proposition—a statement that is either true or false—that depends on elements $x \in A$. We use the following notation sometimes to define sets

$$\{x \in A : P(x)\}.$$

We read this: "the set of elements x in the set A such that $P(x)$ is true. The colon, :, translates as "such that". Some examples:

$$\{x \in \mathbb{R} : 0 \leq x < 1\} = [0, 1)$$
$$\{x \in \mathbb{R} : x^2 = -1\} = \emptyset$$
$$\{x \in \mathbb{R} : x^2 = 1\} = \{-1, 1\}.$$

Some special sets:

$$\begin{aligned}
\text{integers} \quad & \mathbb{Z} := \{\dots, -2, -1, 0, 1, 2, \dots\} \\
\text{natural numbers} \quad & \mathbb{N} := \mathbb{Z}_{\geq 0} := \{0, 1, 2, \dots\} \\
\text{positive natural numbers} \quad & \mathbb{N}^+ := \mathbb{N}_{>0} := \mathbb{Z}_{>0} := \{1, 2, \dots\} \\
\text{rational numbers} \quad & \mathbb{Q} := \left\{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \neq 0\right\} \\
\text{real numbers} \quad & \mathbb{R} \quad \text{(more about these soon)} \\
\text{complex numbers} \quad & \mathbb{C} \quad \text{(more about these soon).}
\end{aligned}$$

Subsets and operations.

Definition. Let A and B be sets. Then A is a *subset* of B, denoted $A \subseteq B$, if every element of A is an element of B. If $A \subseteq B$, we may use the notation $B \supseteq A$ and say that B is a *superset* of A: so A is a subset of B if and only if B is a superset of A.

Time out for a remark about notation: In mathematics, the symbol "$\Rightarrow$" means "implies". Please use this rule: only use this symbol in your writing if it can be replaced by the word "implies". It is common for people who are just beginning to write mathematics to use the symbol to mean "equals" or "the next step in the argument is". That's not acceptable from now on.

Using this notation, we can write that $A \subseteq B$ if

$$x \in A \quad \Rightarrow \quad x \in B.$$

For some examples of subsets, we have

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

Definition. Let A and B be sets. Then $A = B$ if (i) $A \subseteq B$, and (ii) $B \subseteq A$. On the other hand, A is a *proper subset* of B, denoted $A \subsetneq B$, if $A \subseteq B$ and $A \neq B$. In that case, we write $A \subsetneq B$. (You might write $A \subset B$ for $A \subsetneq B$, but some mathematicians make no distinction between the notation $\subset$ and $\subseteq$.)

Examples. We have $A \subseteq A$, and $(0, 1) := \{x \in \mathbb{R} : 0 < x < 1\} \subsetneq \mathbb{R}$. We also have $\emptyset \subseteq A$ since otherwise there would need to be an element of $\emptyset$ that in not in A. That's impossible since $\emptyset$ has no elements.

Warning: Don't confuse $\subseteq$ with $\in$. For example, $1 \in \{1, 2\}$ but it is **not** true that $1 \subseteq \{1, 2\}$. Instead, $\{1\} \subseteq \{1, 2\}$.

Definition. The *intersection* of sets A and B is

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

the set of elements that are in both A **and** B. Their *union* is

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

the set of elements that are in A or B (or both[1]). The *complement*[2] of B in A is

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

the set of elements in A but not in B. The sets A and B are *disjoint* if

$$A \cap B = \emptyset.$$

Proofs for statements involving sets.

Many theorems in mathematics are, abstractly, statements that one set is contained in another. There is a standard style for these proofs that goes like this:

Template.

Proposition. If [some hypotheses go here], then

$$A \subseteq B.$$

Proof. Let $a \in A$. Then [use hypotheses, definitions, calculations here]. Therefore, $a \in B$. □

Here is an example where we put the template to use. I've put comments in red in the proof which are just meant to explain what is going on. These comments would be omitted in an actual proof.

Proposition. Let A, B, C be sets with $A \subseteq B$. Then

$$A \cap C \subseteq B \cap C.$$

Proof. Let $x \in A \cap C$. Then $x \in A$ and $x \in C$. [Here, I've just gone back to the definition of $\cap$.] Since $A \subseteq B$ and $x \in A$, it follows that $x \in B$. [That's from the definition of $\subseteq$.] So now we know that $x \in B$ and $x \in C$. Therefore, $x \in B \cap C$. [Definition of $\cap$, again.] □

Leaving out my comments, the actual proof would be:

[1] "or" is always inclusive in mathematical writing

[2] Note the difference in spelling between "complement" and "compliment"—these words mean different things.

Proof. Let $x \in A \cap C$. Then $x \in A$ and $x \in C$. Since $A \subseteq B$, and $x \in A$ is follows that $x \in B$. So now we know that $x \in B$ and $x \in C$. Therefore, $x \in B \cap C$. □

Here is another simple example of a proof of this form:

Proposition. Let A, B be sets. Then

$$A \cap B \subseteq A \cup B.$$

Proof. Let $x \in A \cap B$. Then $x \in A$ and $x \in B$. Since x is in A, it follows that x is in A or B. Hence, $x \in A \cup B$. □

There is so little to prove in the last proposition, that you might be confused about what you really need to say. The point I am trying to get across is, roughly, start with an element of the set on the left-hand side, spell out the definitions, and show that the element is necessarily in the set on the right-hand side.

Week 2, Monday: More sets, Cartesian products

(Supplemental reading: Sections 2.1 and 2.2 in Swanson.)

Template.

Proposition. If [some hypotheses go here], then

$$A = B.$$

Proof. Let $a \in A$. Then [use hypotheses, definitions, calculations here]. Therefore, $a \in B$. Hence, $A \subseteq B$.

Conversely, let $b \in B$. Then [use hypothesese, definitions, calculations here]. Therefore, $b \in A$. Hence, $B \subseteq A$, too. Therefore, $A = B$. □

An example:

Proposition. Let A and B be sets, and let $C := A \cup B$. Suppose $A \cap B = \emptyset$. Then

$$A = C \setminus B.$$

Proof. Let $a \in A$. Then $a \in C = A \cup B$. Since $A \cap B = \emptyset$ and $a \in A$, it follows that $a \notin B$. In sum, $a \in C$ and $a \notin B$. Therefore, $a \in C \setminus B$. Thus $A \subseteq C \setminus B$.

Conversely, let $x \in C \setminus B$. This means that $x \in C$ and $x \notin B$. But $x \in C = A \cup B$, means that $x \in A$ or $x \in B$. Since $x \notin B$, it follows that $x \in A$. Therefore, $C \setminus B \subseteq A$. □

Indexed unions and intersections. Let I be a set, and suppose that for each $i \in I$, you are given a set A_i. Then by definition,

$$\cup_{i \in I} A_i := \{x : x \in A_i \text{ for some } i \in I\}$$
$$\cap_{i \in I} A_i := \{x : x \in A_i \text{ for all } i \in I\}.$$

If $I = \mathbb{N}$, we might write $\cup_{i=1}^{\infty} A_i$ in place of $\cup_{i \in \mathbb{N}^+} A_i$, and similarly for intersections. In

that case, your can think of these operations as follows:

$$\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$$

$$\cap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots$$

Examples.

1. For each $n \in \mathbb{N}^+$, let $A_n = [0, 1/n)$, an interval in $\mathbb{R}$. Then
   (a) $\cup_{n\in\mathbb{N}^+} A_n = [0, 1)$.
   (b) $\cap_{n\in\mathbb{N}^+} A_n = \{0\}$.
2. $\cup_{r\in\mathbb{R}} \{r\} = \mathbb{R}$.

We will prove $\cap_{n\in\mathbb{N}^+} A_n = \{0\}$. We need to show these two sets are equal, so we show inclusions in both directions. I find that when possible, it helps to first get an intuitive grasp by writing out the indexed intersection long-hand:

$$\cap_{n\in\mathbb{N}^+} A_n = \cap_{n\in\mathbb{N}^+} [0, 1/n) = [0, 1) \cap [0, 1/2) \cap [0, 1/3) \cap \cdots$$

Each successive interval is contained in the preceding one. So the intersection is getting smaller as we go out in the chain. Now for a formal proof:

Proof. Let $x \in \cap_{n\in\mathbb{N}^+} A_n$. Then $x \in A_n = [0, 1/n)$ for all $n \in \mathbb{N}$. Thus,

$$0 \leq x < \frac{1}{n}.$$

This means that $x = 0$ (which we won't prove here). Therefore, $x \in \{0\}$. We have shown the inclusion

$$\cap_{n\in\mathbb{N}^+} A_n \subseteq \{0\}.$$

For the opposite inclusion: there is only one element of $\{0\}$, namely 0, and $0 \in [0, 1/n)$ for $n = 1, 2, \ldots$ Therefore, $0 \in \cap_{n\in\mathbb{N}^+} A_n$, and hence

$$\{0\} \subseteq \cap_{n\in\mathbb{N}^+} A_n.$$

Having shown both inclusions, we know the sets are equal. □

CARTESIAN PRODUCTS

Definition. The *Cartesian product* of sets A and B is

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

By (a, b) we mean an "ordered pair". (Formally, we could define $(a, b) := \{a, \{a, b\}\}$.) For example, $(1, 2) \neq (2, 1)$, whereas $\{1, 2\} = \{2, 1\}$. By definition,

$$(a, b) = (a', b') \quad \text{exactly when} \quad a = a \text{ and } b = b'.$$

Examples.

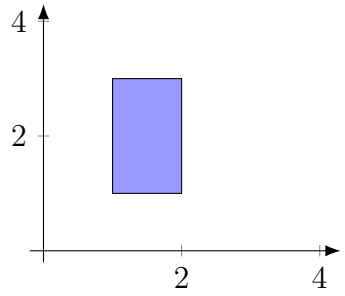
1. Let $A = \{\checkmark, \star\}$ and $B = \{1, 2, 3\}$. Then

$$A \times B = \{(\checkmark, 1), (\checkmark, 2), (\checkmark, 3), (\star, 1), (\star, 2), (\star, 3)\}.$$

2. Let $A = [1, 2] \subset \mathbb{R}$ and $B = [1, 3] \subset \mathbb{R}$. Then

$$A \times B = \{(a, b) : 1 \leq a \leq 2 \text{ and } 1 \leq b \leq 3\}.$$

This is a rectangle in the plane $\mathbb{R}^2$:



$[1, 2] \times [1, 3]$.

3. Let $A = B = \mathbb{R}$. Then $A \times B = \mathbb{R}^2$, the ordinary real plane.

Given sets A, B, C, we can define

$$A \times B \times C := \{(a, b, c) : a \in A, b \in B, \text{ and } c \in C\},$$

the collection of ordered triples. Similarly, one could define ordered quadruples, etc. The *n-fold* Cartesian product of a set A with itself is

$$A^n := \underbrace{A \times \cdots \times A}_{n \text{ times}}.$$

For example, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ is ordinary 3-space.

Week 2, Wednesday: Relations and equivalence relations

(Supplemental reading: Section 2.3 in Swanson.)

Relations.

Definition. A *relation* between sets A and B is a subset R of their Cartesian product:

$$R \subseteq A \times B.$$

If $(a, b) \in R$, we may write aRb. If $A = B$, we say R is a *relation on* A.

Examples.

(a) First, we consider a toy example that does not have much meaning. Let $A = \{\checkmark, \star\}$ and $B = \{1, 2, 3\}$, and $R = \{(\checkmark, 2), (\checkmark, 3), (\star, 2)\}$. Then $\checkmark R\, 2$, $\checkmark R\, 3$, and $\star R\, 2$.

(b) For a more serious example we see how the "less than or equal" relation on the integers, $\mathbb{Z}$, can be thought of as a relation in the technical sense we have just introduced. We just take

$$R = \{(a, b) : a, b \in \mathbb{Z} \text{ and } a \leq b\}.$$

So in this case, we have aRb if and only if $a \leq b$.

A relation R on a set S, i.e., $R \subseteq S \times S$, is an *equivalence relation on* S if for all $x, y, z \in S$:

1. xRx (the relation is reflexive)
2. If xRy, then yRx (the relation is symmetric)
3. If xRy and yRz, then xRz (the relation is transitive).

If R is an equivalence relation on S, we usually write $a \sim b$ instead of aRb. So re-writing the axioms for an equivalence relation, we require for all $x, y, z \in S$:

1. $x \sim x$ (the relation is reflexive)
2. If $x \sim y$, then $y \sim x$ (the relation is symmetric)
3. If $x \sim y$ and $y \sim z$, then $x \sim z$ (the relation is transitive).

Examples.

1. Is $\leq$ an equivalence relation on $\mathbb{Z}$? It's reflexive: $a \leq a$ for all $a \in \mathbb{Z}$. It's transitive: if $a \leq b$ and $b \leq c$, then $a \leq c$. However, it's not symmetric: $a \leq b$ does not necessarily imply $b \leq a$. For example $0 < 1$ but $1 \not< 0$. So $\leq$ is a relation on $\mathbb{Z}$, but it is not an equivalence relation.

2. The relation $=$ is an equivalence relation on any set.

The integers modulo n. We now describe a whole class of equivalence relations on the set of integers, $\mathbb{Z}$—one for each $n \in \mathbb{Z}$. First, fix your favorite integer $n \in \mathbb{Z}$. Then for $a, b \in \mathbb{Z}$ we will say $a \sim b$ if a and b differ by a multiple of n: i.e., if

$$a - b = kn$$

for some $k \in \mathbb{Z}$.

We will prove we get an equivalence relation in the next lecture. For now let's look at some examples. The case of $n = 2$ says $a, b \in \mathbb{Z}$ are equivalent if a and b differ by a multiple of 2. So we have the following equivalences

$$\cdots \sim -4 \sim -2 \sim 0 \sim 2 \sim 4 \sim \cdots .$$

and we have

$$\cdots \sim -3 \sim -1 \sim 1 \sim 3 \sim 5 \sim \cdots .$$

So under this equivalence relation, the even integers are all equivalent to each other, and the odd integers are equivalent to each other. No even integer is equivalent to an odd integer. We say there are two "equivalence classes".

For another example, consider the relation in the case $n = 3$: two integers are equivalent if they differ by a multiple of 3. In that case we have three equivalence classes:

$$\cdots \sim -6 \sim -3 \sim 0 \sim 3 \sim 6 \sim \cdots \tag{5.1}$$

$$\cdots \sim -5 \sim -2 \sim 1 \sim 4 \sim 7 \sim \cdots \tag{5.2}$$

$$\cdots \sim -4 \sim -1 \sim 2 \sim 5 \sim 8 \sim \cdots . \tag{5.3}$$

Equivalence relations and partitions. A *partition* of a set S is a collection of nonempty subsets S_k satisfying:

(a) the S_k are pair-wise disjoint: for all i, j, we have $S_i \cap S_j = \emptyset$, and

(b) the union of the S_k is S: we have $\cup_k S_k = S$.

Example. Let $S = \{1, 2, 3, 4, 5, 6\}$. The following sets form a partition of S:

$$S_1 := \{2, 4\}, S_2 := \{1, 3, 5\}, S_3 := \{6\}.$$

No two of these sets share an element, and their union is all of S.

Fact 1: For every partition of a set S, there is an equivalence relation on S whose equivalence classes are the subsets in the partition. The equivalence relation is defined by requiring the elements in each subset of the partition to be related to each other.

Continuing with the previous example, we are looking for a equivalence relation on $S = \{1, 2, 3, 4, 5, 6\}$ whose equivalence classes are S_1, S_2 and S_3. First considering $S_1 = \{2, 4\}$, we require $2 \sim 2$, $4 \sim 4$, $2 \sim 4$, and $4 \sim 2$. The other subsets, S_2 and S_3 are handled similarly, producing the required equivalence relation on S.

We have a converse to the above fact:

Fact 2: Given an equivalence relation on a set S, its set of equivalence classes partitions S.

As an example, consider the integers modulo 3. In (5.1), above, we saw that we get three equivalence classes: one containing 0, one containing 1, and one containing 2. One may check that these equivalence classes partition the set $\mathbb{Z}$: every integer is in exactly one of these classes.

Facts 1 and 2 show that equivalence relations and partitions are essentially the same thing. The interested reader could attempt to give a formal proof.

Week 2, Friday: Equivalence classes

(Supplemental reading: Section 2.3 in Swanson.)

Definition. Let $\sim$ be an equivalence relation on a set S. The *equivalence class* for $x \in S$ is

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

The *quotient* of S by $\sim$ is the set of equivalence classes for $\sim$:

$$S/\sim := \{[x] : x \in S\}.$$

We also refer to $S/\sim$ as "S modulo $\sim$" or "S mod $\sim$".

Last time, we introduced an equivalence relation on $\mathbb{Z}$ for each $n \in \mathbb{Z}$. Fix your favorite integer n. Then for $a, b \in \mathbb{Z}$ we will say $a \sim b$ if a and b differ by a multiple of n, i.e., if

$$a - b = kn$$

for some $k \in \mathbb{Z}$. We use the notation $\mathbb{Z}/n\mathbb{Z}$ to denote $\mathbb{Z}/\sim$, the set of equivalence classes of $\mathbb{Z}$ modulo n.

Example. Consider the equivalence relation $\mathbb{Z}$ we defined above for the case $n = 2$. There are two equivalence classes:

$$[0] = \{0, \pm 2, \pm 4, \dots\}$$
$$[1] = \{1, \pm 3, \pm 5, \dots\}.$$

The "name" of an equivalence class is not necessarily unique. In this example, we have $[0] = [2]$ and $[1] = [-17]$, for instance. Note that people use these equivalence classes all the time: it's just the notion of even and odd.

Example. What are the equivalence classes when $n = 3$? Looking above, we see that there are three of them:

$$[0] = \{\dots, -6, -3, 0, 3, 6, \dots\}$$
$$[1] = \{\dots, -5, -2, 1, 4, 7, \dots\}$$
$$[2] = \{\dots, -4, -1, 2, 5, 8, \dots\}.$$

It's interesting that, unlike the case $n = 2$, there aren't common words for the three equivalence classes of the integers modulo three.

Proposition. Let $n \in \mathbb{Z}$ and for $a, b \in \mathbb{Z}$, say $a \sim b$ if

$$a - b = kn$$

for some integer k. Then $\sim$ is an equivalence relation.

Proof. We need to show that $\sim$ is reflexive, symmetric, and transitive. Let $a, b, c \in \mathbb{Z}$.

Reflexivity. We have $a \sim a$ since $a - a = 0 \cdot n$. (We are letting $k = 0$ in the definition of $\sim$.)

Symmetry. Suppose that $a \sim b$. This means that there exists a $k \in \mathbb{Z}$ such that

$$a - b = kn.$$

But then

$$b - a = (-k)n.$$

Hence $b \sim a$.

Transitivity. Suppose that $a \sim b$ and $b \sim c$. Then there exist $k, \ell \in \mathbb{Z}$ such that

$$a - b = kn \quad \text{and} \quad b - c = \ell n.$$

Adding these two equations, we get

$$a - c = (a - b) + (b - c) = kn + \ell n = (k + \ell)n.$$

Therefore, $a \sim c$. To help make this last point as clear as possible, we can let $m := k + \ell$. Then $m \in \mathbb{Z}$, and

$$a - c = mn.$$

□

Template. Here is a template for a proof that a given relation is an equivalence relation:

Proposition. Define a relation $\sim$ on a set A by blah, blah, blah. Then $\sim$ is an equivalence relation.

Proof. Let $a, b, c \in A$.

Reflexivity. We have $a \sim a$ since blah, blah, blah. Therefore, $\sim$ is reflexive.

Symmetry. Suppose that $a \sim b$. Then, blah, blah, blah. It follows that $b \sim a$. Therefore $\sim$ is symmetric.

Transitivity. Suppose that $a \sim b$ and $b \sim c$. Then blah, blah, blah. It follows that $a \sim c$. Therefore, $\sim$ is transitive.

Since $\sim$ is reflexive, symmetric, and transitive, it follows that $\sim$ is an equivalence relation. □

Week 3, Monday: Functions

(Supplemental reading: Section 2.4 in Swanson.)

Rule for proof-writing. When proving something is *not* the case, always give as simple a counter-example as possible. This is the opposite of trying to prove something *is* the case, in which you must be as general as possible (i.e., don't give a "proof by example"). For instance, if I want to disprove the statement "There are no prime numbers that are 1 modulo 4", I can say: "The statement is not true. For example, 5 is prime and $5 = 1 \bmod 4$."

FUNCTIONS

What is a function? To work up to the definition, think about some function that you know. For instance, consider the function $f(x) = x^2$ from calculus. The *graph* of f is the set

$$\left\{(x, x^2) \in \mathbb{R}^2 : x \in \mathbb{R}\right\}.$$

Note that f is completely determined by its graph. To see this point clearly, note that if I told you that I am thinking of a function whose graph is

$$\left\{(x, 2\cos(x) + 3) \in \mathbb{R}^2 : x \in \mathbb{R}\right\},$$

you would be able to tell me that the function is $g(x) = 2\cos(x) + 3$.

Next, note that the graph of a function is actually a relation between sets. For instance, in the above two examples, it's a relation between $\mathbb{R}$ and itself, i.e., a relation on $\mathbb{R}$. Is every relation on $\mathbb{R}$ a function? The answer is "no" since a function must be single-valued. For instance, if $h(1) = 4$, we cannot also have $h(1) = 7$. So we can't have both $(1, 4)$ and $(1, 7)$ in the graph of h.

Definition. Let A and B be sets. A *function* f with *domain* A and *codomain* B, denoted $f\colon A \to B$ is a relation

$$R_f \subseteq A \times B$$

such that:

1. For all $a \in A$, there exists $b \in B$ such that $(a, b) \in R_f$.

2. If (a, b) and (a, b') are in R_f, then $b = b'$.

If $(a, b) \in R_f$, then we write $f(a) = b$.

Remarks. Part 1 says that a function needs to be defined at each point in its domain. Part 2 says that a function is single-valued. Note that R_f, which we are using to define f, is exactly the graph of f.

Example. Let $S = \{1, 2, 3\}$ and $T = \{0, 1\}$. Define a function $f\colon S \to T$ by letting $f(1) = 0$ and $f(2) = f(3) = 1$.

1. What is the domain of f?
2. What is the codomain of f?
3. What is the relation R_f defining f?

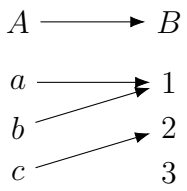
(See the footnote[1] for the solution.)

Definition. Let $f\colon A \to B$ be a function. The *image* or *range* of a function is the subset of B

$$\operatorname{im}(f) := \{f(a) \in B : a \in A\}.$$

Examples.

1. Let $A = \{a, b, c\}$ and let $B = \{1, 2, 3\}$, and define a function $f\colon A \to B$ by $f(a) = f(b) = 1$ and $f(c) = 2$:



   We have

$$\begin{aligned} \text{domain of } f\colon &\quad A = \{a, b, c\} \\ \text{codomain of } f\colon &\quad B = \{1, 2, 3\} \\ \text{image of } f\colon &\quad \{1, 2\}. \end{aligned}$$

   Note: The words "image" and "range" mean the same thing. As is evident in the above example, the image is a subset of the codomain, but not necessarily equal to the codomain.

[1] The domain is $S = \{1, 2, 3\}$, the codomain is $T = \{0, 1\}$, and the relation is $R_f = \{(1, 0), (2, 1), (3, 1)\}$

2. The image of the function

$$\begin{aligned} g\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto |x-3| \end{aligned}$$

is $\mathbb{R}_{\geq 0}$, the set of nonnegative real numbers. The domain and codomain of g are both $\mathbb{R}$.

Definition. Let $f\colon A \to B$ be a function. Then

1. f is *injective* or *one-to-one* if for all $a, a' \in A$:

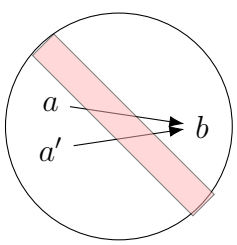
$$f(a) = f(a') \quad \Longrightarrow \quad a = a'.$$

2. f is *surjective* or *onto* if

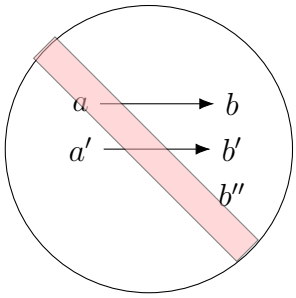
$$\operatorname{im}(f) = B.$$

3. f is *bijective* if it is injective and surjective (one-to-one and onto).

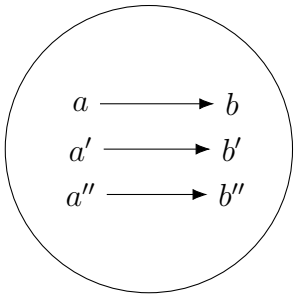
"International symbols":



not injective



injective
not surjective



bijective

Proof templates. Here are templates for proving a function is injective, surjective, or bijective:

Proposition. The function $f\colon A \to B$ is injective.

Proof. Let $x, y \in A$, and suppose that $f(x) = f(y)$. Then blah, blah, blah. It follows that $x = y$. Hence, f is injective. □

Proposition. The function $f\colon A \to B$ is surjective.

Proof. Let $b \in B$. Then blah, blah, blah. Thus, there exists $a \in A$ such that $f(a) = b$. Hence, f is surjective. □

Proposition. The function $f\colon A \to B$ is bijective.

Proof. We first show that f is injective. [Follow the template above to show injectivity.] Next, we show f is surjective. [Follow the template above to show surjectivity.] □

Later, we will see alternative proofs of these propositions using (left and right) inverse functions.

Examples.

1. Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 2x$. Then f is bijective.

   Proof. To see f is injective, let $x, y \in \mathbb{R}$ and suppose that $f(x) = f(y)$. Since $f(x) = f(y)$, we have that $2x = 2y$. Dividing by 2, we see that $x = y$. Hence, f is injective.

   To see that f is surjective, let $z \in \mathbb{R}$ (in the codomain). Then, in the domain, we have $z/2 \in \mathbb{R}$, and $f(z/2) = 2(z/2) = z$. Hence, f is surjective.

   Since f is injective and surjective, it is bijective. □

2. The function

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

   is neither injective nor surjective. It's not injective since, for instance, $f(1) = f(-1)$, and it's not surjective since its image is $\mathbb{R}_{\geq 0}$, which is not equal to the codomain, $\mathbb{R}$. For instance, -1 is not in the image of f.

Challenge. Can you find $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$ such that the function $A \to B$ defined by $x \mapsto x^2$ is injective but not surjective? surjective but not injective? bijective?

Week 3, Wednesday: More functions

More functions

(Supplemental reading: Section 2.4 in Swanson.)

Functions acting on sets.

Definition. Let $f\colon A \to B$ be a function between sets A and B, and let $C \subseteq A$. The *image of C under f* is the subset of B

$$f(C) := \{f(c) : c \in C\} \subseteq B.$$

Let $D \subseteq B$. The *inverse image of D under f* is the subset of A

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

Roughly, the image of C under f is the subset of the codomain, B, obtained by plugging all of the elements of C into f. The inverse image of D is the subset of the domain, A, consisting of all elements that are sent into D by f. Note that the image (or range) of f, defined previously, is the image of the domain, i.e., $\mathrm{im}(f) = f(A)$.

Examples.

1. Let

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2. \end{aligned}$$

   Then

$$f([0,2]) = [0,4],$$

   and

$$f^{-1}([-1,3]) = [-\sqrt{3}, \sqrt{3}].$$

   (Reminder: in this context, $[-1,3] := \{x \in \mathbb{R} : -1 \leq x \leq 3\}$, for example.) For the latter, note that $f^{-1}([-1,3]) = f^{-1}([0,3])$ since the square of a real number is nonnegative. Then notice that $[-\sqrt{3}, \sqrt{3}]$ is exactly the set of real numbers whose squares are in $[0,3]$.

2. Let

$$f\colon \mathbb{Z} \to \mathbb{Z}/4\mathbb{Z}$$
$$a \mapsto [a].$$

Here, $[a]$ denotes the equivalence class for 4, i.e.,

$$[a] := \{b \in \mathbb{Z} : a - b = 4k \text{ for some } k \in \mathbb{Z}\}.$$

Then

$$f(\{0, 2, 4, 6, 8, \dots\}) = \{[0], [2]\}.$$

and

$$f^{-1}([3]) = \{\dots, -9, -5, -1, 3, 7, 11, 15, \dots\}.$$

Proposition. If $f\colon A \to B$ and $C \subseteq A$, then

$$C \subseteq f^{-1}(f(C)).$$

Proof. Let $c \in C$. Then $f(c) \in f(C)$. Hence $c \in f^{-1}(f(C))$. [Note to the reader: Go back to the definition of *image* and *inverse image* to see that what is written here constitutes a proof.] □

Remark. The proposition we just proved is no longer true if "$\subseteq$" is replaced by "$=$". For example, consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Let $C = [0, \infty)$. Then

$$[0, \infty] = C \subsetneq f^{-1}(f(C)) = f^{-1}(f([0, \infty))) = (-\infty, \infty).$$

Proposition. Let $f\colon A \to B$, and let $X, Y \subseteq A$. Then

$$f(X \cap Y) \subseteq f(X) \cap f(Y).$$

Proof. Let $z \in f(X \cap Y)$. Then $z = f(a)$ for some $a \in X \cap Y$. Since $a \in X \cap Y$, it follows that $a \in X$, and hence, $z = f(a) \in f(X)$. Similarly, since $a \in X \cap Y$, it follows that $a \in Y$, and hence $z = f(a) \in f(Y)$. We've shown $z \in f(X)$ and $z \in f(Y)$, and therefore, $z \in f(X) \cap f(Y)$. □

Challenge. Find an example of a function $f\colon A \to B$ and subsets X, Y of A such that

$$f(X \cap Y) \neq f(X) \cap f(Y).$$

Thus, we don't necessarily have equality in the previous proposition.

Composition of functions.

Definition. Let $f\colon A \to B$ and $g\colon B \to C$ be functions. The *composition of f and g* is the function

$$g \circ f\colon A \to C$$

defined by $(g \circ f)(a) := g(f(a))$. (See Figure 8.1.)

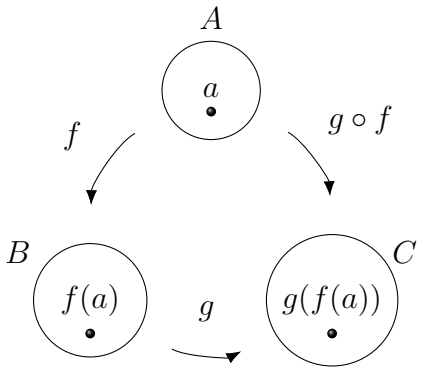
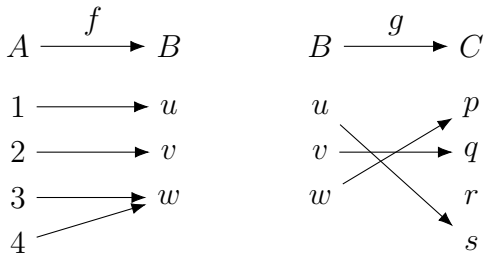
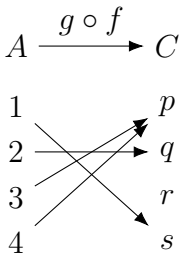


Figure 8.1: The composition of two functions.

Example. Let $A = \{1, 2, 3, 4\}$, $B = \{u, v, w\}$, and $C = \{p, q, r, s\}$, and consider the two functions pictured below:



The composition $g \circ f$ is then:



Note, the composition $g \circ f$ makes sense since codomain of f is B, which is the same as the domain of g. The opposite composition, $f \circ g$ does not makes sense since the codomain of g is C, which is not the domain, A, of f.

Proposition. The composition of surjective functions is surjective.

Proof. Let $f\colon A \to B$ and $g\colon B \to C$ be surjective functions. We want to show that $g \circ f\colon A \to C$ is surjective. Let $c \in C$. Since g is surjective, there exists $b \in B$ such that $g(b) = c$. Since f is surjective, there exists $a \in A$ such that $f(a) = b$. Then,

$$(g \circ f)(a) := g(f(a)) = g(b) = c.$$

We have shown that every element of C is in the image of $g \circ f$, i.e., $g \circ f$ is surjective. Therefore, $g \circ f$ is surjective. □

Proposition. The composition of injective functions is injective.

Proof. Let $f\colon A \to B$ and $g\colon B \to C$ be injective functions. We want to show that $g \circ f$ is injective. So let $a, a' \in A$ and suppose that $(g \circ f)(a) = (g \circ f)(a')$. In other words, $g(f(a)) = g(f(a'))$. Since g is injective, $f(a) = f(a')$. Then, since f is injective $a = a'$. Thus, $g \circ f$ is injective. □

Corollary. The composition of bijective functions is bijective.

Proof. This is an immediate corollary of the preceding two propositions. □

Inverse functions.

Definition. Suppose $f\colon A \to B$ is a bijective function. Then the *inverse of* f is the function $g\colon B \to A$ defined by

$$g(b) = a \quad \text{if} \quad f(a) = b.$$

This function g is denoted by f^{-1}. Hence, $f^{-1}(b) = a$ if and only if $f(a) = b$.

Note that the inverse function is defined only if f is bijective.

Important remark. Let $f\colon A \to B$. Earlier, for every subset $C \subset B$, we defined $f^{-1}(C) = \{a \in A : f(a) \in C\}$. If f is bijective, then we have two different but closely related meanings for f^{-1}. If $b \in B$, then $f^{-1}(\{b\})$ is a subset of A consisting of a single element, say a. So $f^{-1}(\{b\}) = \{a\}$. That's the earlier meaning of f^{-1}, defined for subsets of B. For the inverse of f, just defined, we have $f^{-1}(b) = a$.

Exercise. For each of the following functions, state why there is no inverse, or describe the inverse function.

1.

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto |x|. \end{aligned}$$

2.

$$\begin{aligned} g\colon \mathbb{R}_{\geq 0} &\to \mathbb{R} \\ x &\mapsto |x|. \end{aligned}$$

3.

$$\begin{aligned} h\colon \mathbb{R}_{\geq 0} &\to \mathbb{R}_{\geq 0} \\ x &\mapsto |x|. \end{aligned}$$

Definition. The *identity function* on a set S is the function

$$\begin{aligned} \mathrm{id}_S\colon S &\to S \\ x &\mapsto x. \end{aligned}$$

One may show the following:

1. The function $f\colon A \to B$ has an inverse if and only if there exists a function $g\colon B \to A$ such that
$$f \circ g = \mathrm{id}_B \quad \text{and} \quad g \circ f = \mathrm{id}_A.$$
In this case, $g = f^{-1}$.

2. If g is the inverse of f, then f is the inverse of g.

3. If $f\colon A \to B$ and $g\colon B \to C$ have inverses, then so does $g \circ f\colon A \to C$, and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. (Note the reversal of the order of the compositions. You put your sock on before putting on your shoe. To reverse the process, you reverse the order: first take off your shoe, then take off your sock.)

Week 3, Friday: Modular arithmetic

Modular arithmetic

(Supplemental reading: Example 2.3.10 in Swanson.)

Fix $n \in \mathbb{Z}$, and recall $\mathbb{Z}/n\mathbb{Z}$, the integers modulo n. Its elements are the equivalence classes[1], for the equivalence relation defined by $a \sim b$ if $a - b = kn$ for some $k \in \mathbb{Z}$.

Example. The elements of $\mathbb{Z}/5\mathbb{Z}$ are the equivalence classes for the integers modulo 5:

$$\begin{aligned}
[0] &= \{\dots, -15, -10, -5, 0, 5, 10, 15, \dots\} \\
[1] &= \{\dots, -14, -9, -4, 1, 6, 11, 16, \dots\} \\
[2] &= \{\dots, -13, -8, -3, 2, 7, 12, 17, \dots\} \\
[3] &= \{\dots, -12, -7, -2, 3, 8, 13, 18, \dots\} \\
[4] &= \{\dots, -11, -6, -1, 4, 9, 14, 19, \dots\}
\end{aligned}$$

Note that the equivalence classes *partition* $\mathbb{Z}$: each element of $\mathbb{Z}$ is in exactly one of these sets. We know that must be the case since we are working with an equivalence relation. Also, recall that the equivalence classes do not have unique names. For instance, $[1] = [6] = [-14]$. Any two elements of the same equivalence class may serve as representatives for the equivalence class. We have chosen the "standard representatives" in this case.

There are exactly n equivalence classes for $\mathbb{Z}/n\mathbb{Z}$:

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \dots, [n-1]\}.$$

This follows from a standard result of elementary number theory (which we will assume without proof): Each integer a has a unique remainder $r \in \{0, 1, \dots, n-1\}$ upon divison by n. It follows that $a = r + kn$, and $[a] = [r]$. For instance, in the above example, note that the equivalence class [2] consists of all those integers whose remainer upon division by 5 is equal to 2. To find the remainder upon division by 5, we can add or subtract 5s until we get to a number between 0 and 4. The difference between

[1] Recall that if $\sim$ is an equivalence relation on a set A, then the *equivalence class* for an element $a \in A$ is the subset of A defined by $[a] := \{x \in A : x \sim a\}$.

that number and the original number will be some multiple of 5, and hence the two numbers will be equivalent (and therefore belong to the same equivalence class).

Definition. The numbers $0, \ldots, n-1$ are called the *standard representatives* for the elements of $\mathbb{Z}/n\mathbb{Z}$.

Important notation. Again, fix $n \in \mathbb{Z}$ and let $\sim$ be the equivalence relation on $\mathbb{Z}$ defined by $a \sim b$ if $a - b = nk$ for some $k \in \mathbb{Z}$. Then, if $a \sim b$, we write

$$a = b \bmod n$$

and say *a is equal to b mod (or modulo) n.*

Example. We have

$$28 = 22 \bmod 3, \quad 33333 = 0 \bmod 3, \quad 7 = 12 \bmod 5, \quad 134 = 4 = 9 = -6 \bmod 5.$$

Note that 28 and 22 both have a remainder of 1 upon division by 3. Therefore, they both equal 1 modulo 3. In general, two numbers are the same modulo n if and only if they have the same remainder upon division by n.

Modular addition and multiplication

It turns out the addition and multiplication modulo n have a pleasant and extremely useful property:

Proposition 1. Let $a, a', b, b', n \in \mathbb{Z}$ and suppose that

$$a' = a \bmod n \quad \text{and} \quad b' = b \bmod n.$$

Then

$$a' + b' = a + b \bmod n \quad \text{and} \quad a'b' = ab \bmod n.$$

Proof. Since $a = a' \bmod n$ and $b = b' \bmod n$, there are integers k and ℓ such that

$$a' = a + kn \quad \text{and} \quad b' = b + \ell n.$$

We then have

$$a' + b' = (a + kn) + (b + \ell n) = (a + b) + (k + \ell)n.$$

Thus, $(a' + b') - (a + b)$ is a multiple of n. This means

$$a' + b' = a + b \bmod n.$$

Similarly,

$$a'b' = (a + kn)(b + \ell n) = ab + (a\ell + kb + k\ell n)n.$$

So $a'b'$ and ab differ by a multiple of n. Thus,

$$a'b' = ab \bmod n.$$

□

Example. We have

$$2 + 4 = 1 \bmod 5 \quad \text{and} \quad 2 \cdot 4 = 3 \bmod 5.$$

Now, pick some random integers equivalent to 2 and 4 modulo 5, say 12 and 119, respectively: $2 = 12 \bmod 5$ and $4 = 119 \bmod 5$. Compare the following calculation with the one we just did:

$$12 + 119 = 131 = 1 \bmod 5 \quad \text{and} \quad 12 \cdot 119 = 1428 = 3 \bmod 5.$$

The point is: *it doesn't matter which representatives we pick for equivalences classes when doing arithmetic modulo n.*

Here is another example: it turns out that $6^{1234} = 1 \bmod 5$. There are two ways to see this—one much harder than the other. The first way would be to multiply 6 by itself 1234 times to find

$$6^{1234} = 1732379250784311705176250882257708301496015396192567495371732614880752400871169994529068236589846749362776819680548436903764998373282710761889295067153626224411626400828902110953397848321171649505344345091523804741479399403475492481577374563714344875179653434497389873227035019055954428562022397680573941777462364995879482333585428161616053043017565934169582942626743609111951678685089418568380365655592065035761568184740932154765308074525500549643420339660570902440119972233854360573349265714550130678661859347154787046323758952117165815423383201126915413514946378616402788079378042112779332622008421219605849378881375846449614678547997867027771586431623496521692936194869871189110975700858236818633323144274142613679161190931094133860886932238754383337441616446755433877978151676971336837648216013452195410646025186312119047035620428437436577556869185659352811910096895578226171464048835337975361593424475051581642484227301745969638555856404916265709727645696 = 1 \bmod 5.$$

The other is to first note that $6 = 1 \bmod 5$, and use the fact that multiplication "plays nicely" (to paraphrase Proposition 1) with equivalence modulo 5:

$$6^{1234} = 1^{1234} = 1 \bmod 5.$$

In fact, $6^n = 1 \bmod 5$ for any $n \in \mathbb{N}$. Here is a similar example: since $4 = -1 \bmod 5$, we have

$$4^{1234567} = (-1)^{1234567} = -1 = 4 \bmod 5,$$

and

$$4^{1234568} = (-1)^{1234568} = 1 \bmod 5,$$

Definition (Addition and multiplication for $\mathbb{Z}/n\mathbb{Z}$). For each $[a], [b] \in \mathbb{Z}/n\mathbb{Z}$, define

$$[a] + [b] := [a + b] \quad \text{and} \quad [a][b] = [ab].$$

The first thing to notice in the expression $[a] + [b] = [a + b]$ is that the $+$ on the right-hand side and the $+$ on the left-hand side are different. The $+$ on the right-hand

side is ordinary addition of integers. The + on the left-hand side is a new operation: it defines how to add not integers but equivalence classes of integers. The meaning of the + on the left-hand side is a new thing which we are just now defining. It says this: in order to add two equivalence classes: (i) choose any representative integers for those classes, say a and b, (ii) add the representatives as integers, $a + b \in \mathbb{Z}$, and then (iii) take the equivalence class of the result, $[a+b]$. Similar remarks hold for multiplication.

Since addition and multiplication depend on the representatives a and b that we choose for the equivalence classes, we need to make sure that the resulting sum $[a + b]$ does not depend on that choice. So suppose that a' and b' are different choices for these equivalence classes. In other words, suppose that

$$[a] = [a'] \quad \text{and} \quad [b] = [b'].$$

We need to make sure that

$$[a] + [b] = [a'] + [b'] \quad \text{and} \quad [a][b] = [a'][b'].$$

Now, by the definition given just above, $[a] + [b] := [a + b]$ and $[a'] + [b'] := [a' + b']$, and similarly for multiplication.[2] So we need to check that

$$[a + b] = [a' + b'] \quad \text{and} \quad [ab] = [a'b'].$$

Proposition 1 comes to the rescue: since $[a] = [a']$, we have $a = a' \bmod n$, and similarly, $b = b' \bmod n$. Proposition 1 then says $a + b = a' + b' \bmod n$ and $ab = a'b' \bmod n$. In other words, $a + b$ and $a' + b'$ are both representatives of the same equivalence class and similarly for ab and $a'b'$, as desired.

Example. Here are addition and multiplication tables for $\mathbb{Z}/4\mathbb{Z}$. **Note:** For ease of notation in the tables below, we use standard representatives to represent equivalence classes. In other words, we will write a instead of $[a]$ where $a \in \{0, 1, 2, 3\}$:

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

.

[2] The notation := means "is defined by".

Week 4, Friday: Field axioms

Field axioms

(Supplemental reading: Example 2.6 in Swanson.)

Our next goal is to make a short, complete list of the rules characterizing the real numbers. Anything that can be said about the real numbers will follow from these rules, and, roughly, the real numbers is the only object that satisfies these rules. There will be three parts: (i) the nine field axioms, (ii) the four order axioms, and (iii) the completeness axiom. For right now, we will concentrate on the field axioms.

Definition. A *field* is a set F with two operations[1]:

$$+ : F \times F \to F \text{ (addition)} \quad \text{and} \quad \cdot : F \times F \to F \text{ (multiplication)},$$

satisfying the following axioms:

A1. Addition is *commutative*. For all $x, y \in F$,

$$x + y = y + x.$$

A2. Addition is *associative*. For all $x, y, z \in F$,

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

A3. There is an *additive identity*. There is an element of F, usually denoted 0, such that for all $x \in F$,

$$x + 0 = x.$$

A4. There are *additive inverses*. For all $x \in F$, there is an element $y \in F$ such that

$$x + y = 0.$$

The element y is denoted $-x$. Thus,$-x$ *is the element of* F *which when added to* x *yields* 0. (*Subtraction* is then defined by $x - y := x + (-y)$ for all $x, y \in F$.)

[1] In this context, "operation" is just another word for "function". Our functions take an ordered pair of elements of F and return another element of F.

M1. Multiplication is *commutative*. For all $x, y \in F$,

$$xy = yx.$$

M2. Multiplication is *associative*. For all $x, y, z \in F$,

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

M3. There is a *multiplicative identity*. There is an element, usually denoted 1, such that:

(a) $1 \neq 0$, and

(b) $1x = x$ for all $x \in F$.

M4. There are *multiplicative inverses*. For each *nonzero* $x \in F$, there is a $y \in F$ such that

$$xy = 1.$$

The element y is denoted $1/x$ or x^{-1}. Thus, $1/x$ *is the element of* F *which when multiplied by* x *yields* 1. (*Division* is then defined by $x/y := xy^{-1}$ for nonzero y.)

D. Multiplication *distributes* over addition. For all $x, y, z \in F$,

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

Thus, there are four field axioms governing addition, four governing multiplication, and one dictating how addition and multiplication interact.

Remark. The axioms for addition and multiplication are quite similar. Here are two subtle things to notice, however: (i) by definition, the additive and multiplicative identities are not equal ($0 \neq 1$), and (ii) only *nonzero* elements of a field are required to have multiplicative inverses.

Examples. The examples of fields with which you are most familiar are the rationals, $\mathbb{Q}$, and the reals, $\mathbb{R}$. Later in this course, we will consider the field of complex numbers, $\mathbb{C}$. It turns out that $\mathbb{Z}/n\mathbb{Z}$, with the addition and multiplication we defined earlier, is a field if and only if n is a prime number.[2] The only axiom that is not satisfied by $\mathbb{Z}/n\mathbb{Z}$ for general n is the existence of multiplicative inverses for nonzero elements (M4).

[2]Indeed, although $\mathbb{Z}/n\mathbb{Z}$ has many uses, the only two reasons it was introduced in this course is to help with the understanding of equivalence relations and field axioms.

Consider the addition and multiplication tables for $\mathbb{Z}/5\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$ (where we write k instead of $[k]$ for each equivalence class):

$\mathbb{Z}/5\mathbb{Z}$

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

$\mathbb{Z}/6\mathbb{Z}$

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	5	2	0	4	2
5	0	5	4	3	2	1

.

The additive identity for $\mathbb{Z}/5\mathbb{Z}$ is $[0]$. Note that axiom A3 says the identity is usually denoted 0, and we will often do that in the case of $\mathbb{Z}/n\mathbb{Z}$, but note that in the case of $\mathbb{Z}/5\mathbb{Z}$, for instance, 0 is really $[0] = \{\dots, -10, -5, 0, 5, 10, \dots\}$. The additive identity for $\mathbb{Z}/6\mathbb{Z}$ is $[0] = \{\dots, -12, -6, 0, 6, 12, \dots\}$, again, an infinite set.

Commutativity of addition and multiplication for $\mathbb{Z}/n\mathbb{Z}$ can seen in the above examples via the symmetry of the addition and multiplication tables about the northwest-to-southeast diagonal.

Here is an **important point**: if x is an element of a field, then $-x$ is *defined* to be the additive inverse of x, i.e., the field element which when added to x gives the additive identity, 0. Similarly, if x is nonzero, then $1/x$ is *defined* to be the multiplicative inverse of x, i.e., the field element which when multiplied by x gives the multiplicative identity 1. For instance, since $[2][3] = [1]$ in $\mathbb{Z}/5\mathbb{Z}$, we have that

$$\frac{1}{[2]} = [3] \quad \text{and} \quad \frac{1}{[3]} = [2].$$

If we know that we are working in $\mathbb{Z}/5\mathbb{Z}$, we might abbreviate these to $1/2 = 3$ and $1/3 = 2$.

Note that every nonzero element of $\mathbb{Z}/5\mathbb{Z}$ has a multiplicative inverse. (It turns out that $\mathbb{Z}/5\mathbb{Z}$ satisfies all of the field axioms.) On the other hand, the only nonzero elements of $\mathbb{Z}/6\mathbb{Z}$ with multiplicative inverses are 1 and 5. You can see this in the multiplication table, above: only the columns for 1 and 5 contain 1s. (In general, the nonzero elements of $\mathbb{Z}/n\mathbb{Z}$ with multiplicative inverses turn out to be $[k]$ such that k and n share no prime factors in common.)

A final fun fact: the smallest field is $\mathbb{Z}/2\mathbb{Z}$. Every field needs an additive identity, 0 and a distinct multiplicative identity 1, and $\mathbb{Z}/2\mathbb{Z}$ has no elements besides these.

Non-examples.

- The set of natural numbers, $\mathbb{N} = \{0, 1, 2, \dots\}$ with its usual addition and multiplication does not form a field. It violates axioms A4 and M4 (the existence of additive and multiplicative inverses). For instance, no natural number besides 0 has an additive inverse (0 is its own additive inverse), and no natural number besides 1 has a multiplicative inverse.

- The set of integers, $\mathbb{Z}$, satisfies all of the axioms except M4: no integers besides ± 1 have multiplicative inverses.

- Consider the set $X := \mathbb{R} \setminus \mathbb{Q}$ with ordinary addition and multiplication. Then X is not a field. For instance, it does not have an additive or a multiplicative identity since both 0 and 1 are elements of $\mathbb{Q}$. There is another serious problem. Consider the elements $\pm\sqrt{2} \in X$. We have $-\sqrt{2} + \sqrt{2} = 0 \notin X$. So addition is not defined for X (recall that addition for X would be a function $X \times X \to X$, so the result of the sum of two elements of X must be an element of X).

As said earlier, everything that can be known about the real numbers follows from the fact that the reals satisfy the field axioms (along with the order axioms and the completeness axiom, which we will examine later). For instance, we all know that if x is a real number then $x \cdot 0 = 0$. But why is that? Can you prove it? Note that it is not one of the field axioms. The following proposition shows that this result holds in any field (including, for example, $\mathbb{Z}/5\mathbb{Z}$). One may think of this as a game: the nine field axioms are the rules, and you need to use them to show $x \cdot 0 = 0$. It is surprisingly tricky (see the first displayed line of the proof—how could one know that's a reasonable first step?)!

Proposition. Let F be a field, and let $x \in F$. Then $x \cdot 0 = 0$.

Proof. We have

$$\begin{aligned} x \cdot 0 &= x(0 + 0) && \text{(since 0 is the additive identity)} \\ &= x \cdot 0 + x \cdot 0 && \text{(distributivity).} \end{aligned}$$

Since F is a field, we know $x \cdot 0 \in F$ and hence has an additive inverse $-(x \cdot 0)$.

Continuing from above,

$$
\begin{aligned}
& x \cdot 0 = x \cdot 0 + x \cdot 0 \\
& \quad \Rightarrow \quad -(x \cdot 0) + x \cdot 0 = -(x \cdot 0) + (x \cdot 0 + x \cdot 0) \\
& \quad \Rightarrow \quad 0 = -(x \cdot 0) + (x \cdot 0 + x \cdot 0) && \text{(definition of additive inverse)} \\
& \quad \Rightarrow \quad 0 = (-(x \cdot 0) + x \cdot 0) + x \cdot 0 && \text{(associativity of addition)} \\
& \quad \Rightarrow \quad 0 = 0 + x \cdot 0 && \text{(definition of additive inverse)} \\
& \quad \Rightarrow \quad 0 = x \cdot 0 && \text{(0 is the additive identity).}
\end{aligned}
$$

□

Week 4, Friday: Field axioms addendum

Here we provide several more examples of implications of the field axioms. Each of the results is a basic property of $\mathbb{Q}$ and of $\mathbb{R}$. We would like to see how they follow solely from the field axioms (and, thus, hold for any field).

First, the field axioms dictate the existence of an additive identity, but they nowhere stipulate that the additive identity is *unique*. Could there be two distinct elements of a field, say 0_1 and 0_2, such that both $0_1 + x = x$ and $0_2 + x = x$ for all x in the field? It turns out that the answer is "no":

Proposition. Let F be a field. Then F has a unique additive identity.

Proof. Suppose that 0_1 and 0_2 are both additive identities for F. This means that

$$0_1 + x = x \quad \text{and} \quad 0_2 + x = x$$

for all $x \in F$. Then,

$$\begin{aligned} 0_1 &= 0_1 + 0_2 && \text{(since } 0_2 \text{ is an additive identity)} \\ &= 0_2 && \text{(since } 0_1 \text{ is an additive identity).} \end{aligned}$$

Thus, $0_1 = 0_2$. □

A similar argument proves that a field has a unique multiplicative identity.

Proposition. (Cancellation law for addition.) Let F be a field, and let $x, y, z \in F$. Then

$$x + y = x + z \quad \Rightarrow \quad y = z.$$

Proof. Since F is a field, x has an additive inverse. Therefore,

$$\begin{aligned} x + y = x + z \quad &\Rightarrow \quad -x + (x + y) = -x + (x + z) && \text{(a property of } =) \\ &\Rightarrow \quad (-x + x) + y = (-x + x) + z && \text{(associativity of addition)} \\ &\Rightarrow \quad 0 + y = 0 + z && \text{(definition of } -x) \\ &\Rightarrow \quad y = z && (\ 0 \text{ is the additive identity).} \end{aligned}$$

□

Proposition. Let F be a field, and let $x, y \in F$. Then $xy = 0$ if and only if $x = 0$ or $y = 0$.

Proof. ($\Rightarrow$) Suppose that $xy = 0$ but $x \neq 0$. Since F is a field, x has a multiplicative inverse x^{-1}. Then

$$\begin{aligned} xy = 0 &\Rightarrow x^{-1}(xy) = x^{-1} \cdot 0 && \text{(property of =)} \\ &\Rightarrow x^{-1}(xy) = 0 && \text{(previous proposition)} \\ &\Rightarrow (x^{-1}x)y = 0 && \text{(associativity of } \cdot\text{)} \\ &\Rightarrow 1 \cdot y = 0 && \text{(definition of } x^{-1}\text{)} \\ &\Rightarrow y = 0 && \text{(1 is the mult. identity).} \end{aligned}$$

Similarly, if $y \neq 0$, then $x = 0$.

($\Leftarrow$) Conversely, if either $x = 0$ or $y = 0$, then the previous proposition implies $xy = 0$. □

Example. In contrast to the proposition we just proved, note that in $\mathbb{Z}/6\mathbb{Z}$, we have $[2] \neq [0]$ and $[3] \neq [0]$, but

$$[2][3] = [0].$$

Thus, $\mathbb{Z}/6\mathbb{Z}$ is not a field.

Proposition. Let F be a field, and let $x \in F$. Then

$$(-1)x = -x.$$

Proof. (Note that there is something to prove here! On the left we have the product of the additive inverse of 1 with x, and on the right we have the additive inverse of x. To show they are equal, we need to show $(-1)x$ is the additive inverse of x, i.e., if we add it to x, we get 0.)

Compute:

$$\begin{aligned} x + (-1)x &= 1 \cdot x + (-1)x && \text{(1 is the multiplicative identity)} \\ &= (1 + (-1))x && \text{(distributivity)} \\ &= 0 \cdot x && \text{(definition of } -1\text{)} \\ &= 0 && \text{(proved earlier).} \end{aligned}$$

□

Week 5, Monday: Order axioms

Order axioms

(Supplemental reading: Example 2.7 in Swanson.)

The real numbers have more structure than that described by the field axioms. For instance, we know that, for instance, $\sqrt{2} < 6$, which does not follow from just the field axioms. What does "<" even mean here? A complete answer to that question would require a rigorous definition of $\mathbb{R}$. Instead, we first summarize the essential properties.

Definition. An *ordered field* is a field F with a relation, denoted $<$, satisfying

O1. (Trichotomy) For all $x, y \in F$, exactly one of the following statements is true:

$$x < y, \quad y < x, \quad x = y.$$

O2. (Transitivity) The relation $<$ is transitive, i.e., for all $x, y, z \in F$,

$$x < y \quad \text{and} \quad y < z \quad \Longrightarrow \quad x < z.$$

O3. (Additive translation) For all $x, y, z \in F$,

$$x < y \quad \Longrightarrow \quad x + z < y + z.$$

O4. (Multiplicative translation) For all $x, y, z \in F$,

$$x < y \quad \text{and} \quad 0 < z \quad \Longrightarrow \quad xz < yz.$$

Remark. We write $x > y$ if $y < x$, and we write $x \leq y$ if either $x = y$ or $x < y$, and so on.

In the next proposition, we list several properties of $<$ with which you are familiar in the context of $\mathbb{Q}$ or $\mathbb{R}$. We will see that these properties follow from the order axioms, alone.

Proposition 1. Let x, y, z, w be elements of an ordered field F. Then

1. $x < y \Rightarrow -y < -x$.

2. $x < y$ and $z < 0 \Rightarrow xz > yz$.

3. $x^2 > 0$ if $x \neq 0$.

4. $1 > 0$ and $-1 < 0$.

5. $w < x$ and $y < z \Rightarrow w + y < x + z$.

Proof. 1. We have

$$\begin{array}{lcll} x < y & \Rightarrow & -x + x < -x + y & \text{(additive translation)} \\ & \Rightarrow & -y + (-x + x) < -y + (-x + y) & \text{(additive translation)} \\ & \Rightarrow & -y + (-x + x) < (-y + y) + (-x) & \text{(assoc. and commutativity of } +) \\ & \Rightarrow & -y + 0 < 0 + (-x) & \text{(def. of } -x \text{ and } -y) \\ & \Rightarrow & -y < -x & \text{(0 is the additive identity).} \end{array}$$

2. If $z < 0$, then $-z > 0$ by part 1. Then, by multiplicative translation,

$$\begin{array}{lcll} x < y & \Rightarrow & (-z)x < (-z)y & \\ & \Rightarrow & -(zx) < -(zy) & \text{(exercise: } (-a)b = -(ab)) \\ & \Rightarrow & -(-(zx) > -(-(zy)) & \text{(by part 1)} \\ & \Rightarrow & zx > zy & \text{(exercise: } -(-a) = a). \end{array}$$

3. Suppose that $x \neq 0$. By trichotomy, either $x > 0$ or $x < 0$. We consider these cases separately. If $x > 0$, then use multiplicative translation:

$$x > 0 \quad \Rightarrow \quad x \cdot x > x \cdot 0 \quad \Rightarrow \quad x^2 > 0.$$

   If $x < 0$, use part 2:

$$x < 0 \quad \Rightarrow \quad x \cdot x > x \cdot 0 \quad \Rightarrow \quad x^2 > 0.$$

4. By part 3, we have $1 = 1^2 > 0$. Then applying part 1,

$$0 < 1 \quad \Rightarrow \quad -1 < -0 \quad \Rightarrow \quad -1 < 0.$$

5. Left as an exercise.

□

Exercise. Can the field $\mathbb{Z}/5\mathbb{Z}$ be ordered, i.e., can a relation $<$ on $\mathbb{Z}/5\mathbb{Z}$ be defined that satisfies the order axioms?

Absolute value. Here we define the absolute value function for a ordered field, and prove some of its basic properties. We then prove one of the most important theorems in analysis—the triangle inequality. It will soon become one of our main tools.

Definition. Let F be an ordered field. The *absolute value* of $x \in F$ is

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

Proposition 2. For all x in an ordered field F,

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

Proof. There are three cases (by trichotomy):

$x = 0$: In this case $-|x| = x = |x| = 0$, and the result holds.

$x > 0$: In this case $|x| = x$. Hence, $x \leq x = |x|$, which is the right half of the result. For the left half, note that $|x| = x > 0$, implies $-|x| < 0$. Then, since $-|x| < 0$ and $0 < x$, by transitivity $-|x| \leq x$.

$x < 0$: In this case, $|x| = -x > 0$, which implies $x < 0 < |x|$. By transitivity, $x \leq |x|$. On the other hand $|x| = -x$ implies $-|x| = x$. In particular, $-|x| \leq x$. □

Proposition 3. If x and a are elements of an ordered field F, then $|x| \leq a$ if and only if $-a \leq x \leq a$.

Proof. This proof is similar to that of Proposition 1 and is left as an exercise for the interested reader. □

As mentioned above, the following result will be very important in the sequel. It is worth reading its short proof.

Theorem. (Triangle inequality.) If x and y are elements of an ordered field, then

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

Proof. By Proposition 2, we have

$$-|x| \leq x \leq |x| \quad \text{and} \quad -|y| \leq y \leq |y|.$$

By Proposition 1, part 5, we can add these inequalities to get

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

and, hence,

$$-(|x| + |y|) \leq x + y \leq |x| + |y|.$$

The result now follows by letting $a = |x| + |y|$ in Proposition 3. □

Example. Here are several instances of the triangle inequality in the case of $\mathbb{Q}$ or $\mathbb{R}$:

$$5 = |2 + 3| \leq |2| + |3| = 5$$

$$1 = |-2 + 3| \leq |-2| + |3| = 5$$

$$5 = |-2 - 3| \leq |-2| + |-3| = 5.$$

When, in general, does *equality* hold in the triangle inequality?

Week 5, Wednesday: Completeness

Completeness

(Supplemental reading: Example 2.7 in Swanson.)

We now come to the last axiom we need to characterize the real numbers: completeness. Let F be any ordered field, and let $S \subseteq F$. To state the axiom, we first need some vocabulary:

- $B \in F$ is an *upper bound* for S if $s \leq B$ for all $s \in S$,
- $b \in F$ is an *lower bound* for S if $b \leq s$ for all $s \in S$,
- S is *bounded* if it has both an upper bound and a lower bound.

Example. Let $S := (0,1) := \{x \in \mathbb{R} : 0 < x < 1\}$, a subset of the ordered field $\mathbb{R}$. Then any number greater than or equal to 1 is an upper bound for S. For example, 1, π, 42, and 10^6 are all upper bounds. Similarly, any number less than or equal to 0 is a lower bound. For example, 0, -1, $-\pi$ are all lower bounds. Not every set has a lower bound. For example, if $S = \mathbb{Z}$, the integers, then S is a subset of the ordered field $\mathbb{R}$ which has neither an upper bound nor a lower bound.

Some more vocabulary:

- $B \in F$ is a *supremum* for S if it is a *least upper bound*. This means that B is an upper bound and if B' is any upper bound, then $B \leq B'$. If B exists, then we write $B = \sup(S)$ or $B = \text{lub}(S)$.
- $b \in F$ is an *infimum* for S if it is a *greatest lower bound*. This means that b is a lower bound and if b' is any lower bound, then $b' \leq b$. If b exists, then we write $b = \inf(S)$ or $b = \text{glb}(S)$.

Examples.

1. If $S = (0,1) \subset \mathbb{R}$, then $\sup(S) = 1$ and $\inf(S) = 0$.

   > **Important:** Note that, as in this example, *the supremum and infimum of a set are not necessarily elements in the set.*

2. Let $S = [0, 1) = \{x \in \mathbb{R} : 0 \leq x < 1\} \subset \mathbb{R}$. Then $\sup(S) = 1$ and $\inf(S) = 0$, as before. However, this time $\inf(S) \in S$ while $\sup(S) \notin S$.

3. If $S = (-2, \infty) = \{x \in \mathbb{R} : -2 < x\} \subset \mathbb{R}$, then $\sup S$ does not exist and $\inf S = -2$.

4. If $S = \{1/n : n = 1, 2, 3, \dots\} \subset \mathbb{R}$, then $\sup(S) = 1 \in S$ and $\inf(S) = 0 \notin S$. Note that in this example, finding the infimum is a way to take a limit of the sequence of rationals $1, 1/2, 1/3, 1/4, \dots$

We've just seen that if a set S has a sup or inf, then it may or may not be the case that either is contained in S. This leads to a bit more vocabulary: if S has a supremum B and $B \in S$, then we call B the *maximum* or *maximal element* of S and write $\max(S) = B$. Similarly, if S has an infimum b and $b \in S$, then we call b the *minimum* of *minimal element* of S and write $\min(S) = b$.

Example. If $S = (0, 1] \subset \mathbb{R}$, then $\max(S) = 1$ and $\min(S)$ does not exist and $\inf(S) = 0$.

Definition. An ordered field F is *complete* if every *nonempty* subset of F which is bounded above has a supremum.

Consider the ordered field of rational numbers, $\mathbb{Q}$, and pick any sequence of rational numbers converging to π, say by just truncating the decimal expansion, and put these numbers into a set

$$S = \{3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415,\ 3.14159,\ \dots\}.$$

Each of the numbers in S is rational, e.g., $3.14 = \frac{314}{100}$. Does this set of rational numbers have a supremum? The answer depends on which field we are working in. In $\mathbb{R}$, the answer is "yes": $\sup S = \pi$ (and it's not in S). However, in $\mathbb{Q}$ it does not have a supremum. To see this, suppose that $B \in \mathbb{Q}$. If $\pi < B$, then there is a rational number B' such that $\pi < B' < B$.[1] So B' is an upper bound for S that is smaller than B. Therefore, in this case B is not a least upper bound. On the other hand, if $B < \pi$, then there is an element $s \in S$ such that $B < s$ (since we can get arbitrarily close to π by taking the decimal expansion for π and truncating it sufficiently far out). In this case, B is not an upper bound for S. The rational numbers have this "defect": there are nonempty subsets of $\mathbb{Q}$ that are bounded above but have no least upper bound. Thus, both $\mathbb{Q}$ and $\mathbb{R}$ are ordered fields, but of these two, only $\mathbb{R}$ also satisfies the completeness axiom. (Of course, we have not *proved* these properties of $\mathbb{Q}$ and $\mathbb{R}$ since, in fact, we have not even given definitions for either of these field. Instead, we are appealing to the reader's prior experience with these number systems.)

[1] One way to create B' is to go out in the decimal expansion for π far enough to get a number much closer to π than to B, then slightly round up that number by adding 1 to the last decimal place. The resulting number will be slightly bigger than π and less than B.

Challenge. Show that an order field is complete if and only if every nonempty subset of F which is bounded below has an infimum.

Theorem. There exists a unique complete ordered field. By *uniqueness*, we mean that if F_1 and F_2 are complete ordered fields, then there exists a bijection $g\colon F_1 \to F_2$ such that for all $a, b \in F_1$,

1. $g(a+b) = g(a) + g(b)$,
2. $g(ab) = g(a)g(b)$,
3. if $a > 0$, then $g(a) > 0$.

Proof. We will accept this result on faith for now and possibly return to it later. □

Since g is a bijection, we can think of it as simply a relabeling of the elements of F_1: each element $a \in F_1$ is now called $g(a) \in F_2$, instead. The listed requirements for g amount to saying that F_1 and F_2 are the same ordered fields up to relabeling.

We finally get to our ultimate goal:

Definition. The complete ordered field is called the field of *real numbers*, denoted $\mathbb{R}$.

While we have not given a construction of the real numbers, the previous theorem says that anything we can prove about the real numbers may be derived from the fourteen field axioms, the four order axioms, and the completeness axiom (i.e., that every nonempty subset that is bounded above has a supremum).

Important vocabulary. We will use the vocabulary introduced in this lecture extensively in the second half of this course. Here is a summary: First there are properties pertaining to subsets of an ordered field: upper bound, lower bound, bounded, supremum, infimum, maximum, and minimum. Second, there is a property an ordered field may possess: completeness. Please take a few moments now to review these terms, trying to come up with your own examples.

Week 5, Friday: Extrema

Extrema

(Supplemental reading: Sections 2.6 and 2.7 in Swanson.)

Before getting to the main propositions for today, we provide some templates for proving statements involving bounds for sets.

Proof template. Define a subset $S \subset \mathbb{R}$ by blah, blah, blah, and let B be the real number blah. Then B is an upper bound for S.

Proof. Let $s \in S$. Then blah, blah, blah. It follows that $s \leq B$. □

Proof template. Define a subset $S \subset F$ by blah, blah, blah, and let B be the real number blah. Then $B = \sup S$.

Proof. We first show that B is an upper bound for S. Let $s \in S$. Then blah, blah, blah. It follows $s \leq B$. Next we show that B is a least upper bound for S. Suppose that B' is an upper bound for S. Then blah, blah, blah. It follows that $B \leq B'$. □

An alternative, to the preceding template, which is sometimes easy to use:

Proof template. Define a subset $S \subset \mathbb{R}$ by blah, blah, blah, and let B be the real number blah. Then $B = \sup S$.

Proof. We first show that B is an upper bound for S. Let $s \in S$. Then blah, blah, blah. It follows $s \leq B$. Next we show that B is a least upper bound for S. Suppose $B' < B$. Then B' is not an upper bound for S since blah, blah, blah. □

The templates for lower bounds and infima are similar.

Example. Claim: Let $S = (-\infty, 1)$. Then $\sup(S) = 1$.

Proof. We first show that 1 is an upper bound for S. This follows immediately from the definition of $S = (-\infty, 1) := \{x \in \mathbb{R} : x < 1\}$. So if $s \in S$ then $s < 1$. Next, we show that 1 is a least upper bound for S. Suppose $x < 1$. Then $(x+1)/2$, the midpoint between x and 1, is an element of S and is greater than x, it follows that x is not an upper bound for S. In sum, 1 is an upper bound for S and anything smaller then 1 is not an upper bound. So $1 = \sup(S)$. □

We move on now to our main results for this lecture. The last two are especially interesting and important for what is to come.

Suppose F is an ordered field and $S \subseteq F$. Define the subset $-S \subseteq F$ by

$$-S := \{-s : s \in S\}.$$

For instance, if $S = (2, 5) \subset \mathbb{R}$, then $-S = (-5, -2)$. By our first proposition, below, $\inf(-S) = -5 = -\sup(S)$. Note that, in general, $-(-S) = S$.

Proposition 1. Using the notation from above,

(a) if $\inf(-S)$ exists, then $\sup(S)$ exists and $\sup(S) = -\inf(-S)$;

(b) if $\sup(S)$ exists, then $\inf(-S)$ exists and $\inf(-S) = -\sup(S)$.

Proof. We will just prove the first part, the second being similar. Suppose that $\inf(-S)$ exists, and for ease of notation, let $t = \inf(-S)$. We must show that $\sup(S)$ exists and $\sup(S) = -t$. We first show that $-t$ is an upper bound for S. Take $s \in S$. Then $-s \in -S$, and so it follows that $t \leq -s$ (by definition of $\inf(-S)$). It follows that $-t \geq s$. Hence, $-t$ is an upper bound for S. Next, suppose that x is any upper bound for S. We first claim that $-x$ is a lower bound for $-S$: given $y \in -S$, we have $-y \in S$, and hence, $x \geq -y$ (since x is an upper bound for S). It follows that $-x \leq y$. We have shown that $-x$ is a lower bound for $-S$. Since $t = \inf(-S)$ is the greatest lower bound for $-S$, we have $-x \leq t$. From this, it follows that $x \geq -t$. We have shown that $-t$ is the least upper bound for S, i.e., $-t = -\inf(-S) = \sup(S)$. □

For the next proposition, recall that an ordered field is complete if each of its nonempty subsets that is bounded above has a least upper bound, i.e., a supremum.

Proposition 2. Let F be an ordered field, and suppose that every nonempty subset of F that is bounded below has an infimum. Then F is complete.

Proof. Let $\emptyset \neq S \subseteq F$, and suppose S is bounded above, say by $B \in F$. We would like to show that S has a supremum. First, note that $-B$ is a lower bound for $-S$: if $x \in -S$, then $-x \in S$, and hence, $B \geq -x$. It follows that $-B \leq x$. Next, since $-S$ is bounded below, by hypothesis, $-S$ has an infimum $\inf(-S)$. Then by

Proposition 1, we see that $\sup(S)$ exists (and is equal to $-\inf(-S)$). We have shown that every nonempty subset of F that is bounded above has a least upper bound, i.e., F is complete. □

The next result shows that there are **no infinitely small positive elements** in any ordered field!

Proposition 3 (No infinitesimals.) Let F be an ordered field. Let $x \in F$ and suppose that

$$0 \leq x \leq y$$

for all $y \in F$ with $y > 0$. Then $x = 0$. In other words, the only nonnegative element of F that is less than or equal to all the positive elements of F is 0. No positive element of F can be less than or equal to all the positive elements of F.

Proof. Since $x \geq 0$, by trichotomy there are two possibilities: $x > 0$ or $x = 0$. For sake of contradiction, suppose that $x > 0$. Then let $y := x/2$. Below, using the order axioms, we will prove that $y < x$ and $y > 0$ in contradiction to the hypotheses concerning x, i.e., that $0 \leq x \leq y$ for $y > 0$. The only possibility then left is that $x = 0$, as desired.

The reasoning we just used for y is clearly in agreement with our experience with the rational or real numbers. However, we are working in an arbitrary ordered field. So we now check that, in general, if $x > 0$ and $y = x/2$, then $y < x$ and $y > 0$.

We have already shown that since F is an ordered field, $1 > 0$. By additive translation, adding 1 to both sides of the inequality, we get $2 := 1+1 > 1$. By transitivity, $2 > 1 > 0$ implies $2 > 0$. Therefore, $2 \neq 0$ (again by trichotomy). The point of all of the above is that $2 \neq 0$, and hence has a multiplicative inverse, which we naturally denote by $1/2$. Why is $1/2 > 0$, necessarily? Answer: by trichotomy, either $1/2 = 0$, $1/2 < 0$, or $1/2 > 0$. We rule out the first two possibilities. If $1/2 = 0$, then $1 = 2 \cdot (1/2) = 2 \cdot 0 = 0$. But $1 \neq 0$ in a field. If $1/2 < 0$, then multiplying through $2 > 0$ by $1/2$ would yield $1 < 0$, which we saw in an earlier lecture is impossible.

Continuing: since $2 > 1$, multiplicative translation by $1/2$ gives $1 > 1/2$. Then multiplicative translation by $x > 0$ gives $x > x/2$, i.e., $x > y$. Finally, to see that $y > 0$, start with $x > 0$. Multiplicative translation by $1/2 > 0$ yields $x/2 > (1/2) \cdot 0 = 0$, i.e., $y > 0$. □

The next proposition will be very useful later on. It says that if a set has a supremum, then even if the supremum is not an element of the set, then if can be approximated arbitrarily closely with an element of the set. For example, consider the interval $S := (0, 1) \subset \mathbb{R}$. Then $\sup(S) = 1 \notin S$. Can we approximate $\sup(S)$ to within a tolerance of $\varepsilon := 0.001$ by an element in the set? Sure: take $x := 1 - 0.0001 = 0.9999$, for instance. Then $x \in S$ and is within 0.001 of the supremum, 1.

Proposition 4. Let S be a subset of an ordered field F, and suppose that $M := \sup S$ exists. Given $\varepsilon \in F$ with $\varepsilon > 0$, there exists $s \in S$ such that $M - s < \varepsilon$.

Proof. To help with understanding this proof, here is a picture for the case where S is an interval in $F = \mathbb{R}$:

Let $s \in S$. Note that the statement $M - s < \varepsilon$ is equivalent to the statement $M - \varepsilon < s$. Further, if either of these statements holds, then since $M = \sup(S)$ and $s \in S$ it follows that, in addition to $M - \varepsilon < s$, we have $s \leq M$, i.e., s is in the interval $(M - \varepsilon, M]$. Thus, we are trying to argue that for any $\varepsilon > 0$, we can find an s in S that's between $M - \varepsilon$ and M. Of course, if $M \in S$, we could take $M = s$, but it's not always the case that a set contains its supremum.

The proof starts here. By the additive translation order axiom, we can add $M - \varepsilon$ to both sides of the inequality $0 < \varepsilon$ to get

$$(M - \varepsilon) + 0 < (M - \varepsilon) + \varepsilon.$$

Using associativity and addition and the definition of the additive inverse $-\varepsilon$, we see that

$$M - \varepsilon < M.$$

Since M is the least upper bound of S, we know that if M' is an upper bound of S, then $M \leq M'$. In particular, $M - \varepsilon$ cannot be an upper bound of S because $M - \varepsilon < M$. (In other words, since $M - \varepsilon$ is strictly smaller than the least upper bound, it cannot be an upper bound.) The fact that $M - \varepsilon$ is not an upper bound of S, means there exists some $s \in S$ such that $M - \varepsilon < s$. We now add $\varepsilon - s$ to both sides of this inequality to obtain the desired inequality:

$$M - s = (M - \varepsilon) + (\varepsilon - s) < s + (\varepsilon - s) = \varepsilon.$$

To summarize: since $M - \varepsilon$ is strictly smaller than the least upper bound of S, it cannot be an upper bound for S. This means that there exists some $s \in S$ such that $M - \varepsilon < s$, and the result follows. □

A similar proposition holds for infima.

Week 6, Monday: Complex numbers I

Complex numbers II

(Supplemental reading: Sections 3.2 and 3.3 in Swanson.)

The field $\mathbb{C}$ cannot be ordered. It turns out that $\mathbb{C}$ cannot be ordered. To see this, recall we showed that if x is any nonzero element in an ordered field, then $x^2 > 0$. In particular, letting $x = 1$, we saw that $1 > 0$, and then that $-1 < 0$. Suppose $\mathbb{C}$ could be ordered. Then, since i is nonzero, we would have that $i^2 > 0$, but in fact, $i^2 = -1 < 0$. By the trichotomy axiom, we cannot have both $i^2 > 0$ and $i^2 < 0$.

The triangle inequality for complex numbers. Our goal today is to prove the triangle inequality for the complex numbers. You may recall that we already proved the triangle inequality for any ordered field F:

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

for all $x, y \in F$. A problem arises immediately in trying to extend this result to the complex numbers: our definition of $|x|$ depended on F being ordered, yet, $\mathbb{C}$ can not be ordered. So we start with a new definition of "absolute value" that will work in the context of complex numbers.

Definition. Let $z = a + bi \in \mathbb{C}$. The *conjugate* of z is

$$\overline{z} := a - bi.$$

The *modulus* or *length* of z is

$$|z| := \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}.$$

The *real part* of z is

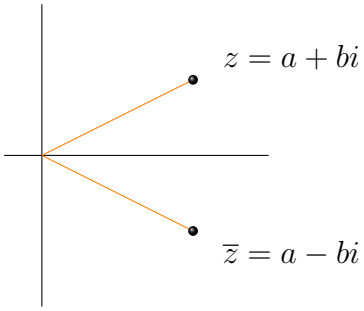
$$\mathrm{Re}(z) := a$$

and the *imaginary part* of z is

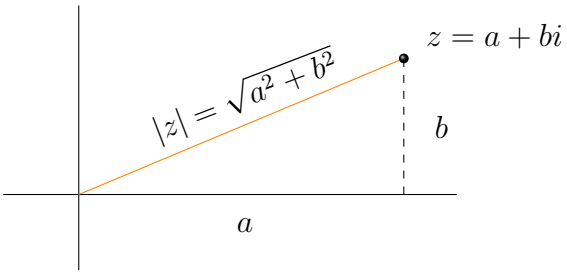
$$\mathrm{Im}(z) := b.$$

Remarks.

1. Geometrically, taking the conjugate amounts to flipping about the x-axis (also known as the *real axis* in the context of complex numbers):



2. The modulus of a complex number is the distance of the number from the origin:



3. Note that both the real and imaginary part of a complex number are real numbers. For instance $\mathrm{Im}(2 + 3i) = 3 \in \mathbb{R}$.

4. If $r \in \mathbb{R}$, we already have a definition of its absolute value

$$|r| := \begin{cases} r & \text{if } r \geq 0, \\ -r & \text{if } r < 0. \end{cases}$$

However, we may now consider $r = r + 0 \cdot i$ as a complex number and compute its modulus:

$$|r| = \sqrt{r^2 + 0^2} = \sqrt{r^2}.$$

Note that the expression on the right, $\sqrt{r^2}$, is the ordinary absolute value of r as a real number. So there is no conflict between $\mathbb{R}$ and $\mathbb{C}$ in our notation for $|r|$. In other words, our new definition of $|\ |$ generalizes our old definition.

Practice problems. (Solutions appear on the last page.)

1. Let $z = 4 + 2i + (5 + i)i$. Find $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$.

2. Let $z = 4 + 5i$. What is $\overline{z}$? What is $|z|$? What are $z + \overline{z}$ and $z - \overline{z}$?

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

1. $\overline{\overline{z}} = z$.
2. $|\overline{z}| = |z|$.
3. $\overline{z + w} = \overline{z} + \overline{w}$.
4. $\overline{zw} = \overline{z}\,\overline{w}$.
5. $|zw| = |z||w|$.
6. $|\mathrm{Re}(z)| \leq |z|$ and $|\mathrm{Im}(z)| \leq |z|$.
7. $z + \overline{z} = 2\mathrm{Re}(z)$ and $z - \overline{z} = 2i\,\mathrm{Im}(z)$.
8. (triangle inequality for complex numbers) $|z + w| \leq |z| + |w|$.
9. If $z \neq 0$, then $z/|z|$ is a *unit vector*, i.e.,

$$\left|\frac{z}{|z|}\right| = 1.$$

Proof.

1. Let $z = a + bi$. Then

$$\overline{\overline{z}} = \overline{\overline{(a + bi)}} = \overline{a - bi} = a + bi = z.$$

2. Using part 1,

$$|\overline{z}| = \sqrt{\overline{z}\,\overline{\overline{z}}} = \sqrt{\overline{z}\,z} = \sqrt{z\,\overline{z}} = z.$$

3. Let $z = a + bi$ and $w = c + di$. Then

$$\begin{aligned}
\overline{z + w} &= \overline{(a + bi) + (c + di)} \\
&= \overline{(a + c) + (b + d)i} \\
&= (a + c) - (b + d)i \\
&= (a - bi) + (c - di) \\
&= \overline{z} + \overline{w}.
\end{aligned}$$

4. Exercise.
5. $$|zw| = \sqrt{(zw)\overline{zw}} = \sqrt{(zw)(\overline{z}\,\overline{w})} = \sqrt{(z\overline{z})(w\overline{w})} = \sqrt{z\overline{z}}\sqrt{w\overline{w}} = |z||w|.$$
6. If $z = a + bi$, we have

$$|z| = \sqrt{a^2 + b^2} \geq \sqrt{a^2} = |a| = |\mathrm{Re}(z)|.$$

The result for the complex part of z is similar.

7. Let $z = a + bi$. Then

$$z + \overline{z} = (a + bi) + (a - bi) = 2a = 2\mathrm{Re}(z),$$

and

$$z - \overline{z} = (a + bi) - (a - bi) = 2bi = 2i\,\mathrm{Im}(z).$$

8.

$$\begin{aligned}
|z+w|^2 &= (z+w)(\overline{z+w}) && \text{(definition of modulus)}\\
&= (z+w)(\overline{z}+\overline{w}) && \text{(part 3)}\\
&= z\,\overline{z} + z\,\overline{w} + \overline{z}\,w + \overline{w}\,w\\
&= |z|^2 + z\,\overline{w} + \overline{z}\,w + |w|^2\\
&= |z|^2 + z\,\overline{w} + \overline{z}\,\overline{\overline{w}} + |w|^2 && \text{(part 1)}\\
&= |z|^2 + z\,\overline{w} + \overline{z\overline{w}} + |w|^2 && \text{(part 4)}\\
&= |z|^2 + 2\mathrm{Re}(z\,\overline{w}) + |w|^2 && \text{(part 6)}\\
&\le |z|^2 + 2|\mathrm{Re}(z\,\overline{w})| + |w|^2\\
&\le |z|^2 + 2|z\overline{w}| + |w|^2 && \text{(part 5)}\\
&= |z|^2 + 2|z||\overline{w}| + |w|^2 && \text{(part 4)}\\
&= |z|^2 + 2|z||w| + |w|^2 && \text{(part 2)}\\
&= (|z| + |w|)^2 .
\end{aligned}$$

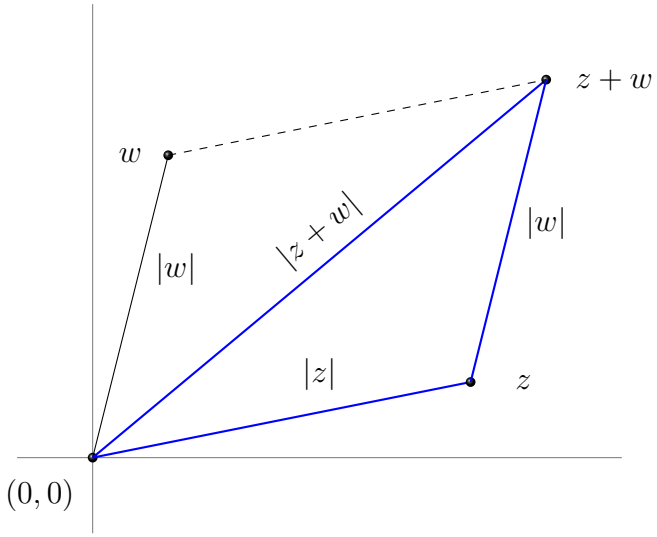
Now take square roots to get the result.

9.

$$\left|\frac{z}{|z|}\right| = \left|z \cdot \frac{1}{|z|}\right| = |z|\left|\frac{1}{|z|}\right| = |z|\frac{1}{|z|} = 1.$$

□

The picture below may help to give an intuitive understanding of the triangle inequality.



Solutions to practice problems.

1. We have
$$z = 4 + 2i + (5 + i)i = 4 + 2i + 5i - 1 = 3 + 7i.$$
Hence, $\mathrm{Re}(z) = 3$ and $\mathrm{Im}(z) = 7$.

2. If $z = 4 + 5i$, then
$$\begin{aligned} \overline{z} &= 4 - 5i \\ |z| &= \sqrt{4^2 + 5^2} = \sqrt{41} \\ z + \overline{z} &= 4 + 5i + 4 - 5i = 8 = 2\,\mathrm{Re}(z) \\ z - \overline{z} &= 4 + 5i - (4 - 5i) = 10i = 2i\,\mathrm{Im}(z). \end{aligned}$$

Week 6, Wednesday: Complex numbers II

Complex numbers II

(Supplemental reading: Sections 3.2 and 3.3 in Swanson.)

$\mathbb{C}$ cannot be ordered. It turns out that the field $\mathbb{C}$ cannot be ordered. To see this, recall we showed that if x is any nonzero element in an ordered field, then $x^2 > 0$. In particular, letting $x = 1$, we saw that $1 > 0$, and then that $-1 < 0$. Suppose $\mathbb{C}$ could be ordered. Then, since i is nonzero, we would have that $i^2 > 0$, but in fact, $i^2 = -1 < 0$. By the trichotomy axiom, we cannot have both $i^2 > 0$ and $i^2 < 0$.

The triangle inequality for complex numbers. Our goal today is to prove the triangle inequality for the complex numbers. You may recall that we already proved the triangle inequality for any ordered field F:

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

for all $x, y \in F$. A problem arises immediately in trying to extend this result to the complex numbers: our definition of $|x|$ depended on F being ordered, yet, $\mathbb{C}$ can not be ordered. So we start with a new definition of "absolute value" that will work in the context of complex numbers.

Definition. Let $z = a + bi \in \mathbb{C}$. The *conjugate* of z is

$$\overline{z} := a - bi.$$

The *modulus* or *length* of z is

$$|z| := \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}.$$

The *real part* of z is

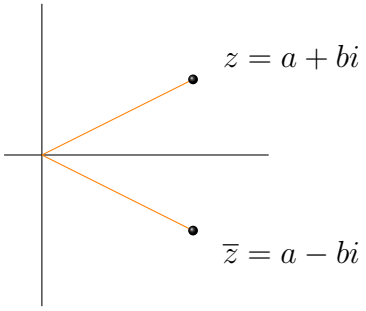
$$\mathrm{Re}(z) := a$$

and the *imaginary part* of z is

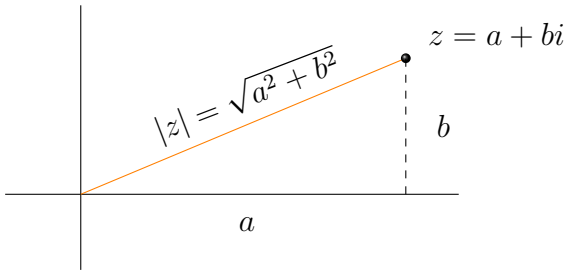
$$\mathrm{Im}(z) := b.$$

Remarks.

1. Geometrically, taking the conjugate amounts to flipping about the x-axis (also known as the *real axis* in the context of complex numbers):



2. The modulus of a complex number is the distance of the number from the origin:



3. Note that both the real and imaginary part of a complex number are real numbers. For instance $\mathrm{Im}(2 + 3i) = 3 \in \mathbb{R}$.

4. If $r \in \mathbb{R}$, we already have a definition of its absolute value

$$|r| := \begin{cases} r & \text{if } r \geq 0, \\ -r & \text{if } r < 0. \end{cases}$$

However, we may now consider $r = r + 0 \cdot i$ as a complex number and compute its modulus:

$$|r| = \sqrt{r^2 + 0^2} = \sqrt{r^2}.$$

Note that the expression on the right, $\sqrt{r^2}$, is the ordinary absolute value of r as a real number. So there is no conflict between $\mathbb{R}$ and $\mathbb{C}$ in our notation for $|r|$.

Practice problems. (Solutions appear on the last page.)

1. Let $z = 4 + 2i + (5 + i)i$. Find $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$.

2. Let $z = 4 + 5i$. What is $\overline{z}$? What is $|z|$? What are $z + \overline{z}$ and $z - \overline{z}$?

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

1. $\overline{\overline{z}} = z$.
2. $|\overline{z}| = |z|$.
3. $\overline{z+w} = \overline{z} + \overline{w}$.
4. $\overline{zw} = \overline{z}\,\overline{w}$.
5. $|zw| = |z||w|$.
6. $|\mathrm{Re}(z)| \leq |z|$ and $|\mathrm{Im}(z)| \leq |z|$.
7. $z + \overline{z} = 2\mathrm{Re}(z)$ and $z - \overline{z} = 2i\,\mathrm{Im}(z)$.
8. (triangle inequality for complex numbers) $|z+w| \leq |z| + |w|$.
9. If $z \neq 0$, then $z/|z|$ is a *unit vector*, i.e.,

$$\left|\frac{z}{|z|}\right| = 1.$$

Proof.

1. Let $z = a + bi$. Then
$$\overline{\overline{z}} = \overline{\overline{(a+bi)}} = \overline{a - bi} = a + bi = z.$$
2. Using part 1,
$$|\overline{z}| = \sqrt{\overline{z}\,\overline{\overline{z}}} = \sqrt{\overline{z}\,z} = \sqrt{z\,\overline{z}} = z.$$
3. Let $z = a + bi$ and $w = c + di$. Then
$$\begin{aligned}\overline{z+w} &= \overline{(a+bi)+(c+di)} \\ &= \overline{(a+c)+(b+d)i} \\ &= (a+c)-(b+d)i \\ &= (a-bi)+(c-di) \\ &= \overline{z} + \overline{w}.\end{aligned}$$
4. Exercise.
5. $$|zw| = \sqrt{(zw)\overline{zw}} = \sqrt{(zw)(\overline{z}\,\overline{w})} = \sqrt{(z\overline{z})(w\overline{w})} = \sqrt{z\overline{z}}\sqrt{w\overline{w}} = |z||w|.$$
6. If $z = a + bi$, we have
$$|z| = \sqrt{a^2 + b^2} \geq \sqrt{a^2} = |a| = |\mathrm{Re}(z)|.$$

The result for the complex part of z is similar.

7. Let $z = a + bi$. Then

$$z + \overline{z} = (a + bi) + (a - bi) = 2a = 2\mathrm{Re}(z),$$

and

$$z - \overline{z} = (a + bi) - (a - bi) = 2bi = 2i\,\mathrm{Im}(z).$$

8.

$$\begin{aligned}
|z+w|^2 &= (z+w)(\overline{z+w}) && \text{(definition of modulus)}\\
&= (z+w)(\overline{z}+\overline{w}) && \text{(part 3)}\\
&= z\,\overline{z} + z\,\overline{w} + \overline{z}\,w + \overline{w}\,w\\
&= |z|^2 + z\,\overline{w} + \overline{z}\,w + |w|^2\\
&= |z|^2 + z\,\overline{w} + \overline{z}\,\overline{\overline{w}} + |w|^2 && \text{(part 1)}\\
&= |z|^2 + z\,\overline{w} + \overline{z\overline{w}} + |w|^2 && \text{(part 4)}\\
&= |z|^2 + 2\mathrm{Re}(z\,\overline{w}) + |w|^2 && \text{(part 6)}\\
&\le |z|^2 + 2|\mathrm{Re}(z\,\overline{w})| + |w|^2\\
&\le |z|^2 + 2|z\overline{w}| + |w|^2 && \text{(part 5)}\\
&= |z|^2 + 2|z||\overline{w}| + |w|^2 && \text{(part 4)}\\
&= |z|^2 + 2|z||w| + |w|^2 && \text{(part 2)}\\
&= (|z| + |w|)^2 .
\end{aligned}$$

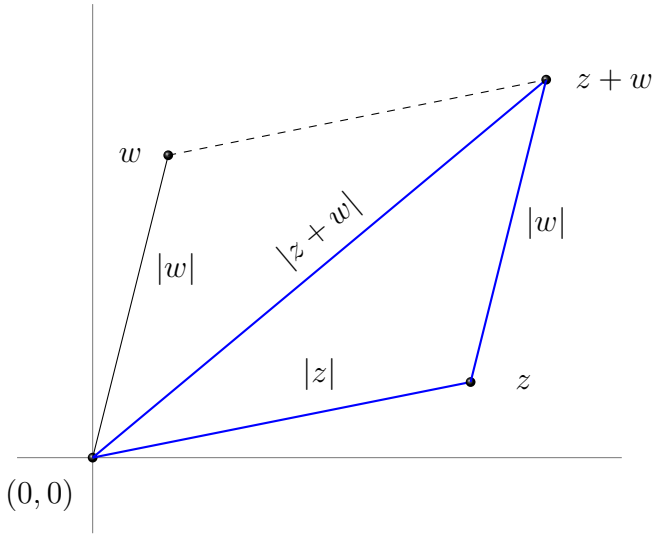
Now take square roots to get the result.

9.

$$\left|\frac{z}{|z|}\right| = \left|z \cdot \frac{1}{|z|}\right| = |z|\left|\frac{1}{|z|}\right| = |z|\frac{1}{|z|} = 1.$$

□

The picture below may help to give an intuitive understanding of the triangle inequality.



Solutions to practice problems.

1. We have
$$z = 4 + 2i + (5 + i)i = 4 + 2i + 5i - 1 = 3 + 7i.$$
Hence, $\mathrm{Re}(z) = 3$ and $\mathrm{Im}(z) = 7$.

2. If $z = 4 + 5i$, then
$$\begin{aligned} \overline{z} &= 4 - 5i \\ |z| &= \sqrt{4^2 + 5^2} = \sqrt{41} \\ z + \overline{z} &= 4 + 5i + 4 - 5i = 8 = 2\,\mathrm{Re}(z) \\ z - \overline{z} &= 4 + 5i - (4 - 5i) = 10i = 2i\,\mathrm{Im}(z). \end{aligned}$$

Week 6, Friday: Complex numbers III

Complex numbers III

(Supplemental reading: Section 3.4 in Swanson.)

polar form for a complex numbers

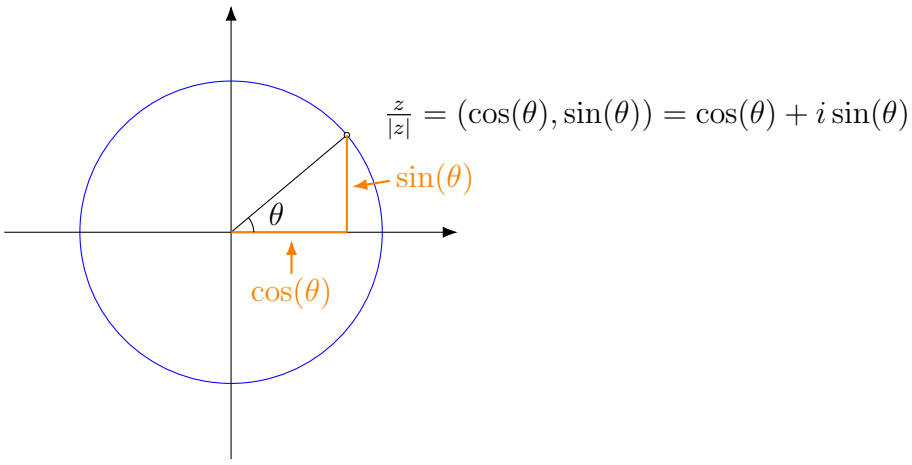
Let z be a nonzero complex number. Then its length, $|z|$, is a nonzero real number, and we can write

$$z = |z| \cdot \frac{z}{|z|}.$$

The number $z/|z|$ has unit length:

$$\left|\frac{z}{|z|}\right| = \left|z \cdot \frac{1}{|z|}\right| = |z| \cdot \left|\frac{1}{|z|}\right| = |z| \cdot \frac{1}{|z|} = 1.$$

So $z/|z|$ is a point in the plane with unit length, i.e., it's a point on the unit circle:



Therefore, we can write

$$\frac{z}{|z|} = (\cos(\theta), \sin(\theta)) = \cos(\theta) + i\sin(\theta)$$

for some angle θ.

Definition. The *angle* or *argument* of $z \in \mathbb{C} \setminus \{0\}$ is

$$\arg(z) := \theta$$

where θ is the angle shown above. The *polar form* for z is

$$z = |z|(\cos(\theta) + i\sin(\theta)).$$

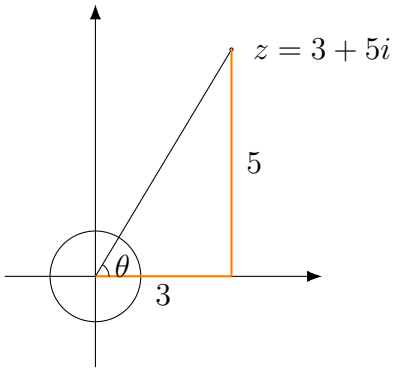
The polar form describes z in terms of its length $|z|$ and angle θ.

Examples.

1. Let $z = 7 + 7i$. Then
$$|z| = \sqrt{7^2 + 7^2} = 7\sqrt{2}$$
and
$$\frac{z}{|z|} = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i.$$
Plotting this complex number, it's clear angle is $45° = \pi/4$. So the polar form for z is
$$z = 7\sqrt{2}\left(\cos\left(\frac{\pi}{4}\right) + \sin\left(\frac{\pi}{4}\right)i\right).$$

2. Let $z = \sqrt{3} + i$. The length of z is
$$|z| = \sqrt{\sqrt{3}^2 + 1^2} = \sqrt{3+1} = 2.$$
Hence,
$$\frac{z}{|z|} = \frac{\sqrt{3}}{2} + \frac{1}{2}i.$$
You should notice that $(\sqrt{3}/2, 1/2)$ is the point on the unit circle with angle $30° = \pi/6$. Hence, $\arg(z) = \pi/6$, and the polar form for z is
$$z = 2\left(\cos\left(\frac{\pi}{6}\right) + \sin\left(\frac{\pi}{6}\right)i\right).$$

3. Let $z = 3 + 5i$:



The length of z is

$$|z| = \sqrt{3^2 + 5^2} = \sqrt{34},$$

the angle of z is

$$\arg(z) = \tan^{-1}\left(\frac{5}{3}\right) \approx 59^\circ \approx 1.03\,\text{rad}.$$

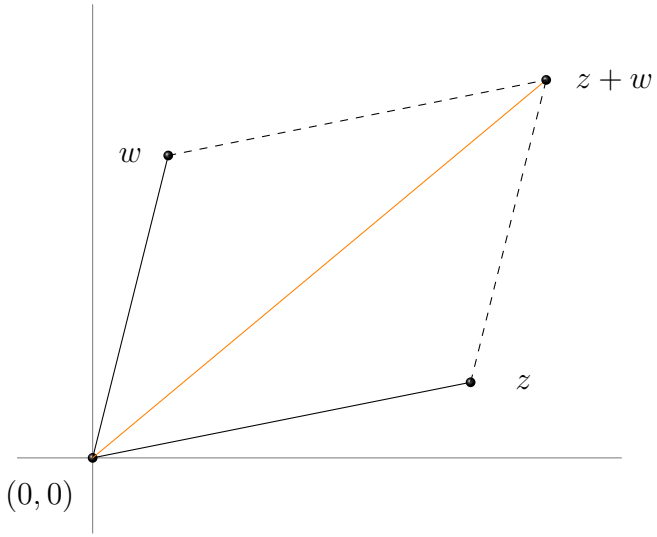
Therefore, the polar form for z is

$$z = \sqrt{34}(\cos(\theta) + \sin(\theta)i)$$

where $\theta = \tan^{-1}(5/3)$.

GEOMETRY OF ADDITION AND MULTIPLICATION IN $\mathbb{C}$

The geometry of addition in $\mathbb{C}$ is just the geometry of vector addition. It is given by the "parallelogram rule":



To see the geometry of multiplication, the secret is to use polar form. Given two nonzero complex numbers $z, w \in \mathbb{C}$, write each in polar form

$$\begin{aligned} z &= |z|\,(\cos(\theta) + \sin(\theta)i) \\ w &= |w|\,(\cos(\psi) + \sin(\psi)i)\,. \end{aligned}$$

Their product is then

$$\begin{aligned} zw &= |z|\,(\cos(\theta) + \sin(\theta)i)\,|w|\,(\cos(\psi) + \sin(\psi)i) \\ &= |z||w|\,(\cos(\theta) + \sin(\theta)i)\,(\cos(\psi) + \sin(\psi)i) \\ &= |zw|\,(\cos(\theta) + \sin(\theta)i)\,(\cos(\psi) + \sin(\psi)i) \\ &= |zw|\,(\cos(\theta)\cos(\psi) - \sin(\theta)\sin(\psi) + (\cos(\theta)\sin(\psi) + \cos(\psi)\sin(\theta))\,i) \end{aligned}$$

$$= |zw| \left(\cos(\theta + \psi) + \sin(\theta + \psi)i\right).$$

The last step uses the angle sum formulas for cosine and sine. The last line gives the polar form for zw and reveals the geometry: when multiplying two complex numbers, **angles add and lengths multiply**.

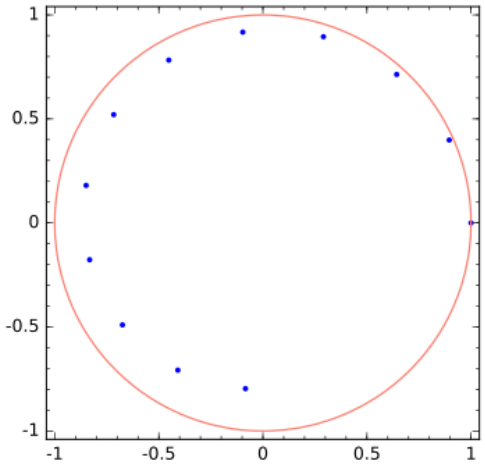
Example. Consider the complex number

$$\alpha = 0.95\left(\cos\left(\frac{2\pi}{15}\right) + \sin\left(\frac{2\pi}{15}i\right)\right).$$

What will the sequence of complex numbers

$$1 = \alpha^0,\ \alpha,\ \alpha^2,\ \alpha^3,\ \ldots$$

look like in the plane? The length of α is 0.95, and so the length of $\alpha^n = (0.95)^n$. Since 0.95 is less than 1, its powers will get smaller and smaller, which means that this sequence of points will get closer and closer to the origin. The argument of α is an angle that is 1/15-th of the circle. To get to each successive number in the sequence, we multiply by α, which adds that angle to each succesive number, and scale by 0.95. So the sequence spirals around the origin, getting closed to the origin with each step.



The Sage code on cocalc.com that produced this image is on the next page.

Code on cocalc.com:

```
s = 15
a = 0.98*exp(I*2*pi/s)
p = list_plot([a^n for n in range(12)])
q = parametric_plot\left( cos(x),sin(x)),(x,0,2*pi),color='salmon')
(p+q).show(aspect_ratio=1,axes=false,frame=true)
```

Week 7, Monday: Topology

Topology

(Supplemental reading: Section 3.5 in Swanson.)

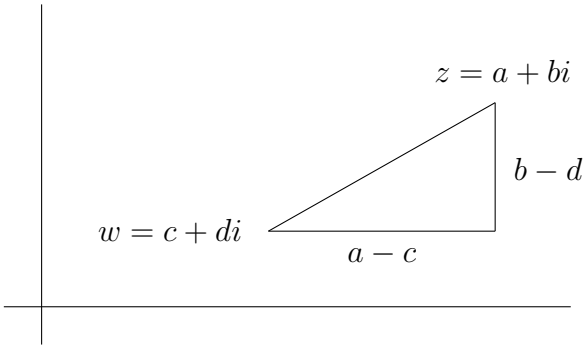
For the following, let $F = \mathbb{R}$ or $\mathbb{C}$.

Definition. The *distance* between $z, w \in F$ is

$$d(z, w) := |z - w|.$$

Example. In $\mathbb{C}$, with $z = a + bi$ and $w = c + di$,

$$d(z, w) = \sqrt{(a - c)^2 + (b - d)^2}.$$



Definition. The *open ball centered at* $z \in F$ *of radius* $r \in \mathbb{R} > 0$ is the subset

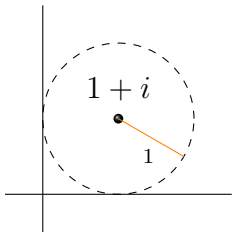
$$B(z; r) := \{w \in F : |w - z| < r\}.$$

Example.

1. If $F = \mathbb{R}$, then what we have just called a ball is an open interval. For example, in $\mathbb{R}$, we have $B(3; 1) = (2, 4)$:

$$B(3; 1) = (2, 4): \qquad \text{(}\quad\text{)} \qquad .$$
$$2 \quad 3 \quad 4$$

2. Here is a picture of the ball of radius 1 in $\mathbb{C}$ centered at $1+i$:

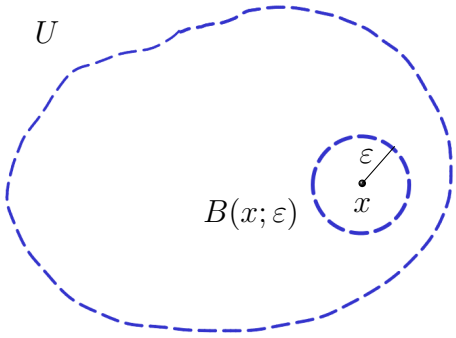
$B(1+i;1)$: 

.

Definition. A subset $U \subseteq F$ is *open* if it contains an open ball about each of its points. This means that for all $u \in U$, there exists $\varepsilon > 0$ such that

$$B(u;\varepsilon) \subseteq U.$$

A key point here is that the open ball must be entirely contained inside U:



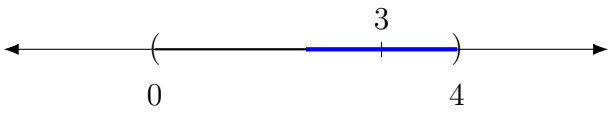
In the above picture, $U \subset \mathbb{C}$ is the set of all points inside the outer dashed line. Inside of U, we show a typical point x and an example of an open ball about x completely contained inside of U. As we test points nearer the border of U, the choice for ε will need to become smaller.

Example.

1. The empty set $\emptyset \subset F$ is open.

2. The set F, itself, is open (whether $F = \mathbb{R}$ or $\mathbb{C}$). Proof: Given $x \in F$, the open ball $B(x;1)$ is contained F.

3. An open interval $(a,b) \subseteq \mathbb{R}$ is open. Given $c \in (a,b)$, let ε be either $c-a$ or $b-c$, whichever is smallest. Then the open interval centered at c and with radius ε is contained in (a,b). For instance, take the interval $(0,4)$, and let $c = 3$. Then $(0,4)$ contains the open ball of radius 1 about c, i.e., it contains the interval $(2,4)$. This open ball is shown in blue below:

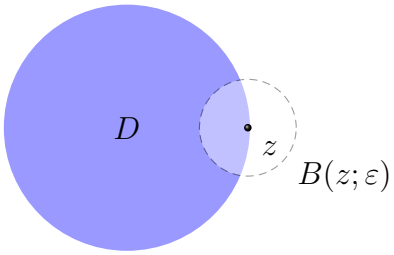


4. A union of open intervals in $\mathbb{R}$ is an open set. In fact, every subset of $\mathbb{R}$ is a union of open intervals. We prove this and the analogous result for $\mathbb{C}$ in a proposition below.

5. In $\mathbb{C} = \mathbb{R}^2$ the Cartesian product $(a, b) \times (c, d)$ is open. (This is a rectangle in $\mathbb{R}^2$ not containing its boundary.)

6. The closed interval $[1, 2)$ is *not* open. The problem is the endpoint 1. We have $1 \in [1, 2]$ but there is no open ball (interval) centered at 1 and contained in $[1, 2)$. Any open interval about 1 will contain points less that 1 and hence not in the set $[1, 2)$.

7. The closed disc $D := \{z \in \mathbb{C} : |z| \leq 1\}$ is not open. The problem is each of the points with modulus 1 sitting on the boundary of D. Any open ball about one of these points will contain points that are not in D, i.e., it will not be completely contained in D:



Proposition. Every open set in F is a union of open balls.

Proof. Let $U \subseteq F$ be open. Since U is open, for each $u \in U$, there exists $\varepsilon_u > 0$ such that

$$B(u; \varepsilon_u) \subseteq U.$$

We claim

$$\bigcup_{u \in U} B(u; \varepsilon_u) = U.$$

To see this, first let $v \in \cup_{u \in U} B(u; \varepsilon_u)$. Then $v \in B(u; \varepsilon_u)$ for some $u \in U$ (in fact, $v \in B(v; \varepsilon_v)$), and since $B(u; \varepsilon_u) \subseteq U$, we have $v \in U$. To see the reverse inclusion let $v \in U$. Then $v \in B(v; ve_v)$, and hence, v is in the union $\cup_{u \in U} B(u; \varepsilon_u)$. □

Proposition. Every nonempty open subset of F has infinitely many elements.

Proof. Let U be a nonempty open subset of F. Since U is nonempty, it contains some point u. Since U is open, it must contain an open ball $B(u;\varepsilon)$ about U. But every open ball contains infinitely many points. For instance $B(u;\varepsilon)$ contains the points $\{u+\varepsilon/n : n = 2,3,4,\dots\}$. □

Corollary. No nonempty finite set of points is open. In particular, for each $z \in F$, the set $\{z\}$ is not open.

Definition. The set of open sets of F forms a *topology* on F. This means that

1. $\emptyset$ and F are open.
2. An arbitrary union of open sets is open.
3. A finite intersection of open sets is open.

Regarding part 3, note that an infinite intersection of open sets might not be open. For example,

$$\bigcap_{n=1}^{\infty} B(z;1/n) = \{z\}$$

which is not open (since it is finite). For a more concrete example, note that

$$\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}\,.$$

Definition. A subset $C \subseteq F$ is *closed* if its complement $F \setminus C$ is open.

Example.

1. F is closed since $F \setminus F = \emptyset$, and $\emptyset$ is open.
2. Similarly $\emptyset$ is closed since $F \setminus \emptyset = F$, and F is open. (Thus, we have seen that F and $\emptyset$ are both open and closed.[1])
3. If $B \subseteq F$ is an open ball, the $F \setminus B$ is closed. For instance, if $F = \mathbb{R}$ and $B = (0;1)$, then
$$\mathbb{R} \setminus (-1,1) = (-\infty,-1] \cup [1,\infty)$$
is closed.

[1] If it seems paradoxical to you that there are sets, like the empty set, that are both open and closed, you are not alone (cf. Hitler learns topology where "null set"= "empty set", "neighborhood"="open ball").

4. A closed interval $[a, b]$ in $\mathbb{R}$ is closed. That's because its complement

$$\mathbb{R} \setminus [a, b] = (-\infty, a) \cup (b, \infty)$$

is open.

5. If $z \in F$, then the set $\{z\}$ is closed. (Exercise.)

Definition. The *closed ball (disc) of radius r centered at $z \in F$* is

$$D(z; r) := \{w \in F : |w - z| \leq r\}.$$

Example. In $\mathbb{R}$, closed discs are the same as closed intervals $[a, b]$. In $\mathbb{C}$ taking an open ball and adding its boundary gives a closed ball.

The following records an argument explained in the accompanying video lecture:

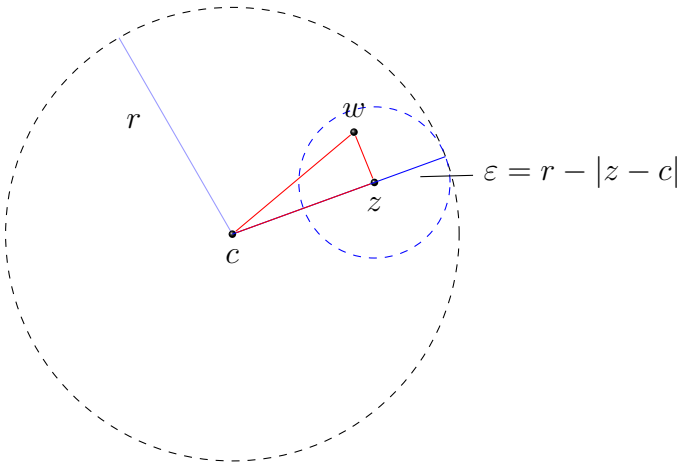
Proposition. An open ball in F is open.

Proof. Let $B(c; r)$ be an open ball in F, and let $z \in B(c; r)$. We must show there is an open ball about z that is completely contained in $B(c; r)$. Let $\varepsilon := r - |z - c|$, and note that $\varepsilon > 0$ since $|z - c| < r$.
We claim that $B(z; \varepsilon) \subseteq B(c; r)$. To see this, $w \in B(z; \varepsilon)$. Then

$$\begin{aligned} d(c, w) = |w - c| &= |(w - z) + (z - c)| \\ &\leq |w - z| + |z - c| \\ &< \varepsilon + |z - c| \\ &= r. \end{aligned}$$

□

Here is a picture motivating the above proof:



Week 7, Wednesday: Sequences I

Sequences I

(Supplemental reading: Start reading Chapter 8 in Swanson.)

A *sequence of complex numbers* (a *complex sequence*) is a function

$$\mathbb{N}_{>0} \to \mathbb{C}.$$

If s is such a function, instead of $s(n)$, we usually write s_n. Essentially, the function s is just an unending ordered sequence of numbers:

$$s_1, s_2, s_3, \ldots$$

The sequence s can be notated in various ways, including

$$\{s_n\}_{n>0}\,,\ \{s_n\}_{n=1}^{\infty}\,,\ \{s_n\}_n\,,\ \{s_n\}\,,$$

among others. A *real sequence* is a special case of a complex sequence in which the image of s is in $\mathbb{R} \subset \mathbb{C}$.

Example. Some examples of sequences:

1.
$$\{1\}_{n=1}^{\infty} = 1, 1, 1, \ldots$$

2. The *first term* in the following sequence is 4:
$$\{n\}_{n\geq 4} = 4, 5, 6, \ldots$$

3.
$$\{(-1)^n\}_{n\geq 0} = 1, -1, 1, -1, \ldots$$

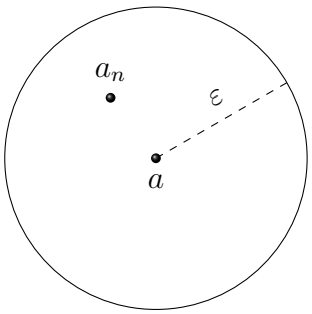
4.
$$\left\{\frac{1}{n^2}\right\}_{n=1}^{\infty} = 1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \ldots$$

Definition. Let $\{a_n\}$ be a sequence of complex numbers, and let $a \in \mathbb{C}$. Then the *limit of the sequence* $\{a_n\}$ *as* n *goes to infinity is* a, denoted $\lim_{n\to\infty} a_n = a$, if for all $\varepsilon > 0$, there exists $N \in \mathbb{R}$ such that $n > N$ implies $|a - a_n| < \varepsilon$. If $\lim_{n\to\infty} a_n = a$ for some $a \in \mathbb{C}$, we say $\{a_n\}$ is *convergent*, and if there is no such a, we say the sequence *diverges*.

Here is a shorthand for writing the definition: $\lim_{n\to\infty} a_n = a$ if $\forall \varepsilon > 0$, $\exists N \in \mathbb{R}$ such that $n > N \Rightarrow |a - a_n| < \varepsilon$. The symbols $\forall$ and $\exists$ are called *quantifiers*. They stand for "for all" and "there exists", respectively.

It is notoriously difficult to fully appreciate all that is packed into the definition of the limit of a sequence. So the reader is encouraged to take their time and be patient! Here are some pointers to get started:

1. Most importantly, the condition $|a - a_n| < \varepsilon$ means that a_n *is within a distance of* ε *of* a, i.e., it's in the ball of radius ε centered at a:



   If we are dealing with a real sequence, then $|a - a_n| < \varepsilon$ means that a_n is in the interval $(a - \varepsilon, a + \varepsilon) \subset \mathbb{R}$. Note that when a is real, this interval is the intersection of the above ball with the real number line in $\mathbb{C}$.

2. The rough idea of the limit is that as n get large, a_n gets close to a. By itself, this characterization is too vague. What is meant by "gets close to"? Also, is it OK if some but not all a_n get close to a?

3. The number ε is a challenge: "Can you make the distance between a_n and a less than ε? The number N is the response: "Yes, if you go out further than N steps in the sequence, then all of the numbers in the sequence past that point are within a distance ε of a."

4. The number N is a function of ε. If $\varepsilon > 0$ is made smaller, then N usually needs to be made larger—you need to go out further in the sequence to get closer to the limit. So it might be better to write $N(\varepsilon)$ or N_ε instead of just N to highlight this dependence.

5. This might be very helpful in understanding our definition: $\lim_{n\to\infty} a_n = a$ is equivalent to saying that for every $\varepsilon > 0$, all but a finite number of the a_n are inside the ball of radius ε centered at a.

The following is a template for a limit proof that works directly from the definition of the limit:

Proof Template. $\lim_{n\to\infty} a_n = a$.

Proof. Given $\varepsilon > 0$, let $N =$ blah. Then if $n > N$, we have

$$\begin{aligned} |a - a_n| &= \text{ blah} \\ &= \\ &\leq \\ &= \\ &< \\ &= \varepsilon. \end{aligned}$$

□

For the proof to be valid, at least one strict inequality,"<", is required.

Example. Let $a \in \mathbb{C}$, and consider the constant sequence $\{a\}_n = a, a, a, \ldots$. Then $\lim_{n\to\infty} a = a$.

Proof. We are considering the sequence $\{a_n\}$ such that $a_n = a$ for all n. Given $\varepsilon > 0$, let $N = 0$. Then $n > N$ implies

$$|a - a_n| = |a - a| = 0 < \varepsilon.$$

Thus, $\lim_{n\to\infty} a = a$. □

Note that in the above proof, we could have chosen any number $N \in \mathbb{R}$. The distance of *every* term in the sequence is a distance of 0 from a, and $0 < \varepsilon$ by choice of ε.

Example. $\lim\limits_{n\to\infty} \dfrac{1}{n} = 0$.

Proof. Given $\varepsilon > 0$, let $N = 1/\varepsilon$. Then $n > N$ implies

$$\begin{aligned} \left|0 - \frac{1}{n}\right| &= \left|\frac{1}{n}\right| \\ &= \frac{1}{n} \qquad\qquad (\text{since } n > 0) \end{aligned}$$

$$\begin{aligned} &< \frac{1}{N} && \text{(since } n > N\text{)} \\ &= \varepsilon && \text{(since } N = 1/\varepsilon\text{).} \end{aligned}$$

□

The above proof is typical in that the value for N is unmotivated. The reader can verify each step of the proof, but may be mystified by the choice of N. That's because one usually constructs a limit proof by starting at the end of the proof (on scratch paper): we want $|a - a_n| < \varepsilon$, and then work backwards to find N. For instance, in this last proof, in the end, we want

$$\left|0 - \frac{1}{n}\right| < \varepsilon.$$

This statement is equivalent to

$$\frac{1}{n} < \varepsilon,$$

which is equivalent to

$$\frac{1}{\varepsilon} < n.$$

Thus, if we take $N = \frac{1}{\varepsilon}$ and assume $n > N$, we are OK:

$$\frac{1}{\varepsilon} = N < n.$$

All of our steps are reversible, so things are going to work out.

Example. Let's apply that same reasoning to construct a proof that

$$\lim_{n\to\infty} (1 + 1/\sqrt{n}) = 1.$$

Here $a_n = 1 + 1/\sqrt{n}$ and $a = 1$. In the end, we will want $|a - a_n| < \varepsilon$, i.e.,

$$|1 - (1 + 1/\sqrt{n})| < \varepsilon.$$

On scratch paper we work out

$$\begin{aligned} |1 - (1 + 1/\sqrt{n})| < \varepsilon &\Leftrightarrow |-1/\sqrt{n}| < \varepsilon \\ &\Leftrightarrow \frac{1}{\sqrt{n}} < \varepsilon \\ &\Leftrightarrow \frac{1}{\varepsilon} < \sqrt{n} \\ &\Leftrightarrow \frac{1}{\varepsilon^2} < n. \end{aligned}$$

Thus, we can take $N = 1/\varepsilon^2$. Now we are ready to write the formal proof:

Claim. $\lim_{n\to\infty}(1+1/\sqrt{n}) = 1$

Proof. Given $\varepsilon > 0$, let $N = 1/\varepsilon^2$ and suppose that $n > N$. It follows that

$$\begin{aligned}\left|1 - \left(1 + \frac{1}{\sqrt{n}}\right)\right| &= \frac{1}{\sqrt{n}} \\ &< \frac{1}{\sqrt{N}} && \text{(since } n > N\text{)} \\ &= \varepsilon && \text{(since } N = 1/\varepsilon^2\text{).}\end{aligned}$$

□

(Note that $n > N \Rightarrow \sqrt{n} > \sqrt{N} \Rightarrow \frac{1}{\sqrt{N}} > \frac{1}{\sqrt{n}}$.)

Week 7, Friday: Sequences II

Sequences II

(Supplemental reading: Sections 8.1 and 8.2 in Swanson.)

Recall the definition of the limit of a sequence of complex numbers: we say $\lim_{n\to\infty} a_n = a$ if for all $\varepsilon > 0$ there exists $N \in \mathbb{R}$ such that $n > N$ implies $|a - a_n| < \varepsilon$.

To prove a sequence $\{a_n\}$ doesn't have a limit, we need to show that for all $a \in \mathbb{C}$, the limit of $\{a_n\}$ is not a. To make sense of what this means in terms of the definition of the limit, one pointer is that the negation of "for all" is "there exists", and vice versa: if it is not true that"for $\varepsilon > 0$, there exists $N\ldots$", then there exists $\varepsilon > 0$ such that for all $N\ldots$ The following proposition illustrates this principle.

Proposition. The sequence $\{(-1)^n\} = -1, 1, -1, 1, \ldots$ diverges.

Proof. Let a be any complex number. We claim that $\{(-1)^n\}$ does not converge to a. To see this, let $\varepsilon = 1$, and let N be any real number. There is some even number $n > N$, and for this n,

$$|a - (-1)^n| = |a - 1|.$$

Similarly, these is some odd number $n > N$, and for this n,

$$|a - (-1)^n| = |a + 1|.$$

Next, using the triangle inequality, we see

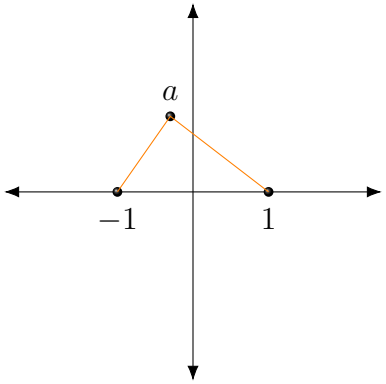
$$2 = |(a - 1) - (a + 1)| \leq |a - 1| + |-(a + 1)| = |a - 1| + |a + 1|.$$

Since $2 \leq |a-1|+|a+1|$, at least one of $|a-1|$ or $|a+1|$ is greater than or equal to 1. So it follows that there exists some $n > N$ such that

$$|a - (-1)^n| \not< \varepsilon = 1.$$

□

The motivation for the above proof is the following picture:



There is no way to find $a \in \mathbb{C}$ that is both close to -1 and close to 1. In the proof, we applied the triangle inequality to the triangle in the picture to conclude that the distance from a to at least one of 1 or -1 must be at least 1. That leads to taking $\varepsilon = 1$. (The argument works by taking any ε such that $0 < \varepsilon \leq 1$.)

We next prove a result that will be of great use later on. Start with a number $\alpha \in \mathbb{C}$, and consider the sequence $\{\alpha^n\} = \alpha, \alpha^2, \alpha^3, \ldots$ What is the behavior of this sequence? Does it converge? The answer depends on the initial choice of α. The key to unlocking the behavior of the sequence is to consider the polar form of α. Say $\alpha = |\alpha|(\cos(\theta) + i\sin(\theta))$. Then $\alpha^n = |\alpha|^n(\cos(n\theta) + i\sin(n\theta))$. So the length of α^n is $|\alpha|^n$ and the angle (or *argument*) of α^n is n times the angle of α. So it seems clear that if $|\alpha| > 1$, the sequence "diverges to ∞", spinning as it goes (unless the angle of α is 0). Similarly, if $|\alpha| < 1$, the sequence is sucked into the origin. What if $|\alpha| = 1$? Then α is a point on the unit circle, and unless $\arg(\alpha) = 0$, the sequence will perpetually spin around the circle, never converging. See Figure 20.1.

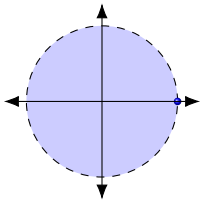


Figure 20.1: The sequence $\{\alpha^n\}$ converges to 1 if $\alpha = 1$, to 0 if $|\alpha| < 0$, and diverges, otherwise. The α for which the sequence converges appear in blue.

Proposition. Let $\alpha \in \mathbb{C}$ and consider the sequence $\{\alpha_n\}$. There are three cases:

1. $\lim_{n\to\infty} \alpha^n = 1$ if $\alpha = 1$;

2. $\lim_{n\to\infty} \alpha^n = 0$ if $|\alpha| < 1$;

3. $\{\alpha^n\}$ diverges, otherwise, i.e., if $\alpha \neq 1$ and $|\alpha| \geq 1$.

Proof. For case 1, where $\alpha = 1$, we get the constant sequence $1, 1, 1, \ldots$ Last time we saw that the limit of a constant sequence is just the constant (given $\varepsilon > 0$, we can take N to be any real number).

For case 2, suppose that $|\alpha| < 1$. If $\alpha = 0$, then α^n is the constant sequence $0, 0, 0, \ldots$, with limit 0. Otherwise, given $\varepsilon > 0$, we need to find $N \in \mathbb{R}$ such that if $n > N$, then

$$|0 - \alpha^n| < \varepsilon.$$

This is equivalent to showing that

$$|\alpha|^n < \varepsilon.$$

To find N, we solve for n in the above equation:

$$|\alpha|^n < \varepsilon \quad \Leftrightarrow \quad \ln(|\alpha|^n) < \ln(\varepsilon) \quad \Leftrightarrow \quad n \ln(|\alpha|) < \ln(\varepsilon) \quad \Leftrightarrow \quad n > \ln(\varepsilon)/\ln(|\alpha|).$$

There are a couple of subtleties in the above calculation. In the first step, we used the fact that taking logs preserves inequalities. This is because the log is an increasing function: $\ln(x) < \ln(y)$ if and only if $x < y$. In the last step of the calculation, the inequality is reversed since $|\alpha| < 1$ means that $\ln(|\alpha|) < 0$. To complete the proof that $\lim_{n\to\infty} \alpha^n = 0$ in this case, let $N := \ln(\varepsilon)/\ln(|\alpha|)$ and suppose that $n > N$. Running the string of implications displayed above in reverse, we see that it follows that $n > N$ implies

$$|0 - \alpha^n| = |\alpha|^n < \varepsilon,$$

as required.

The last case is the toughest. Suppose that $\alpha \neq 1$ and $|\alpha| \geq 1$. We must show that $\{\alpha^n\}$ is divergent. We prove this by contradiction. Suppose that $\lim_{n\to\infty} \alpha^n = \lambda$ for some $\lambda \in \mathbb{C}$. Define

$$\varepsilon := \frac{|\alpha - 1|}{2}.$$

Note that $\varepsilon > 0$ since $\alpha \neq 1$. We'll see below that this particular ε will give us insurmountable problems, forcing the desired contradiction. Since $\lim_{n\to\infty} \alpha^n = \lambda$, there exists $N \in \mathbb{R}$ such that

$$n > N \quad \Longrightarrow \quad |\lambda - \alpha^n| < \varepsilon = \frac{|\alpha - 1|}{2}.$$

Now, if $n > N$, it follows that $n + 1 > N$, too. Hence,

$$|\lambda - \alpha^{n+1}| < \frac{|\alpha - 1|}{2}.$$

Therefore, if $n > N$, it follows that

$$|\alpha^{n+1} - \alpha^n| = |(\lambda - \alpha^n) - (\lambda - \alpha^{n+1})|$$

$$\begin{aligned} &\leq |\lambda - \alpha^n| + |\lambda - \alpha^{n+1}| && \text{(triangle inequality)} \\ &< \frac{|\alpha - 1|}{2} + \frac{|\alpha - 1|}{2} \\ &= |\alpha - 1|, \end{aligned}$$

i.e.,

$$|\alpha^{n+1} - \alpha^n| < |\alpha - 1|$$

for all $n > N$. On the other hand,

$$\begin{aligned} |\alpha^{n+1} - \alpha^n| &= |\alpha|^n |\alpha - 1| \\ &\geq |\alpha - 1| && \text{since } |\alpha| \geq 1. \end{aligned}$$

Thus, our assumption that $\lim_{n\to\infty} \alpha^n = \lambda$ has allowed us to show that there is an N such that $n > N$ implies both

$$|\alpha^{n+1} - \alpha^n| < |\alpha - 1|$$

and

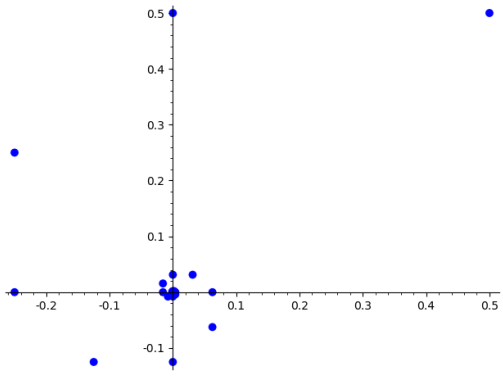
$$|\alpha^{n+1} - \alpha^n| \geq |\alpha - 1|.$$

That's not impossible. So there is no such λ. □

Example. The sequence $\left\{\left(\frac{1+i}{2}\right)^n\right\}$ converges since

$$\left|\frac{1+i}{2}\right| = \frac{1}{2}|i+1| = \frac{1}{2}\sqrt{1^2+1^2} = \frac{\sqrt{2}}{2} < 1.$$

The sequence $\{i^n\}$ does not converge since $|i| = 1$ and $i \neq 1$. A picture of the first sequence appears below:



Week 8, Monday: Sequences II

Sequences II

(Supplemental reading: Sections 8.1 and 8.2 in Swanson.)

Recall the definition of the limit of a sequence of complex numbers: we say $\lim_{n\to\infty} a_n = a$ if for all $\varepsilon > 0$ there exists $N \in \mathbb{R}$ such that $n > N$ implies $|a - a_n| < \varepsilon$.

To prove a sequence $\{a_n\}$ doesn't have a limit, we need to show that for all $a \in \mathbb{C}$, the limit of $\{a_n\}$ is not a. To make sense of what this means in terms of the definition of the limit, one pointer is that the negation of "for all" is "there exists", and vice versa: if it is not true that"for $\varepsilon > 0$, there exists $N\ldots$", then there exists $\varepsilon > 0$ such that for all $N\ldots$ The following proposition illustrates this principle.

Proposition. The sequence $\{(-1)^n\} = -1, 1, -1, 1, \ldots$ diverges.

Proof. Let a be any complex number. We claim that $\{(-1)^n\}$ does not converge to a. To see this, let $\varepsilon = 1$, and let N be any real number. There is some even number $n > N$, and for this n,

$$|a - (-1)^n| = |a - 1|.$$

Similarly, these is some odd number $n > N$, and for this n,

$$|a - (-1)^n| = |a + 1|.$$

Next, using the triangle inequality, we see

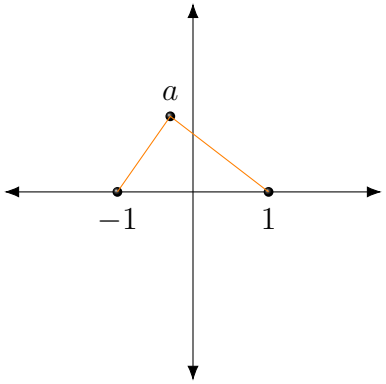
$$2 = |(a - 1) - (a + 1)| \leq |a - 1| + |-(a + 1)| = |a - 1| + |a + 1|.$$

Since $2 \leq |a-1|+|a+1|$, at least one of $|a-1|$ or $|a+1|$ is greater than or equal to 1. So it follows that there exists some $n > N$ such that

$$|a - (-1)^n| \not< \varepsilon.$$

□

The motivation for the above proof is the following picture:



There is no way to find $a \in \mathbb{C}$ that is both close to -1 and close to 1. In the proof, we applied the triangle inequality to the triangle in the picture to conclude that the distance from a to at least one of 1 or -1 must be at least 1. That leads to taking $\varepsilon = 1$. (The argument works by taking any ε such that $0 < \varepsilon \leq 1$.)

We next prove a result that will be of great use later on. Start with a number $\alpha \in \mathbb{C}$, and consider the sequence $\{\alpha^n\} = \alpha, \alpha^2, \alpha^3, \ldots$. What is the behavior of this sequence? Does it converge? The answer, of course, depends on the initial choice of α. The key to unlocking the behavior of the sequence is to consider the polar form of α. Say $\alpha = |\alpha|(\cos(\theta) + i\sin(\theta))$. Then $\alpha^n = |\alpha|^n(\cos(n\theta) + i\sin(n\theta))$. So the length of α^n is $|\alpha|^n$ and the angle (or *argument*) of α^n is n times the angle of α. So it seems clear that if $|\alpha| > 1$, the sequence "diverges to ∞", spinning as it goes (unless the angle of α is 0). Similarly, of $|\alpha|$, the sequence is sucked into the origin. What if $|\alpha| = 1$? Then α is a point on the unit circle, and unless $\arg(\alpha) = 0$, the sequence will perpetually spin around the circle, never converging. See Figure 21.1.

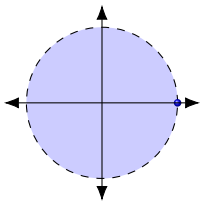


Figure 21.1: If α is one of the blue points, above, then sequence $\{\alpha^n\}$ converges. It converges to 1 if $\alpha = 1$, to 0 if $|\alpha| < 0$, and diverges, otherwise.

Proposition. Let $\alpha \in \mathbb{C}$ and consider the sequence $\{\alpha_n\}$. There are three cases:

1. $\lim_{n\to\infty} \alpha^n = 1$ if $\alpha = 1$;
2. $\lim_{n\to\infty} \alpha^n = 0$ if $|\alpha| < 1$;
3. $\{\alpha^n\}$ diverges, otherwise, i.e., if $\alpha \neq 1$ and $|\alpha| \geq 1$.

Proof. For case 1, where $\alpha = 1$, we get the constant sequence $1, 1, 1, \ldots$ Last time we saw that the limit of a constant sequence is just the constant (given $\varepsilon > 0$, we can take N to be any real number).

For case 2, suppose that $|\alpha| < 1$. If $\alpha = 0$, then α^n is the constant sequence $0, 0, 0, \ldots$, with limit 0. Otherwise, given $\varepsilon > 0$, we need to find $N \in \mathbb{R}$ such that if $n > N$, then

$$|0 - \alpha^n| < \varepsilon.$$

This is equivalent to showing that

$$|\alpha|^n < \varepsilon.$$

To find N, we solve for n in the above equation:

$$|\alpha|^n < \varepsilon \quad \Leftrightarrow \quad \ln(|\alpha|^n) < \ln(\varepsilon) \quad \Leftrightarrow \quad n \ln(|\alpha|) < \ln(\varepsilon) \quad \Leftrightarrow \quad n > \ln(\varepsilon)/\ln(|\alpha|).$$

There are a couple of subtleties in the above calculation. In the first step, we used the fact that taking logs preserves inequalities. This is because the log is an increasing function: $\ln(x) < \ln(y)$ if and only if $x < y$. In the last step of the calculation, the inequality is reversed since $|\alpha| < 1$ means that $\ln(|\alpha|) < 0$. To complete the proof that $\lim_{n\to\infty} \alpha^n = 0$ in this case, let $N := \ln(\varepsilon)/\ln(|\alpha|)$ and suppose that $n > N$. Running the string of implications displayed above in reverse, we see that it follows that $n > N$ implies

$$|0 - \alpha^n| = |\alpha|^n < \varepsilon,$$

as required.

The last case is the toughest. Suppose that $\alpha \neq 1$ and $|\alpha| \geq 1$. We must show that $\{\alpha^n\}$ is divergent. We prove this by contradiction. Suppose that $\lim_{n\to\infty} \alpha^n = \lambda$ for some $\lambda \in \mathbb{C}$. Define

$$\varepsilon := \frac{|\alpha - 1|}{2}.$$

Note that $\varepsilon > 0$ since $\alpha \neq 1$. We'll see below that this particular ε will give us insurmountable problems. Since $\lim_{n\to\infty} \alpha^n = \lambda$, there exists $N \in \mathbb{R}$ such that

$$n > N \quad \Longrightarrow \quad |\lambda - \alpha^n| < \varepsilon = \frac{|\alpha - 1|}{2}.$$

Now, if $n > N$, it follows that $n + 1 > N$, too. Hence,

$$|\lambda - \alpha^{n+1}| < \frac{|\alpha - 1|}{2}.$$

Therefore, if $n > N$, it follows that

$$\begin{aligned} |\alpha^{n+1} - \alpha^n| &= |(\lambda - \alpha^n) - (\lambda - \alpha^{n+1})| \\ &\leq |\lambda - \alpha^n| + |\lambda - \alpha^{n+1}| \qquad \text{(triangle inequality)} \end{aligned}$$

$$\begin{aligned} &< \frac{|\alpha - 1|}{2} + \frac{|\alpha - 1|}{2} \\ &= |\alpha - 1|, \end{aligned}$$

i.e.,

$$|\alpha^{n+1} - \alpha^n| < |\alpha - 1|$$

for all $n > N$. On the other hand,

$$\begin{aligned} |\alpha^{n+1} - \alpha^n| &= |\alpha|^n |\alpha - 1| \\ &\geq |\alpha - 1| \qquad\qquad \text{since } |\alpha| \geq 1. \end{aligned}$$

Thus, our assumption that $\lim_{n\to\infty} \alpha^n = \lambda$ has allowed us to show that there is an N such that $n > N$ implies both

$$|\alpha^{n+1} - \alpha^n| < |\alpha - 1|$$

and

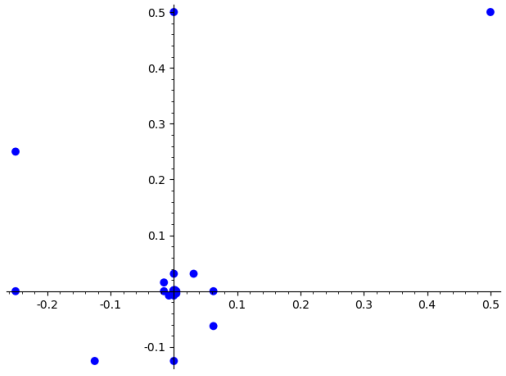
$$|\alpha^{n+1} - \alpha^n| \geq |\alpha - 1|.$$

That's not impossible. So there is no such λ. □

Example. The sequence $\left\{\left(\frac{1+i}{2}\right)^n\right\}$ converges since

$$\left|\frac{1+i}{2}\right| = \frac{1}{2}|i+1| = \frac{1}{2}\sqrt{1^2+1^2} = \frac{\sqrt{2}}{2} < 1.$$

The sequence $\{i^n\}$ does not converge since $|i| = 1$ and $i \neq 1$. A picture of the first sequence appears below:



Week 9, Monday: New-from-old limit theorem

Math 112 lecture for Monday, Week 9

New-from-old limit theorem

(Supplemental reading: Theorem 8.4.3 in Swanson.)

In general, it is difficult to give limit proofs using ε-N arguments. We give an example of this difficulty, and then develop a technique for a much easier, high-level approach.

Example. $\lim_{n\to\infty} \dfrac{n^3+2n}{5n^3+2} = \dfrac{1}{5}$.

Proof. Given $\varepsilon > 0$, let $N = 2/\varepsilon$, and suppose $n > N$. Then

$$\begin{aligned}
\left|\frac{n^3+2n}{5n^3+2} - \frac{1}{5}\right| &= \left|\frac{5(n^3+2n)-(5n^3+2)}{5(5n^3+2)}\right| \\
&= \left|\frac{10n-2}{5(5n^3+2)}\right| \\
&< \left|\frac{10n-2}{5n^3+2}\right| \\
&< \left|\frac{10n-2}{5n^3}\right| \\
&< \left|\frac{10n}{5n^3}\right| \\
&= \left|\frac{2}{n^2}\right| \\
&< \left|\frac{2}{n}\right| \\
&= \frac{2}{n} < \frac{2}{N}
\end{aligned}$$

$$= \frac{2}{2/\varepsilon}$$

$$= \varepsilon.$$

□

Limit theorems. Often, a sequence can constructed from simpler sequences using algebraic operations (addition, substraction, multiplication, and division). In this case, the following theorem is useful:

Theorem. (New-from-old limit theorem.) Suppose that $\lim_{n\to\infty} s_n = s$ and that $\lim_{n\to\infty} t_n = t$. Then,

(1) $\lim_{n\to\infty}(s_n + t_n) = \lim_{n\to\infty} s_n + \lim_{n\to\infty} t_n = s + t.$

(2) $\lim_{n\to\infty}(s_n t_n) = (\lim_{n\to\infty} s_n)(\lim_{n\to\infty} t_n) = st.$

(3) $\lim_{n\to\infty} \frac{s_n}{t_n} = \frac{\lim_{n\to\infty} s_n}{\lim_{n\to\infty} t_n} = \frac{s}{t}$ provided $t \neq 0$.

Corollary. If a and b are real or complex constants, then $\lim_{n\to\infty}(as_n) = a\lim_{n\to\infty} s_n = as$ and, more generally, $\lim_{n\to\infty}(as_n + bt_n) = a\lim_{n\to\infty} s_n + b\lim_{n\to\infty} t_n = as + bt$.

We will provide proofs of these results at the end of these notes. (**Note:** Pay particular attention to the proof of part (1) given there. It contains some important, generally useful ideas.) First, though, we will illustrate the use of our theorem.

Building blocks. We have seen that for $c \in \mathbb{C}$, the constant sequence has limit c, i.e., $\lim_{n\to\infty} c = c$. And we have also seen that

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

In the following examples, we will see how complicated sequences can be evaluated using these building blocks and the new-from-old limit theorem (LT). Note the important trick in part 4.

Examples.

1.

$$\lim_{n\to\infty}\left(5 + \frac{1}{n}\right) = \lim_{n\to\infty} 5 + \lim_{n\to\infty} \frac{1}{n} \qquad \text{(LT 1)}$$

$$= 5 + 0 = 5.$$

2.

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n^2} &= \lim_{n\to\infty}\left(\frac{1}{n}\cdot\frac{1}{n}\right)\\ &= \left(\lim_{n\to\infty}\frac{1}{n}\right)\left(\lim_{n\to\infty}\frac{1}{n}\right) && \text{(LT 2)}\\ &= 0\cdot 0 = 0.\end{aligned}$$

3.

$$\begin{aligned}\lim_{n\to\infty}\frac{5}{n^2} &= \lim_{n\to\infty}\left(5\cdot\frac{1}{n}\cdot\frac{1}{n}\right)\\ &= (\lim_{n\to\infty}5)\left(\lim_{n\to\infty}\frac{1}{n}\right)\left(\lim_{n\to\infty}\frac{1}{n}\right) && \text{(LT 2)}\\ &= 5\cdot 0 = 0.\end{aligned}$$

4.

$$\begin{aligned}\lim_{n\to\infty}\frac{n^3+2n}{5n^3+2} &= \lim_{n\to\infty}\frac{\frac{1}{n^3}(n^3+2n)}{\frac{1}{n^3}(5n^3+2)} && \text{(NB: important trick)}\\ &= \lim_{n\to\infty}\frac{1+\frac{2}{n^2}}{5+\frac{2}{n^3}}\\ &= \frac{\lim_{n\to\infty}\left(1+\frac{2}{n^2}\right)}{\lim_{n\to\infty}\left(5+\frac{2}{n^3}\right)} && \text{(LT3)}\\ &= \frac{\lim_{n\to\infty}1+\lim_{n\to\infty}\frac{2}{n^2}}{\lim_{n\to\infty}5+\lim_{n\to\infty}\frac{2}{n^3}} && \text{(LT1)}\\ &= \frac{1+\lim_{n\to\infty}\frac{2}{n^2}}{5+\lim_{n\to\infty}\frac{2}{n^3}} && \text{(constant sequences)}\\ &= \frac{1+0}{5+0} && \text{(as in example 3)}\\ &= \frac{1}{5}.\end{aligned}$$

Proof of new-from-old limit theorem.

For part (1), we need to show that

$$\lim_{n\to\infty}(s_n+t_n) = \lim_{n\to\infty}s_n + \lim_{n\to\infty}t_n = s+t.$$

Let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = s$, there exists N_s such that $n > N_s$ implies

$$|s - s_n| < \frac{\varepsilon}{2}.$$

Similarly, since $\lim_{n\to\infty} t_n = t$, there exists N_t such that $n > N_t$ implies

$$|t - t_n| < \frac{\varepsilon}{2}.$$

Define N to be the maximum of N_s and N_t:

$$N := \max\{N_s, N_t\}.$$

Thus, $N \geq N_s$ and $N \geq N_t$.
Suppose $n > N$. Then, since $n > N \geq N_s$, it follows that

$$|s - s_n| < \frac{\varepsilon}{2}.$$

Similarly, since $n > N \geq N_t$, it follows that

$$|t - t_n| < \frac{\varepsilon}{2}.$$

Hence, if $n > N$,

$$|(s+t) - (s_n + t_n)| = |(s - s_n) + (t - t_n)| \leq |s - s_n| + |t - t_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

That completes the proof of part (1).
For part (2), we show that

$$\lim_{n\to\infty}(s_n t_n) = (\lim_{n\to\infty} s_n)(\lim_{n\to\infty} t_n) = st.$$

Given $\varepsilon > 0$, define

$$\eta = \min\left\{1, \frac{\varepsilon}{1 + |s| + |t|}\right\}.$$

This means η is the minimum of the two displayed quantities. So $\eta \leq 1$ and $\eta \leq \varepsilon/(1 + |s| + |t|)$ (and equality holds in at least one of these). As above, since $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$, we can find a single N such that $n > N$ implies

$$|s - s_n| < \eta \quad \text{and} \quad |t - t_n| < \eta. \tag{22.1}$$

Now for the creative part: we can write

$$s_n t_n - st = (s - s_n)(t - t_n) - s(t - t_n) - t(s - s_n).$$

(Check this by multiplying out the expression on the right.) It follows that

$$|st - s_n t_n| = |s_n t_n - st|$$

$$
\begin{aligned}
&= |(s - s_n)(t - t_n) - s(t - t_n) - t(s - s_n)| \\
&\leq |(s - s_n)(t - t_n)| + |s(t - t_n)| + |t(s - s_n)| && \text{(triangle inequality)} \\
&= |(s - s_n)||(t - t_n)| + |s||(t - t_n)| + |t||(s - s_n)| \\
&< \eta \cdot \eta + |s|\eta + |t|\eta && \text{(Equation (22.1))} \\
&\leq \eta + |s||\eta + |t|\eta && \text{(since } \eta \leq 1\text{)} \\
&= \eta(1 + |s| + |t|) \\
&< \varepsilon && \left(\text{since } \eta \leq \frac{\varepsilon}{1 + |s| + |t|}\right).
\end{aligned}
$$

The proof for part (3) is similar, and we leave it as an exercise for the interested reader. □

Proof of the corollary to the new-from-old limit theorem.

Suppose a and b are constants. Then

$$
\begin{aligned}
\lim_{n\to\infty}(as_n + bt_n) &= \lim_{n\to\infty}(as_n) + \lim_{n\to\infty}(bt_n) && \text{(LT1)} \\
&= \lim_{n\to\infty} a \lim_{n\to\infty} s_n + \lim_{n\to\infty} b \lim_{n\to\infty} t_n && \text{(LT2)} \\
&= a \lim_{n\to\infty} s_n + b \lim_{n\to\infty} t_n && \text{(limit of constant seq.)} \\
&= as + bt.
\end{aligned}
$$

To see that $\lim_{n\to\infty}(as_n) = a \lim_{n\to\infty} s_n$, let $b = 0$, above. □

Week 9, Wednesday: Monotone Convergence Theorem

Math 112 lecture for Wednesday, Week 9

Monotone Convergence Theorem

(Supplemental reading: Section 8.6 Swanson.)

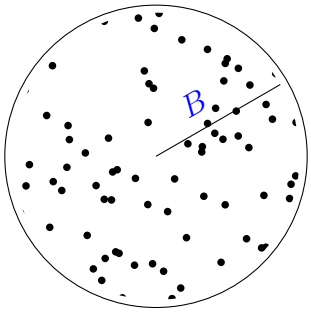
Today, we will prove two results: (i) convergent sequences are bounded (but not conversely), and (ii) the Monotone Convergence Theorem.

Definition. A sequence $\{s_n\}$ of complex numbers is *bounded* if there exists $B \in \mathbb{R}$ such that

$$|s_n| \leq B$$

for all n.

If the s_n are all real numbers, then this definition of boundedness coincides with the one we gave earlier. In general, it means that all of the numbers in the sequence are containing in a closed ball of finite radius centered at the origin:



Theorem. Every convergent sequence is bounded.

Proof. Let $\{s_n\}$ be a convergent sequence, and say $\lim_{n\to\infty} s_n = s$. Apply the definition of the limit with $\varepsilon = 1$ to find an $N \in \mathbb{R}$ such that $n > N$ implies

$$|s - s_n| < 1.$$

By taking N larger, if necessary, we may assume N is an integer greater than or equal to 1. This will be usual later on in the proof. We apply the reverse triangle inequality:

$$1 > |s - s_n| = |s_n - s| \geq |s_n| - |s| \quad \Rightarrow \quad 1 + |s| > |s_n|.$$

Thus, if $n > N$ then $|s_n| < 1 + |s|$. So we have accomplished the hardest part by bounding all but finitely many points in the sequence by the real number $1 + |s|$, but it could be that $s_1, s_2, \ldots, s_N$ are further away from the origin. So let

$$B = \max\{|s_1|, |s_2|, \ldots, |s_N|, 1 + |s|\}.$$

This means that we have set B equal to the maximum number in the given set. It follows that

$$|s_n| \leq B$$

for all n: If $n > N$, then $|s_n| < 1 + |s| \leq B$, and if $n \leq N$, then $|s_n| \leq |s_n| \leq B$. □

Remark. The converse of the above theorem does not hold. For instance, consider the sequence $1, -1, 1, -1, \ldots$ It is bounded but does not converge.

MONOTONE CONVERGENCE THEOREM

Definition. A sequence of real numbers $\{s_n\}$ is *monotone increasing* if $s_n \leq s_{n+1}$ for all n. It is *monotone decreasing* if $s_n \geq s_{n+1}$ for all n.

Note that *monotone increasing* is synonymous with *non-decreasing* since the condition is $s_n \leq s_{n+1}$, not $s_n < s_{n+1}$. An analogous remark holds for monotone decreasing sequences.

Example. The constant sequence $1, 1, 1, \ldots$ is both monotone increasing and monotone decreasing.

Theorem. (Monotone Convergence Theorem, MCT). If $\{s_n\}$ is a monotone increasing sequence that is bounded above or a monotone decreasing sequence that is bounded below, then $\{s_n\}$ converges. (And it converges to $\sup\{s_n\}$ or $\inf\{s_n\}$, respectively.)

Proof. Suppose that $\{s_n\}$ is monotone increasing and bounded above. (The case where $\{s_n\}$ is monotone decreasing and bounded below is similar, or it can be reduced to the increasing case by considering $\{-s_n\}$.) By completeness of the real numbers, the *set* $\{s_n\}$ has a supremum. Say $s = \sup\{s_n\}$. We claim that $\lim_{n\to\infty} s_n = s$. To see this, let $\varepsilon > 0$. Then $s - \varepsilon < s$, i.e., $s - \varepsilon$ is strictly less than the least upper bound of $\{s_n\}$. This means that $s - \varepsilon$ is not an upper bound for that set. Therefore, there exists some s_N such that $s - \varepsilon < s_N$, i.e., $s - s_N < \varepsilon$. However, since the sequence $\{s_n\}$ is a monotone increasing sequence, if $n > N$, it follows that $s_n \geq s_N$,

and hence, $-s_n \leq -s_N$. Adding s to both sides yields $s - s_n \leq s - s_N$. Now, since s is an upper bound for the set, we also have $0 \leq s - s_n$. Putting this all together:

$$0 \leq s - s_n \leq s - s_N < \varepsilon$$

for all $n > N$. It follows that $|s - s_n| < \varepsilon$ for all $n > N$. We have shown that $\lim_{n\to\infty} s_n = s$, as claimed. □

Example. Here is an example of the monotone convergence theorem in use. Define $s_1 = 1$, and for each $n \geq 1$, define $s_{n+1} := \sqrt{1 + s_n}$. The first few terms of the sequence are

$$1,\ \sqrt{1+1},\ \sqrt{1+\sqrt{1+1}}, \ldots$$

To find the limit, we first appeal to the monotone convergence theorem. We show the sequence is monotone increasing, i.e., that $s_n \leq s_{n+1}$ by induction. For the base case, $n = 1$, note that $s_1 = 1 \leq s_2 = \sqrt{2}$. Now suppose that $s_n \leq s_{n+1}$ for some $n \geq 1$. It follows that

$$s_{n+1} = \sqrt{1 + s_n} \leq \sqrt{1 + s_{n+1}} = s_{n+2}.$$

(Here, we have used that $f(x) = \sqrt{x}$ is an increasing function: if $x < y$, then $\sqrt{x} < \sqrt{y}$.) The result now follows for all n by induction.

To use the monotone convergence theorem, we must also verify that the sequence is bounded above. We show it's bounded above by 2 by induction. For the base case, note that $s_1 = 1 \leq 2$. Suppose that $s_n \leq 2$ for some $n \geq 1$. Then

$$s_{n+1} = \sqrt{1 + s_n} \leq \sqrt{1 + 2} = \sqrt{3} \leq 2.$$

The result holds for all n by induction.

The monotone convergence theorem now tells us the sequence has a limit. Say that $\lim_{n\to\infty} s_n = s$. We would like to evaluate s. We will appeal to a result we have not yet shown: if $f(x)$ is a continuous function, then $\lim_{n\to\infty} f(s_n) = f(\lim_{n\to\infty} s_n) = f(s)$. We apply this in the case $f(x) = \sqrt{x}$:

$$s_{n+1} = \sqrt{1 + s_n} \quad \Rightarrow \quad \lim_{n\to\infty} s_{n+1} = \sqrt{\lim_{n\to\infty} (1 + s_n)} = \sqrt{1 + s}.$$

It is an easy exercise to show that $\lim_{n\to\infty} s_{n+1} = \lim_{n\to\infty} s_n = s$.[1] Therefore, we see

$$s = \lim_{n\to\infty} s_n = \lim_{n\to\infty} s_{n+1} = \sqrt{1 + s}.$$

Squaring both sides,

$$s^2 = 1 + s \quad \Rightarrow \quad s^2 - s - 1 = 0 \quad \Rightarrow \quad s = \frac{1 \pm \sqrt{5}}{2}$$

[1] Here we are comparing the sequence $\{s_n\}$ to the "shifted sequence" $\{s_{n+1}\} = s_2, s_3, \ldots$. If s_n is within ε of s for all $n > N$, then so is s_{n+1} since $n > N$ implies that $n + 1 > N$.

by the quadratic equation. By squaring, we introduced an extraneous answer. Since $0 \leq s_n$ for all n, it follows that $0 = \lim 0 \leq \lim s_n = s$. [2] Hence, s must be nonnegative. It follows that

$$s = \frac{1+\sqrt{5}}{2}.$$

[2] In general, taking limits "preserves inequalities". We will discuss this later and assume it for now.

Week 9, Friday: Misc. limit theorems, infinite limits, and Cauchy sequences

(Supplemental reading: Sections 8.3, 8.4, 8.7, and 8.8 in Swanson.)

Miscellaneous limit theorems

We state several theorems here about limits. The numbering refers to Swanson's text, where the interested reader may find the proofs.

Theorem 8.4.1. Limits are unique: if $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} s_n = s'$, then $s = s'$.

Proof. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} s_n = s'$, there exists an N such that $n > N$ implies that $|s - s_n| < \varepsilon/2$ and $|s' - s_n| < \varepsilon/2$. As usual, we can take N large enough so that it applies to both limits simultaneously. By the triangle inequality, we then have that for $n > N$,

$$|s - s'| = |(s - s_n) - (s' - s_n)| \leq |s - s_n| + |s' - s_n| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

We have shown that $|s - s'| < \varepsilon$ for all $\varepsilon > 0$. Since there are no infinitely small positive numbers (proved in the earlier *Extrema* lecture), it follows that $|s - s'| = 0$, and hence, $s = s'$. □

Theorem 8.4.11. (Squeeze theorem.) Let $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ be *real* sequences, and suppose that $a_n \leq b_n \leq c_n$ for all n. Then, if $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n$, it follows that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \lim_{n\to\infty} c_n.$$

Example. Here, we prove that $\lim_{n\to\infty} \dfrac{n^2 - 3}{n^3 + 6n + 1} = 0$ using the squeeze theorem. We have

$$\frac{n^2 - 3}{n^3 + 6n + 1} \leq \frac{n^2}{n^3 + 6n + 1} \leq \frac{n^2}{n^3} = \frac{1}{n};$$

Thus,

$$0 \leq \frac{n^2 - 3}{n^3 + 6n + 1} \leq \frac{1}{n},$$

for $n > 1$. Given that $\lim_{n\to\infty} 0 = 0$ and $\lim_{n\to\infty} \frac{1}{n} = 0$, it follows that

$$\lim_{n\to\infty} \frac{n^2 - 3}{n^3 + 6n + 1} = 0$$

by the squeeze theorem.

Example. Consider the sequence $\left\{\frac{\sin(n)}{n}\right\}$. Since $|\sin(n)| \leq 1$, we have

$$-\frac{1}{n} \leq \frac{\sin(n)}{n} \leq \frac{1}{n}.$$

Then, since

$$\lim_{n\to\infty} -\frac{1}{n} = \lim_{n\to\infty} \frac{1}{n} = 0,$$

it follows that

$$\lim_{n\to\infty} \frac{\sin(n)}{n} = 0$$

by the squeeze theorem.

Theorem 8.4.10. The operation of taking limits preserves inequalities: Suppose that $\{a_n\}$ and $\{b_n\}$ are convergent *real* sequences and that $a_n \leq b_n$ for all n (or for all n past a certain point) then

$$\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n.$$

(Recall that $\mathbb{C}$ cannot be ordered. So to above theorem does not make sense for complex sequences in general.)

Remark. The operation of taking limits does not preserve strict inequality. For example, compare the constant sequence $\{0\}$ with the sequence $\left\{\frac{1}{n}\right\}$. We have

$$0 < \frac{1}{n}$$

for all n, but

$$\lim_{n\to\infty} 0 = \lim_{n\to\infty} \frac{1}{n}.$$

Subsequences

We form a subsequence from a giving sequence by dropping terms from the sequence (but leaving an infinite number):

sequence: $s_1, \underline{s_2}, \underline{s_3}, \underline{s_4}, s_5, \underline{s_6}, s_7, s_8, \underline{s_9}, s_{10}, \underline{s_{11}}, \ldots$

subsequence: $s_2, s_3, s_4, s_6, s_9, s_{11}, \ldots$

Definition. Let $\{s_n\}$ be a sequence, and let $n_0 < n_1 < n_2 < \ldots$ be any sequence of natural numbers. Then the sequence $\{s_{n_k}\}_{k=0}^{\infty}$ is called a *subsequence* of $\{s_n\}$.

Example. In the previous example,

$$n_0 = 2, n_1 = 3, n_2 = 4, n_3 = 6, n_4 = 9, n_5 = 11, \ldots$$

Example. Let $s_n = 1/n$ for $n = 1, 2, \ldots$. Then the following is a subsequence of $\{s_n\}$:

$$\{s_{2k}\}_{k=1}^{\infty} = s_2, s_4, s_6, s_8, \ldots$$

$$= \frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \ldots$$

Theorem. (Main theorem for subsequences.) If $\{s_n\}$ is a sequence converging to s and $\{s_{n_k}\}$ is any subsequence, then $\{s_{n_k}\}$ also converges to s. (Every subsequence of a convergent sequence is convergent and has the same limit.)

Example. The previous theorem is especially useful for proving non-convergence. For instance, consider the sequence $\{(-1)^n\}$. It has the constant sequence $\{1\}$ and the constant sequence $\{-1\}$ as subsequences. The former converges to 1 and the latter to -1. We can deduce that $\{(-1)^n\}$ diverges (since otherwise all subsequence would need to converge to the same value).

Infinite limits

Definition. A *real* sequence $\{a_n\}$ *diverges to* ∞ if for all $B \in \mathbb{R}$, there exists $N \in \mathbb{R}$ such that $n > N$ implies

$$a_n > B.$$

The sequence *diverges to* $-\infty$ if for all $B \in \mathbb{R}$, there exists $N \in \mathbb{R}$ such that $n > N$ implies

$$a_n < B.$$

We write

$$\lim_{n\to\infty} a_n = \infty \quad \text{or} \quad \lim_{n\to\infty} a_n = -\infty,$$

respectively.

What this means: If $\lim_{n\to\infty} s_n = \infty$, and you come up with a number B, no matter how large, eventually, all of the terms in the sequences are larger than B.

If $\lim_{n\to\infty} s_n = -\infty$, and you come up with a number B, no matter how negative, eventually, all of the terms in the sequences are less than B.

Example. To prove $\lim_{n\to\infty} n^2 = \infty$, let $B \in \mathbb{R}$. Given B, let $N = \sqrt{|B|}$. Then

$$n > N \quad \Rightarrow \quad n > \sqrt{|B|} \quad \Rightarrow \quad n^2 > |B| \quad \Rightarrow \quad n^2 > B.$$

Cauchy sequences.

The real numbers fill the "holes" in $\mathbb{Q}$. For instance, consider the set of rational numbers

$$S = \left\{x \in \mathbb{Q} : x^2 \leq 2\right\}.$$

This is a nonempty set of $\mathbb{Q}$ and bounded above, but it does not have a supremum in $\mathbb{Q}$. One way to construct $\mathbb{R}$ is to use decimals. However that approach is a little tougher than it might first seem. (To point out one difficulty, two different decimals can be equal, e.g., $0.999\cdots = 1.000\ldots$.) We will now outline a different approach. There are sequences of rationals that "want to converge". For example, take the sequence we get by truncating the decimal expansion of $\sqrt{2}$:

$$1, 1.4, 1.41, 1.414, 1.4142, \ldots$$

However, without real numbers, this sequence has nowhere to converge to.

A key idea in one approach to the construction of $\mathbb{R}$ is to *think of sequences of rational numbers* as real numbers. For instance, we can think of the above sequence as $\sqrt{2}$. A real number that is already a rational number, e.g., $1/2$, can be thought of as the constant sequence $1/2, 1/2, 1/2, \ldots$ There are some obvious problems with this approach:

1. We only want to consider sequences of rationals that "want to converge". What could this mean?

2. There are many different sequences that want to converge to the same point. For instance, the constant sequence $\{0\}$ and the sequence $\{1/n\}$ both converge to 0, as do infinitely many other sequences.

We fix the first problem with the following definition:

Definition. A sequence of numbers (rational, real, or complex) $\{s_n\}$ is a *Cauchy sequence* if for $\varepsilon > 0$, there exists $N \in \mathbb{R}$ such that $m, n > N$ implies

$$|s_m - s_n| < \varepsilon.$$

Remark. The rough idea behind this definition is that as you go out far in the sequence, the points in the sequence start to clump together—the distance between all the remaining points is small.

Proposition. Every convergent sequence is a Cauchy sequence.

Proof. Suppose that $\{s_n\}$ is a convergent sequence, and let $\varepsilon > 0$. Say $\lim_{n\to\infty} s_n = s$. Then there exists N such that $n > N$ implies

$$|s - s_n| < \frac{\varepsilon}{2}.$$

But then, if $m, n > N$, by the triangle inequality,

$$|s_m - s_n| = |(s - s_n) - (s - s_m)| \leq |s - s_n| + |s - s_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

We've shown that $|s_m - s_n| < \varepsilon$ for all $m, n > N$. Hence $\{s_n\}$ is Cauchy. □

Theorem. A sequence of real or complex numbers is convergent if and only if it is a Cauchy sequence.

Proof. See Swanson's text, Section 8.7. □

Remark. The above theorem does not hold for $\mathbb{Q}$. For instance, take a sequence of rational numbers that converges to $\sqrt{2}$ in $\mathbb{R}$. That sequence will be a Cauchy sequence of rational numbers, but it will not converge in the rational numbers.

Definition of the real numbers

Let $\mathcal{C}$ be the set of all Cauchy sequences of rational numbers. As a first approximation, we could try to define $\mathbb{R} := \mathcal{C}$. The problem with this is that multiple Cauchy sequences will want to have the same limit. Again, for instance, consider the constant sequence 0 and the sequence $\{1/n\}$. To fix that we define an equivalence relation:

Definition. Let $\{s_n\}, \{t_n\} \in \mathcal{C}$. We say

$$\{s_n\} \sim \{t_n\}$$

if

$$\lim_{n\to\infty} (s_n - t_n) = 0.$$

Any easy check shows that $\sim$ is an equivalence relation on $\mathcal{C}$. An equivalence class for the relation is the set of all Cauchy sequences of rational numbers that want to converge to the same thing. We can then define the real numbers to be the sets of these equivalence classes:

Definition. The *real numbers* are

$$\mathbb{R} := \mathcal{C}/\sim .$$

with addition and multiplication defined by

$$[\{s_n\}] + [\{t_n\}] := [\{s_n + t_n\}]$$

$$[\{s_n\}][\{t_n\}] := [\{s_n t_n\}].$$

There are a lot of details to check:

1. Are addition and multiplication well-define or do they depend on the choice of representatives for the equivalence classes? (Recall we had the same consideration when defining addition and multiplication for equivalence classes of integers modulo n.)
2. Do we get a field?
3. How do we define an order relation on $\mathbb{R}$?

We'll stop here, though.

Week 10, Monday: Series

Series

(Supplemental reading: Section 9.1 in Swanson.)

Definition. Let $\{a_n\}$ be a sequence of real or complex numbers. The *n-th partial* sum of $\{a_n\}$ is

$$s_n := \sum_{i=1}^{n} a_i.$$

Example. The third partial sum of $\{1/n\}_{n\geq 1}$ is

$$s_3 = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}.$$

Note in the following definition that *a series is a special kind of sequence*, and, hence, all our earlier results about sequences apply.

Definition. Let $\{a_n\}$ be a sequence of real or complex numbers. The *infinite series* whose n-th term is a_n is the sequence of partial sums

$$\{s_n\} = a_1,\ a_1 + a_2,\ a_1 + a_2 + a_3, \ldots$$

If $\{s_n\}$ converges, say $\lim s_n = s$, then we write

$$\sum_{i=1}^{\infty} a_i := \lim_{n\to\infty} s_n = s,$$

and s is called the sum of the series. If $\{s_n\}$ diverges, we say the series diverges.

Example. Consider the series

$$\sum_{i=1}^{\infty} \frac{1}{n(n+1)}.$$

The first few partial sums are

$$s_1 = \frac{1}{1(1+1)} = \frac{1}{2}$$

$$s_2 = \frac{1}{1(1+1)} + \frac{1}{2(2+1)} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

$$s_3 = \frac{1}{1(1+1)} + \frac{1}{2(2+1)} + \frac{1}{3(3+1)} = \frac{3}{4}.$$

It looks like

$$s_n = \frac{n-1}{n}. \tag{25.1}$$

If that's the case, then

$$\sum_{i=1}^{\infty} \frac{1}{n(n+1)} = \lim_{n\to\infty} \frac{n-1}{n} = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right) = 1.$$

We could easily establish equation (25.1) by induction, but here is another (somewhat tricky) approach: note that

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

Thus, the n-th partial sum for the series is

$$s_n = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) + \left(\frac{1}{n} - \frac{1}{n+1}\right)$$

$$= \frac{1}{1} - \frac{1}{n+1}.$$

The above sum is called a *telescoping sum*—all its intermediate terms collapse. We have

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} := \lim_{n\to\infty} s_n = \lim_{n\to\infty} \left(\frac{1}{1} - \frac{1}{n+1}\right) = 1.$$

Proposition. (Limit theorem for series.) Suppose that $\sum_{n=1}^{\infty} a_n = a$ and $\sum_{n=1}^{\infty} b_n = b$, and let r a real or complex number. Then

1. $\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n = a + b.$
2. $\sum_{n=1}^{\infty}(ra_n) = r\sum_{n=1}^{\infty} a_n = ra.$

Proof. Let s_n and t_n be the n-th partial sums for $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$, respectively. We are given that $\lim_{n\to\infty} s_n = a$ and $\lim_{n\to\infty} t_n = b$. From our limit theorems for ordinary limits, we have $\lim_{n\to\infty}(s_n + t_n) = a + b$ and $\lim_{n\to\infty}(rs_n) = ra$. But $s_n + t_n$ is the n-th partial sum for $\sum_{n=1}^{\infty}(a_n + b_n)$ and rs_n is the n-th partial sum for $\lim_{n=1}^{\infty}(ra_n) = ra$. The result follows. □

Figure 25.1: A collapsible telescope.

GEOMETRIC SERIES

Definition. A *geometric series* is a series of the form

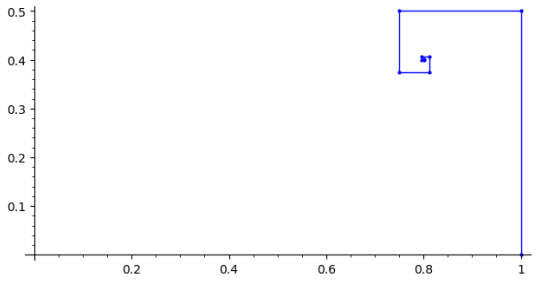
$$\sum_{n=0}^{\infty} r^n$$

where r is a real or complex number.

Example. Consider the geometric series $\displaystyle\sum_{n=0}^{\infty} \left(\frac{i}{2}\right)^n$. The index starts at $n = 0$, so the first few terms of the sequence of partial sums is

$$s_0 = 1,\ s_1 = 1 + \frac{i}{2},\ s_2 = 1 + \frac{i}{2} - \frac{1}{4},\ s_3 = 1 + \frac{i}{2} - \frac{1}{4} - \frac{i}{8},\ s_4 = 1 + \frac{i}{2} - \frac{1}{4} - \frac{i}{8} + \frac{1}{16}.$$

Here is a picture of the sequence of partial sums (connected by lines)



Theorem. (Geometric series) Let $r \in \mathbb{C}$.

1. If $|r| < 1$, then

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

   More generally if $|r| < 1$ and $m \in \mathbb{N}$, then

$$\sum_{n=m}^{\infty} r^n = \frac{r^m}{1-r}.$$

2. If $|r| \geq 1$, then $\sum_{n=0}^{\infty} r^n$ diverges.

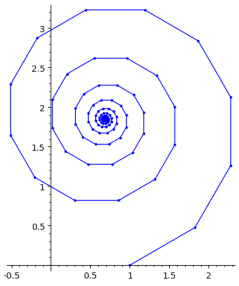
Example. Let

$$\alpha = (20/21)\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) = \frac{1}{2.1} + \frac{\sqrt{3}}{2.1}i,$$

a point with argument 30° and length slightly less than 1. According to the geometric series theorem, the series converges to

$$\sum_{n=0}^{\infty} \alpha^n = \frac{1}{1-\alpha} \approx 0.68 + 1.85\,i.$$

Here is a picture of the partial sums for $\sum_{n=0}^{\infty} \alpha^n$



Note that the angle at which each successive line rotates is $\arg(\alpha) = 30°$. Can you see why?

Example Compute the following sums:

(a) $\sum_{n=0}^{\infty} \left(\frac{i}{2}\right)^n$ (b) $\sum_{n=0}^{\infty} 5\left(\frac{2}{3}\right)^n$ (c) $\sum_{n=2}^{\infty} \left(\frac{3}{4}\right)^n$

(d) $\displaystyle\sum_{n=0}^{\infty} (4i)^n$ (e) $\displaystyle\sum_{n=3}^{\infty} 7\frac{2^{2n}}{10^n}$.

Note that in (c), the sum starts with $n = 2$, and in (e), the sum starts with $n = 3$.

SOLUTION: (a) Since $|i/2| = 1/2 < 1$, the formula for the sum of a geometric series applies, and we get

$$\sum_{n=0}^{\infty} \left(\frac{i}{2}\right)^n = \frac{1}{1-\frac{i}{2}} = \frac{2}{2-i} = \frac{2}{2-i} \cdot \frac{2+i}{2+i} = \frac{4+2i}{5}.$$

(b) Using the limit theorems for series and the formula for the sum of a geometric series, we have

$$\sum_{n=0}^{\infty} 5\left(\frac{2}{3}\right)^n = 5\sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = 5 \cdot \frac{1}{1-\frac{2}{3}} = 15.$$

(c) Using the formula for a geometric series, we have

$$\sum_{n=2}^{\infty} \left(\frac{3}{4}\right)^n = \frac{(3/4)^2}{1-\frac{3}{4}} = \frac{9}{4}.$$

Note: Another way to approach this problem is to first compute the sum starting with $n = 0$:

$$\sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n = \frac{1}{1-\frac{3}{4}} = 4,$$

and then subtract the terms corresponding to $n = 0, 1$:

$$4 - 1 - \frac{3}{4} = \frac{9}{4}.$$

This method takes more steps and, in practice, is a likely source of arithmetic errors.

(d) Since $|4i| = 4 \geq 1$, the series $\sum_{n=0}^{\infty} (4i)^n$ diverges.

(e) We have

$$\begin{aligned}
\sum_{n=3}^{\infty} 7\frac{2^{2n}}{10^n} &= \sum_{n=3}^{\infty} 7\frac{4^n}{10^n} \\
&= 7\sum_{n=3}^{\infty} \left(\frac{4}{10}\right)^n \\
&= 7\sum_{n=3}^{\infty} \left(\frac{2}{5}\right)^n \\
&= 7\left(\frac{2}{5}\right)^3 \sum_{n=0}^{\infty} \left(\frac{2}{5}\right)^n \\
&= 7\left(\frac{2}{5}\right)^3 \frac{1}{1-2/5} \\
&= 7\left(\frac{2}{5}\right)^3 \frac{5}{3} = \frac{7 \cdot 8}{25 \cdot 3} = \frac{56}{75}.
\end{aligned}$$

Proof of the geometric series theorem. An easy induction argument shows that

$$\sum_{i=0}^{n} r^n = \frac{1 - r^{n+1}}{1 - r}$$

for $r \neq 1$. Suppose that $|r| < 1$. By the above formula, the n-th partial sum of the series is

$$s_n = \frac{1 - r^{n+1}}{1 - r}.$$

Using our limit theorems, we get

$$\begin{aligned}
\lim_{n\to\infty} s_n &= \lim_{n\to\infty} \frac{1 - r^{n+1}}{1 - r} \\
&= \frac{1}{1 - r} \cdot \lim_{n\to\infty} (1 - r^{n+1}) && \text{(since } \frac{1}{1-r} \text{ is a constant)} \\
&= \frac{1}{1 - r} \left(\lim_{n\to\infty} 1 - \lim_{n\to\infty} r^{n+1} \right) \\
&= \frac{1}{1 - r} \left(\lim_{n\to\infty} 1 - 0 \right) && \text{(since } |r| < 1) \\
&= \frac{1}{1 - r}.
\end{aligned}$$

Next, use our limit theorem for series to get

$$\sum_{n=m}^{\infty} r^n = r^m + r^{m+1} + r^{m+2} + \cdots = r^m(1 + r + r^2 + \dots) = r^m \sum_{n=0}^{\infty} r^n = \frac{r^m}{1 - r}.$$

Now suppose that $|r| \geq 1$. If $r = 1$, then the n-th partial sum of the series is $s_n = n$, which gives a divergent sequence. Next, suppose that $|r| \geq 1$ and $r \neq 1$. Recall that in that case, we showed earlier that $\lim_{n\to\infty} r^n$ diverges. For the sake of contradiction suppose the series converges. Say $\sum_{n=0}^{\infty} r^n = s$. From our earlier formula, we know

$$s_n = \frac{1 - r^{n+1}}{1 - r}.$$

Solve for r^{n+1} to get $r^{n+1} = 1 - (1 - r)s_n$, and hence,

$$r^n = \frac{1}{r}(1 - (1 - r)s_n).$$

We are assuming $\lim_{n\to\infty} s_n = s$, and therefore, using our limit theorems

$$\lim_{n\to\infty} r^n = \frac{1}{r}(1 - (1 - r)s).$$

However, we know $\lim_{n\to\infty} r^n$ diverges when $|r| \geq 1$ and $r \neq 1$. This contradiction shows that $\sum_{n=0}^{\infty} r^n$ must diverge. □

Week 10, Wednesday: Series tests I

Series tests I

(Supplemental reading: Section 9.2 in Swanson.)

Our next goal is to present the standard collection of tests for determining whether a series converges.[1] Here is the list:

1. **the geometric series test**
2. **the n-th term test**
3. **the comparison test**
4. the limit comparison test
5. the alternating series test
6. the absolute convergence test
7. the ratio test
8. the root test
9. the integral test
10. the p-series test.

We have already discussed the geometric series test, but we will state it again here for completeness and for review.

1. The geometric series test. Let $r \in \mathbb{C}$. The series $\sum_{n=0}^{\infty} r^n$ converges if and only if $|r| < 1$. When $|r| < 1$,

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

More generally, for $a \in \mathbb{C}$ and $k \in \mathbb{N}$, if $|r| < 1$, then

$$\sum_{n=k}^{\infty} ar^n = \frac{ar^k}{1-r}.$$

Examples. (See earlier in our notes for more examples.)

[1] Note that knowing that a series converges is different from knowing its limit.

1.
$$\sum_{n=0}^{\infty}\left(\frac{i}{2}\right)^n = \frac{1}{1-\frac{i}{2}} = \frac{2}{2-i} = \frac{2}{2-i}\cdot\frac{2+i}{2+i} = \frac{4}{5}+\frac{1}{5}i.$$

2.
$$\sum_{n=2}^{\infty} 8\left(\frac{2^{n+2}}{5^n}\right) = 8\sum_{n=2}^{\infty}\frac{4\cdot 2^n}{5^n} = 32\sum_{n=2}^{\infty}\left(\frac{2}{5}\right)^n = 32\left(\frac{2}{5}\right)^2\frac{1}{1-\frac{2}{5}} = \frac{128}{15}.$$

3. $\sum_{n=1}^{\infty}(4i)^n$ diverges since $|4i| = 4 \geq 1$.

2. The n-th term test. The n-th term test is a criterion for divergence of a complex series. It says that if $\lim_{n\to\infty} a_n \neq 0$, then $\sum a_n$ diverges. If $\lim_{n\to\infty} a_n = 0$, we cannot conclude that $\sum a_n$ converges, as we will see with the harmonic series, below.

Proposition. If $\lim_{n\to\infty} a_n \neq 0$, then $\sum a_n$ diverges.

Proof. To prove this result, we prove its contrapositive[2]: if $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. To see this, suppose that $\sum a_n = s$, i.e., $\lim s_n = s$ where $s_n = \sum_{k=1}^{n} a_k$ is the n-th partial sum of $\{a_n\}$. We have that $s_n - s_{n-1} = a_n$. Therefore, using our limit theorems

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty}(s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = s - s = 0.$$

□

Examples.

1. The series $\sum_{n=1}^{\infty}(-1)^n$ diverges since $\lim_{n\to\infty}(-1)^n$ does not exist. In particular, $\lim_{n\to\infty}(-1)^n \neq 0$.

2. The series $\sum_{n=1}^{\infty}\frac{3n^2-2n+1}{2n^2+5}$ does not converge since $\lim_{n\to\infty}\frac{3n^2-2n+1}{2n^2+5} = \frac{3}{2} \neq 0$.

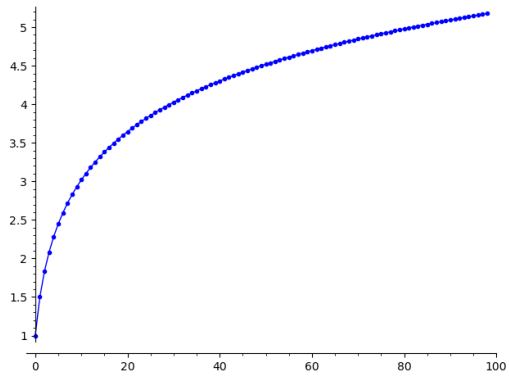
3. Consider series $\sum_{n=1}^{\infty}\frac{1}{n}$. We have $\lim\frac{1}{n} = 0$, however, as we will see below, this series *diverges*. So *the converse of the proposition does not hold*. The proposition can only be used to prove divergence (not convergence).

The harmonic series. The *harmonic series* is the series

$$\sum_{n=1}^{\infty}\frac{1}{n} = 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\dots$$

Here is a plot of its partial sums, $1,\ 1+1/2,\ 1+1/2+1/3,\dots$:

[2]The *contrapositive* of a statement "P implies Q" is "not Q implies not P".



It looks like these partial sums might be converging, but in fact they do not. You may notice that this plot is reminiscent of the graph of the logarithm, which hints that to see substantial growth in the partial sums, we will need to consider partial sums of an exponentially-growing number of terms of the series. The first proof that the harmonic series diverges, which we present below, is due to Nicole Orseme around 1350. The theorem is a special case of the p-series test which we will discuss later and easily prove using the integral test.

Theorem. The harmonic series diverges.

Proof. The first partial sum of the harmonic series is

$$\sum_{n=1}^{1} \frac{1}{n} = 1.$$

The second partial sum is

$$\sum_{n=1}^{2} \frac{1}{n} = 1 + \frac{1}{2}$$

The fourth:

$$\sum_{n=1}^{2^2} \frac{1}{n} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) \geq 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 1 + \frac{1}{2} + \frac{1}{2} = 1 + 2 \cdot \frac{1}{2}.$$

The eighth:

$$\begin{aligned}\sum_{n=1}^{2^3} \frac{1}{n} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \\ &\geq 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) \\ &= 1 + 3 \cdot \frac{1}{2}.\end{aligned}$$

An induction proof shows that

$$s_{2^k} = \sum_{n=1}^{2^k} \frac{1}{n} \geq 1 + k \cdot \frac{1}{2}.$$

for all $k \geq 0$. Thus, the sequence of partial sums for the harmonic series is unbounded and, hence, diverges. □

3. The comparison test. In the following proposition, notice that the test only applies to *nonnegative real sequences*. Also, notice the crucial role of the monotone convergence theorem in the proof. (Recall that MCT says that if a real sequence is monotone and bounded above, then it converges.)

Proposition (series comparison test). Suppose that $\{a_n\}$ and $\{b_n\}$ are real sequences with

$$0 \leq a_n \leq b_n$$

for all n. Then

(a) $\sum b_n$ converges $\Rightarrow \sum a_n$ converges.

(b) $\sum a_n$ diverges $\Rightarrow \sum b_n$ diverges.

Proof. Let

$$s_n = a_1 + a_2 + \cdots + a_n$$
$$t_n = b_1 + b_2 + \cdots + b_n$$

be the respective partial sums. Since $a_n \leq b_n$ for all n, we have $s_n \leq t_n$ for all n. Since both sequences have nonnegative terms, their sequences of partial sums are monotone increasing. To prove part (a), suppose that $\sum b_n$ converges, and say $\sum b_n = t$. This means that $\lim t_n = t$. By the monotone convergence theorem, we know that $t = \sup\{t_n\}$. Thus,

$$s_n \leq t_n \leq t = \sup\{t_n\}.$$

Hence, $\{s_n\}$ is both monotone increasing and bounded above. Again by the monotone convergence theorem, $\{s_n\}$ converges, i.e., $\sum a_n$ converges.

Part (b) is the contrapositive of part (a) and, hence, follows immediately. □

Examples.

1. For all $n \geq 1$

$$0 \leq \frac{1}{(n+1)^2} \leq \frac{1}{n(n+1)}.$$

We showed in an earlier lecture that $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges to 1. Thus, by comparison to $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$,

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)^2} = \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots$$

converges. It then follows that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots$$

(The partial sums of the last two displayed series differ by 1, so the convergence of one implies the convergence of the other.)

2. The series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2 + \sqrt{n}}$ converges by comparison with $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ since

$$0 \leq \frac{1}{n^2 + \sqrt{n}} \leq \frac{1}{n^2}.$$

3. The series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ diverges by comparison with $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ since

$$0 \leq \frac{1}{n} \leq \frac{1}{\sqrt{n}}$$

for all n, and the harmonic series, $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$, diverges.

We will continue with our list of tests next time.

Week 10, Friday: Series tests II

Series tests II

(Supplemental reading: Section 9.2 in Swanson.)

We continue our discussion of the standard tests for determining whether a series converges.

1. the geometric series test
2. the n-th term test
3. the comparison test
4. **the limit comparison test**
5. **the alternating series test**
6. **the absolute convergence test**
7. the ratio test
8. the root test
9. the integral test
10. the p-series test.

4. The limit comparison test. In the following, note the key hypotheses that the sequences have *positive* terms and that the limit of their quotient is *nonzero*. Since a sequence converges if and only if its tail converges, one may also apply the limit comparison test to sequences that are positive after a finite number of terms.

Proposition (limit comparison test). Suppose $\{a_n\}$ and $\{b_n\}$ are real sequences of *positive* terms and that

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L \neq 0.$$

Then $\sum a_n$ converges if and only if $\sum b_n$ converges.

Remark. Whether $\sum_n a_n$ converges depends upon how quickly the terms a_n die off. For instance, $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges but $\sum_{n=1}^{\infty} \frac{1}{n}$ does not. Even though $\lim \frac{1}{n^2} = \lim \frac{1}{n} = 0$, the terms of $\left\{\frac{1}{n}\right\}$ do not die off quickly enough. The condition $\frac{a_n}{b_n} = L \neq 0$ means

that a_n and b_n grow or diminish at comparable rates, unlike the case of $\sum_{n=1}^{\infty} \frac{1}{n^2}$ and $\sum_{n=1}^{\infty} \frac{1}{n}$.

Example. Consider the series

$$\sum_{n=1}^{\infty} \frac{6n+7}{2n^2-4}.$$

For large n,

$$\frac{6n+7}{2n^2-4} \approx \frac{6n}{2n^2} = \frac{3}{n}.$$

Since the harmonic series, $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges, we expect $\sum_{n=1}^{\infty} \frac{6n+7}{2n^2-4}$ will also diverge. *To make this intuition precise*, we use the limit comparison test with $a_n = \frac{6n+7}{2n^2-4}$ and $b_n = \frac{1}{n}$:

$$\frac{a_n}{b_n} = \frac{6n+7}{2n^2-4} \cdot \frac{n}{1} = \frac{6n^2+7n}{2n^2-4} \longrightarrow 3 \neq 0$$

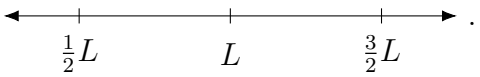
as $n \to \infty$. Thus, $\sum_{n=1}^{\infty} \frac{6n+7}{2n^2-4}$ diverges by limit comparison with the harmonic series.

Exercise. By a similar argument, show that $\sum_{n=1}^{\infty} \frac{7}{2n^2-4}$ converges using limit comparison with $\sum_{n=1}^{\infty} \frac{1}{n^2}$.

Proof of the limit comparison test. Suppose that $\lim \frac{a_n}{b_n} = L \neq 0$. Since $a_n > 0$, $b_n > 0$, and $L \neq 0$, we have $L > 0$. Apply the definition of the limit with $\varepsilon = \frac{1}{2}L > 0$ to find N such that $n > N$ implies

$$\left| L - \frac{a_n}{b_n} \right| < \frac{1}{2}L.$$

Thus, $\frac{a_n}{b_n}$ is within a distance of $L/2$ from L, i.e., it is in the interval pictured below:



Then

$$\frac{1}{2}L < \frac{a_n}{b_n} < \frac{3}{2}L \quad \Rightarrow \quad \frac{1}{2}Lb_n < a_n < \frac{3}{2}Lb_n.$$

If $\sum a_n$ converges, the ordinary comparison theorem implies that $\sum \frac{1}{2}Lb_n$ converges. Hence, using our earlier limit theorems[1], so does $\frac{2}{L}\sum \frac{1}{2}Lb_n = \sum b_n$. Similarly, if $\sum b_n$ converges, then so does $\frac{3}{2}L\sum b_n = \sum \frac{3}{2}Lb_n$. Then since $a_n < \frac{3}{2}Lb_n$, the comparison theorem says $\sum a_n$ converges. □

[1] If $\sum c_n$ converges and k is a constant, then $\sum kc_n$ converges and equals $k\sum c_n$.

5. The alternating series test.

We saw earlier that if $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. We also saw that the converse does not hold, in general. For instance $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges even though $\lim_{n\to\infty} \frac{1}{n} = 0$. We now discuss the case of a special type of series for which the converse does hold.

Proposition (alternating series test). Let $\{a_n\}$ be a monotonically decreasing sequence of positive terms. Then the series $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges if and only if $\lim_{n\to\infty} a_n = 0$.

Examples. The *alternating harmonic series* is the series

$$\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

To see that it converges, apply the alternating series test with $a_n = \frac{1}{n}$. (Check that the test applies: $\left\{\frac{1}{n}\right\}$ is a *monotonically decreasing* sequence of *positive* terms, and $\lim_{n\to\infty} \frac{1}{n} = 0$.)

The alternating harmonic series is the canonical example of what is called a *conditionally convergent* series: it converges, but the series formed by taking the absolute values of it terms, $\sum_{n=1}^{\infty} |(-1)^{n+1}\frac{1}{n}| = \sum_{n=1}^{\infty} \frac{1}{n}$, diverges.

Proof of the alternating series test.

($\Rightarrow$) Suppose that $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges. By the n-th term test, it follows that $\lim_{n\to\infty}(-1)^{n+1}a_n = 0$. We have seen that if $\{b_n\}$ is any sequence, real or complex, that $\lim_{n\to\infty} b_n = 0$ if and only if $\lim_{n\to\infty} |b_n| = 0$. Applying that here, we conclude that $\lim_{n\to\infty} a_n = 0$.

($\Leftarrow$) Now suppose that $\lim_{n\to\infty} a_n = 0$. We must argue that the sequence of partial sums $\{s_n\}$ for $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges. Our strategy is to divide these partial sums into two subsequences, $\{s_{2n}\}$ and $\{s_{2n+1}\}$—the even- and the odd-indexed terms—and to argue that these two subsequences converge to the same value. First consider $\{s_{2n}\}$:

$$s_{2n} = (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}).$$

The terms of s_{2n} are grouped as above to make it clear that since $\{a_n\}$ is monotonically decreasing, each of the $a_{2k-1} - a_{2k}$ is nonnegative. Thus, $\{s_{2n}\}$ is monotonically increasing. Further, looking at this sequence a different way makes it clear that s_{2n} is bounded above by a_1:

$$s_{2n} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - (a_{2n-2} - a_{2n-1}) - a_{2n} \leq a_1.$$

By the monotone convergence theorem, we conclude that

$$\lim_{n\to\infty} s_{2n} = s$$

for some $s \in \mathbb{R}$. It then follows that

$$\lim_{n\to\infty} s_{2n+1} = \lim_{n\to\infty} (s_{2n} + a_{n+1}) = \lim_{n\to\infty} s_{2n} + \lim_{n\to\infty} a_{n+1} = s + 0 = s.$$

(We know that $\lim_{n\to\infty} a_{n+1} = 0$ by the n-term test.) Since

$$\lim_{n\to\infty} s_{2n} = \lim_{n\to\infty} s_{2n+1} = s,$$

it looks like there is some hope of showing that $\lim_{n\to\infty} s_n = s$. In fact, that is true, and we can give a straight ε-N proof. Let $\varepsilon > 0$, and then take N so that $n > N$ implies.

$$\lim_{n\to\infty} |s - s_{2n}| < \varepsilon \quad \text{and} \quad \lim_{n\to\infty} |s - s_{2n+1}| < \varepsilon$$

simultaneously. But this just says that every term of the sequence $\{s_n\}$ is within ε of s: $n > 2N + 1$ implies $|s - s_n| < \varepsilon$. □

6. The absolute convergence test.

Definition. A series of complex numbers $\sum a_n$ is *absolutely convergent* if $\sum |a_n|$ is convergent. If $\sum a_n$ converges but $\sum |a_n|$ does not, then $\sum a_n$ is *conditionally convergent*.

Example. To emphasize the example presented earlier: the alternating harmonic series, $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$, converges (by the alternating series test), but $\sum_{n=1}^{\infty} \left|(-1)^{n+1} \frac{1}{n}\right| = \sum_{n=1}^{\infty} \frac{1}{n}$ does not. So the alternating harmonic series is conditionally convergent.

Many of our series tests apply to only series whose terms are nonnegative reals. The following proposition is of central importance: it shows how these series tests can say something about arbitrary complex series (since the absolute value of a complex number is nonnegative and real).

Proposition (absolute convergence test). Let $\sum a_n$ be a complex series. Then if $\sum a_n$ is absolutely convergent, it is convergent:

$$\sum |a_n| \text{ convergent} \quad \Longrightarrow \quad \sum a_n \text{ convergent.}$$

Examples. The series

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n^2} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \cdots$$

converges by the absolute convergence test since

$$\sum_{n=1}^{\infty} \left|(-1)^{n+1} \frac{1}{n^2}\right| = \sum_{n=1}^{\infty} \frac{1}{n^2},$$

and we have seen that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

Similarly,

$$\sum_{n=1}^{\infty} \frac{\cos(n) + i\sin(n)}{n^2}$$

is absolutely convergent since

$$\sum_{n=1}^{\infty} \left| \frac{\cos(n) + i\sin(n)}{n^2} \right| = \sum_{n=1}^{\infty} \frac{|\cos(n) + i\sin(n)|}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Proof of the absolute convergence test. Suppose that $\sum_{n=1}^{\infty} |a_n|$ converges. We will use the fact that a real or complex series converges if and only it is a Cauchy sequence in order to prove that $\sum_{n=1}^{\infty} a_n$ converges.

Define

$$s_n = a_1 + \cdot + a_n \quad \text{and} \quad \tilde{s}_n = |a_1| + \cdot + |a_n|,$$

the partial sums for $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} |a_n|$, respectively. We are given that $\{\tilde{s}_n\}$ converges and must show that $\{s_n\}$ converges. Let $\varepsilon > 0$. Since $\{\tilde{s}_n\}$ converges, it is a Cauchy sequence. Thus, there exists N such that $m, n > N$ implies

$$|\tilde{s}_m - \tilde{s}_n| < \varepsilon.$$

Without loss of generality, suppose that $m \geq n$. Then, using the triangle inequality,

$$\begin{aligned} |s_m - s_n| &= |a_{n+1} + \cdots + a_m| \\ &\leq |a_{n+1}| + \cdots + |a_m| \\ &= ||a_{n+1}| + \cdots + |a_m|| \\ &= |\tilde{s}_m - \tilde{s}_n| \\ &< \varepsilon. \end{aligned}$$

Thus, $\{s_n\}$ is a Cauchy sequence and therefore converges. In other words, $\sum_{n=1}^{\infty} a_n$ converges. □

A peculiar property of conditionally convergent series. What happens if you rearrange the terms of an infinite series? To be precise, define a *rearrangement* of a series $\sum a_n$ to be a series $\sum b_n$ where the elements of $\{a_n\}$ and $\{a_n\}$ are in bijection—same sums but in different orders. By the commutative law for addition, one might expect to get the same behavior: one of the series converges if and only if the other does, and it they do converge, they converge to the same value. It turns out that *this is true for absolutely convergent series but not for conditionally convergent ones.* For instance, all rearrangements of the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ will converge to the same value. On the other hand the value of a rearrangement of $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$ will depend on the rearrangement—even whether the series will converge. Something even wilder is true:

Proposition. Let $\sum a_n$ be a conditionally convergent real series, and let a be any real number. Then there is a rearrangement of $\sum a_n$ that converges to a. Further, there are rearrangements of $\sum a_n$ that diverge to ∞, that diverge to $-\infty$, and that fail to have any limit.

The idea behind the proof is the following. Say $\sum a_n$ is conditionally convergent. Then let $\sum p_n$ be the same as $\sum a_n$ after setting all negative terms, $a_n < 0$, equal to zero. Similarly, let $\sum q_n$ be the same as $\sum a_n$ after setting all positive terms equal to zero. Using the monotone convergence theorem, one may show that since $\sum a_n$ is conditionally convergent, $\sum p_n$ diverges to ∞ and $\sum q_n$ diverges to $-\infty$. To get a rearrangement of $\sum a_n$ that converges to an arbitrary real number a do the following. Assume $a > 0$, the case of $a \leq 0$ being similar. Create the rearrangement of $\sum a_n$ in steps. First add enough terms of $\sum p_n$ until we first get a number bigger than a. That's possible since $\sum p_n$ diverges to ∞. Next, add enough terms from $\sum q_n$ until we first get a number less than a, possible since $\sum q_n$ diverges to $-\infty$. Continue now by adding further terms from $\sum p_n$ until we first get a number above a, and then add further terms from $\sum q_n$ until we first get below a. Continue ad infinitum. The next part of the argument is to show the resulting rearrangement converges to a. A similar argument holds in the cases where $a = \pm\infty$.

Week 11, Monday: Series tests III

Series tests III

(Supplemental reading: Section 9.2 in Swanson.)

We continue our discussion of the standard tests for determining whether a series converges:

1. the geometric series test
2. the n-th term test
3. the comparison test
4. the limit comparison test
5. the alternating series test
6. the absolute convergence test
7. **the ratio test**
8. **the root test**
9. **the integral test**
10. **the p-series test**.

7. The ratio test.

Proposition (ratio test). Let $\{a_n\}$ be a sequence of positive real numbers, and suppose that

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = R.$$

Then

1. if $R < 1$, then $\sum a_n$ converges;
2. if $R > 1$ or $R = \infty$, then $\sum a_n$ diverges;
3. if $R = 1$, the test is inconclusive.

Example. The ratio test shows that $\sum_{n=0}^{\infty} \frac{1}{n!}$ converges since, as $n \to \infty$,

$$\frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \frac{n!}{(n+1)!} = \frac{1}{n+1} \longrightarrow 0 < 1.$$

Example. To see that the case $R = 1$ is inconclusive, consider the series $\sum_{n=1}^{\infty} \frac{1}{n}$ and $\sum_{n=1}^{\infty} \frac{1}{n^2}$. The former series diverges and the latter converges, yet we have

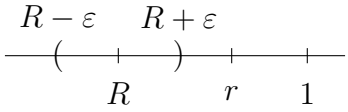
$$\lim_{n\to\infty} \frac{\frac{1}{n+1}}{\frac{1}{n}} = 1 \quad \text{and} \quad \lim_{n\to\infty} \frac{\frac{1}{(n+1)^2}}{\frac{1}{n^2}} = 1.$$

Example. Even if a series does not have positive terms, one can use the ratio test to consider if a series is absolutely convergent. For instance, we can see that the series $\sum_{n=}^{\infty} (-1)^n n \left(\frac{1}{2^n}\right)$ is absolutely convergent (hence, convergent), since

$$\lim_{n\to\infty} \frac{(n+1)\left(\frac{1}{2^{n+1}}\right)}{n\left(\frac{1}{2^n}\right)} = \lim_{n\to\infty} \frac{n}{n+1} \cdot \frac{1}{2} = \frac{1}{2} < 1.$$

Note. The tell-tale sign that the ratio test might apply is the presence of factorials and exponents.

Proof of the ratio test. First suppose that $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = R < 1$. Fix any real number r such that $0 \le R < r < 1$. Our goal, roughly, is to apply the comparison test to $\sum a_n$, comparing it with the convergent geometric series $\sum r^n$. Applying the definition of the limit, we can find N such that $n > N$ implies $|R - \frac{a_{n+1}}{a_n}| < \varepsilon$ where ε has been choose small enough so that this condition forces $\frac{a_{n+1}}{a_n} < r$:



Thus, $n > N$ implies that $0 \le \frac{a_{n+1}}{a_n} < r$, i.e.,

$$0 \le a_{n+1} < a_n r.$$

In particular, taking $n = N + 1$, this means

$$a_{N+2} < a_{N+1} r$$

and then

$$a_{N+3} < a_{N+2} r < a_{N+1} r^2.$$

Continuing,

$$a_{N+4} < a_{N+3} r < a_{N+1} r^3,$$

and so on. Letting $c := a_{N+1}$, we may show by induction that

$$a_{N+k} < cr^{k-1}$$

for $k \geq 2$. Since $|r| = r < 1$, the series

$$\sum_{k=2}^{\infty} cr^{k-1} = cr \sum_{k=0}^{\infty} r^k$$

converges. By the comparison test, the series $\sum_{n=N+2}^{\infty} a_n$ converges. Since this is a tail of our original series $\sum a_n$, the original series converges.

Now suppose that $R > 1$ or $R = \infty$. Then we can take N such that

$$n > N \quad \Rightarrow \quad \frac{a_{n+1}}{a_n} > 1 \qquad a_{n+1} > a_n.$$

By transitivity of $\geq$, we have $a_n \geq a_{N+1} > 0$ for all $n > N$. Thus, $\lim_{n\to\infty} a_n \neq 0$, and the series diverges by the n-th term test. □

8. The root test. We will not discuss this test, but include it here as one of the standard tests.

Proposition (root test). Let $\sum a_n$ be a series of nonnegative terms. Suppose that $\lim a_n^{1/n} = b$. If $b < 1$, the series converges. If $b > 1$, the series diverges. If $b = 1$, the test is inconclusive.

9. The integral test.

Proposition (integral test). Suppose $f(x)$ is a continuous, positive, decreasing function whose domain contains $(0, \infty)$. Then $\sum f(n)$ converges if and only if $\lim_n (\int_1^n f(x)\,dx)$ converges, i.e., if and only if $\int_1^\infty f(x)\,dx$ converges.

Example. We can use the integral test to give a quick proof that the harmonic series diverges:

$$\int_1^\infty \frac{1}{x}\,dx = \lim_{n\to\infty} \int_1^n \frac{1}{x}\,dx = \lim_{n\to\infty} \ln x\Big|_1^n = \lim_{n\to\infty} (\ln n - \ln 1) = \lim_{n\to\infty} \ln n = \infty.$$

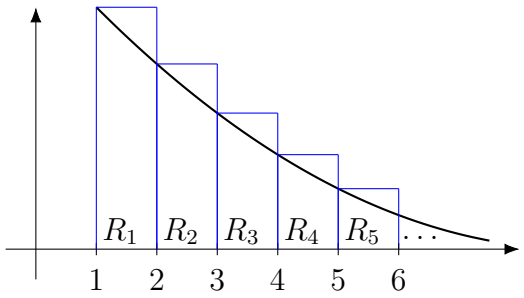
Thus, $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges by the integral test.

Similarly, we may show that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges:

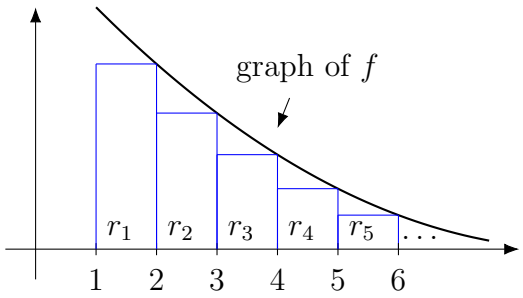
$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{n\to\infty} \int_1^n \frac{1}{x^2}\,dx = \lim_{n\to\infty} -\frac{1}{x}\Big|_1^n = \lim_{n\to\infty} (-\frac{1}{n} + 1) = 1.$$

Therefore, $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges by the integral test.

Proof of the integral test. Consider the standard pictures for the upper and lower Riemann sums for the integral of f:



upper sum



lower sum

For each $n = 1, 2, \ldots$, we have

$$\text{area}(R_n) = \text{base}(R_n) \cdot \text{height}(R_n) = 1 \cdot f(n) = f(n). \tag{28.1}$$

Similarly,

$$\text{area}(r_n) = \text{base}(r_n) \cdot \text{height}(r_n) = 1 \cdot f(n+1) = f(n+1). \tag{28.2}$$

Further, since f is decreasing and nonnegative

$$\text{area}(r_n) \leq \int_n^{n+1} f(x)\,dx \leq \text{area}(R_n),$$

and thus, using (28.1) and (28.2),

$$f(n+1) \leq \int_n^{n+1} f(x)\,dx \leq f(n). \tag{28.3}$$

Summing, we get

$$\sum_{n=1}^{k} f(n+1) \leq \sum_{n=1}^{k} \int_n^{n+1} f(x)\,dx \leq \sum_{n=1}^{k} f(n).$$

Then note that

$$\sum_{n=1}^{k} \int_n^{n+1} f(x)\,dx = \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx + \cdots + \int_k^{k+1} f(x)\,dx = \int_1^{k+1} f(x)\,dx$$

since we are just adding areas under the graph of f. So the result follows from the ordinary comparison theorem applied to (28.3). □

10. The p-series test.

Proposition (p-series test). Let $p \in \mathbb{R}$. Then the series

$$\sum \frac{1}{n^p}$$

converges if and only if $p > 1$.

Example. The p-series tests says these series converge:

$$\sum_{n=1}^{\infty} \frac{1}{n^{3/2}}, \qquad \sum_{n=1}^{\infty} \frac{1}{n^{1.000001}}$$

and these series do not:

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt[3]{n}}, \qquad \sum_{n=1}^{\infty} \frac{1}{n^{0.999999}}.$$

Proof of the p-series test. Apply the integral test. We did the case $p = 1$ earlier. So assume $p \neq 1$. In that case, we get

$$\begin{aligned}
\int_1^{\infty} \frac{1}{x}\,dx &= \int_1^{\infty} x^{-1}\,dx \\
&= \lim_{n\to\infty} \int_1^{n} x^{-1}\,dx \\
&= \lim_{n\to\infty} \frac{1}{1-p} x^{1-p}\Big|_1^n \\
&= \frac{1}{1-p} \lim_{n\to\infty} (n^{1-p} - 1) \\
&= \begin{cases} \frac{1}{p-1} & \text{if } p > 1 \\ \infty & \text{if } p < 1. \end{cases}
\end{aligned}$$

□

Interesting question. Does the following sum converge:

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+\frac{1}{n}}} \ ?$$

The p-series test does not apply since the exponent is not constant.

Week 11, Wednesday: Limits of functions

Limits of functions

(Supplemental reading: Sections 4.1 and 4.2 in Swanson.)

We now switch our focus from limits of sequences and series to limits of *functions*. Let $F = \mathbb{R}$ or $\mathbb{C}$.

Recall that the definition of the derivative of a function f at a point a looks something like this:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Here we are taking the limit of the function

$$g(h) := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

and that function is not defined at exactly the point of interest, i.e., at $h = 0$:

$$g(0) = \frac{f(a+0) - f(a)}{0} = \frac{0}{0} = \text{ undefined.}$$

In this way, a main use of the limit of a function is to determine what value a function should have at a point at which it is not defined. We are able to make this determination since the function is nicely behaved at all nearby points. The following definition characterizes the types of points, called *limit points*, at which we might hope to compute a limit.

Definition. Let $A \subseteq F$. A point $x \in F$ is a *limit point* of A if every open ball centered at x contains a point of A not equal to x. In other words, for all $r > 0$ there exists $y \in B(x, r) \cap A$ such that $x \neq y$.

Roughly, a limit point x of a set A can be approximated arbitrary closely by points besides x that are contained in A. The limit point, itself may or may not be in A.

Examples.

1. The limit points of $A = (0, 1) \subset \mathbb{R}$ are all points in the closed interval $[0, 1]$. Note that 0 and 1 are limit points of A that are not in A.

2. The limit points of $A = (0,1) \cup \{7\} \subset \mathbb{R}$ are again the points of $[0,1]$. That's because there is an open interval (ball) about 7 that contains no points of A besides 7 (for instance, the interval $(6,8)$ contains no points in A besides 7). We naturally call 7 an *isolated point* of A.

3. The set of limit points of the "punctured ball" $B(0;1) \setminus 0$ of radius 1 centered at the origin in $\mathbb{C}$ is the closed ball

$$\overline{B(0;1)} := \{z \in \mathbb{C} : |z| \leq 1\}.$$

   Every point in the punctured open ball is a limit point, but so are the points on the boundary and the origin.

Definition. Let $F = \mathbb{R}$ or $\mathbb{C}$, as usual. Let $A \subseteq F$ and $f\colon A \to F$. Let $a \in F$ be a limit point of A. Then the *limit of $f(x)$ as x approaches a is $L \in F$* if for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $x \in A$ and $0 < |x - a| < \delta$, then

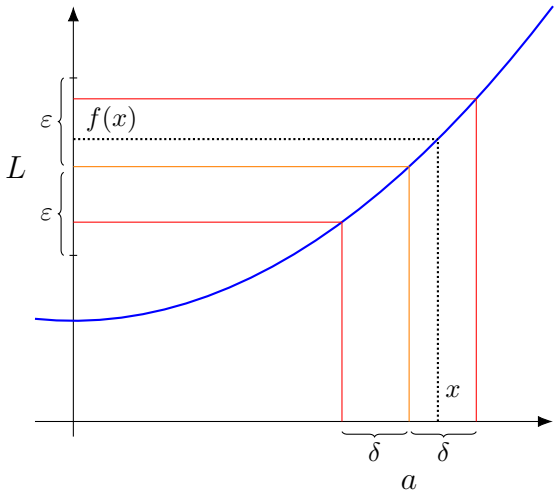
$$|f(x) - L| < \varepsilon.$$

If the limit is L, then we write $\lim_{x \to a} f(x) = L$.

Remark. If you wanted to pack the definition of $\lim_{x \to a} f(x) = L$ into symbols, you could write

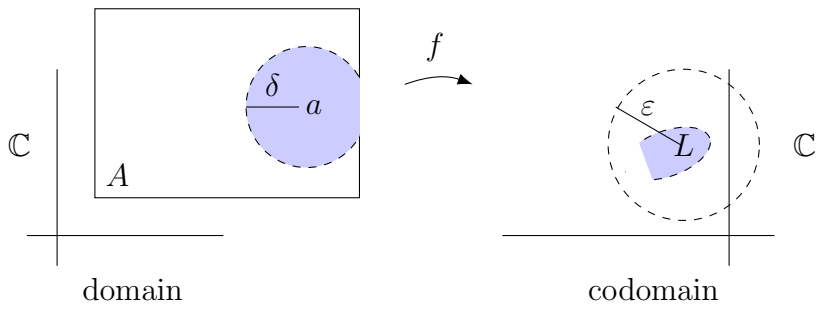
$$\forall \varepsilon > 0, \exists \delta > 0, \text{ s.t. } f\left((B(a;\delta) \setminus \{a\}) \cap A\right) \subseteq B(L;\varepsilon).$$

In the case where $F = \mathbb{R}$, we can interpret the definition of the limit using the graph of f:



Meaning of ε and δ in the definition of the limit when $F = \mathbb{R}$.

Consider now the case where the domain and codomain of f are in $\mathbb{C}$. The graph of f consist of points $(z, f(z))$ where z and $f(z)$ are in $\mathbb{C}$. So to picture the graph we would need four dimensions—two for z and two for $f(z)$. An alternative is to use two copies of the plane, $\mathbb{C}$, and picture the domain and codomain separately. We then try to picture how f moves points in the domain over to points in the codomain. With this set-up, here is the relevant picture for understanding the limit definition:



On the left, we have the codomain A of f, and the blue shaded region is the intersection of an open ball of radius δ with A. Note that for the purposes of the definition of the limit, we should remove the center of the ball, a. On the right, we have the ε-ball about L as the "target", and the blue shaded region is the image of the blue shaded region on the left. The picture shows that for the chosen value of ε, we were able to find a suitable δ.

Remarks.

- We are interested in the behavior of the function f near the point a, but not exactly at the point a. In fact, f need not even be defined at a. For example, consider the function
$$f(x) = \frac{x^2 - x}{x}.$$
If we try to evaluate f at 0, we get $f(0) = \frac{0}{0}$, which does not makes sense (you can't divide by 0), i.e., $\frac{0}{0}$ is not a number. However, the limit is exists at $x = 0$ (and is equal to -1).

- When you see the absolute values in the definition, you should think "distance". The *distance* between the numbers u and v is $|u - v|$. So you should translate $|f(x) - L| < \varepsilon$ as "the distance between $f(x)$ and L is less than ε".

- Consider the part of the definition that says $0 < |x - a| < \delta$. If the expression had just been $|x - a| < \delta$, without the "$0 <$" part, the requirement would be that the distance between the number x and a is less than δ. What does $0 < |x - a|$ add? The only way the absolute value of a number such as $x - a$ can be 0 is if $x - a = 0$ or, equivalently, $x = a$. Thus, requiring $0 < |x - a|$ is just requiring that x not equal a. This is just what we need since, after all, the function f may not be defined at a.

- Note the *quantifiers* "for all" and "there exists" in the definition. Just as with the definition of limits of sequences, it takes a while to appreciate their importance, but they are essential. First take the "for all" part. The definition say that for all $\varepsilon > 0$, we are going to want $|f(x) - L| < \varepsilon$. Translating: for all $\varepsilon > 0$, we will want to make the distance between $f(x)$ and L less that $\varepsilon > 0$. Our goal is to make f close to L, and the ε is a measure of how close. By making ε small and requiring $|f(x) - L| < \varepsilon$, we are ensuring that $f(x)$ is within a distance of ε from L.

  Next, consider the "there exists" part of the definition. It says that if you want $f(x)$ to be within a distance of ε of L, then it suffices to make $0 < |x - a| < \delta$. In other words, you can to make x within a distance of δ of a (remembering that we don't care what happens when $x = a$).

  Given any $\varepsilon > 0$ (a challenge to make $f(x)$ close to L), you want to find an appropriate distance $\delta > 0$ (so that if x is δ-close to a, then $f(x)$ is ε-close to L).

  With sequences, the game was: given ε find N. With functions, the game becomes: given ε, find δ.

Warning. At some point, in proving a statement of the form $\lim_{x\to a} f(x) = L$, you will be tempted to have δ be a function of x. That is not allowed! (On the other hand, δ is typically a function of ε, just as with sequences, N is typically a function of ε (but cannot be a function of n).)

Example. Consider the function

$$\begin{aligned} f\colon F &\to F \\ x &\mapsto 5x + 3. \end{aligned}$$

Then $\lim_{x\to 2} 5x + 3 = 13$.

Proof. Given $\varepsilon > 0$, let $\delta = \varepsilon/5$. Then if

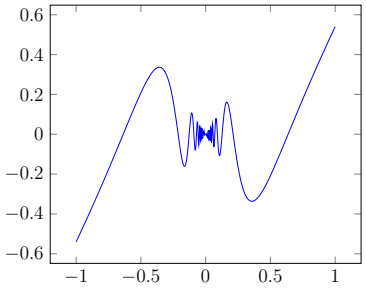
$$0 < |x - 2| < \delta,$$

it follows that

$$\begin{aligned} |f(x) - 13| &= |(5x + 3) - 13| \\ &= |5x - 10| \\ &= 5|x - 2| \\ &< 5\delta \\ &= 5 \cdot \frac{\varepsilon}{5} \\ &= \varepsilon. \end{aligned}$$

□

Example. Prove that $\lim_{x\to 0} x\cos(1/x) = 0$.



Graph of $f(x) = x\cos(1/x)$.

Proof. Given $\varepsilon > 0$, let $\delta = \varepsilon$. Suppose that $0 < |x - 0| < \delta$; in other words, suppose that $0 < |x| < \varepsilon$. Then, since $|\cos(y)| \leq 1$ for all y, we have

$$\begin{aligned} |x\cos(1/x)| &= |x||\cos(1/x)| \\ &\leq |x| \\ &< \delta \\ &= \varepsilon. \end{aligned}$$

□

Example. Proving limit statements for simple function like $f(x) = x^2$ can be a challenge. For instance, here we prove the "obvious" statement that $\lim_{x\to 5} x^2 = 25$. This example points out the need for better tools (e.g., a new-from-old limit theorem like we had with sequences) allowing us to avoid going down to the level of using ε-δ arguments.

Proof. Given $\varepsilon > 0$, let $\delta = \min\{1, \varepsilon/11\}$, i.e., δ is the minumum of 1 and $\varepsilon/11$. So $\delta \leq 1$ and $\delta \leq \varepsilon/11$ (with equality holding in at least one of these). Suppose that x satisfies $0 < |x - 5| < \delta$. Since $\delta \leq 1$, the distance between x and 5 is less than 1. It follows $4 < x < 6$, and hence, adding 5 across this string of inequalities, we get $9 < x + 5 < 11$. In particular, $|x + 5| < 11$. Therefore,

$$|x^2 - 25| = |(x + 5)(x - 5)| = |x + 5||x - 5| < 11|x - 5|.$$

Now, since $\delta \leq \varepsilon/11$ and $|x - 5| < \delta$, it follows that

$$|x^2 - 25| < 11|x - 5| < 11 \cdot \frac{\varepsilon}{11} = \varepsilon,$$

as required. □

Example. Claim: $\lim_{x\to 16} \sqrt{x} = 4$.

Proof. Given $\varepsilon > 0$, let $\delta = 4\varepsilon$, and suppose that

$$0 < |x - 16| < \delta = \frac{\varepsilon}{4}.$$

Then

$$\begin{aligned}
|\sqrt{x} - 4| &= \left|(\sqrt{x} - 4) \cdot \frac{\sqrt{x} + 4}{\sqrt{x} + 4}\right| \\
&= \frac{|x - 16|}{|\sqrt{x} + 4|} \\
&< \frac{|x - 16|}{4} \\
&= \frac{1}{4}\,|x - 16| \\
&< \frac{1}{4}\,\delta \\
&= \varepsilon.
\end{aligned}$$

□

Week 11, Friday: Continuity and derivatives

Continuity and derivatives

(Supplemental reading: Sections 5.1, 6.1, and 6.2 in Swanson.)

Our results will hold for real and complex functions unless noted. To treat both cases simultaneously, let $F = \mathbb{R}$ or $\mathbb{C}$.

New-from-old limit theorems.

Functions can sometimes be decomposed into sums and products of simpler functions. Taking a cue from what we did earlier with sequences, we first find limits of some simple functions using ε-δ proofs, and then we prove a general result that allows us to compute limits of more complicated functions built from these without resorting to ε-δ proofs.

Recall the definition of a limit of a function:

Definition. Let $A \subseteq F$ and $f\colon A \to F$. Let $a \in F$ be a limit point of A. Then the *limit of $f(x)$ as x approaches a is $L \in F$* if for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $x \in A$ and $0 < |x - a| < \delta$, then

$$|f(x) - L| < \varepsilon.$$

If the limit is L, then we write $\lim_{x\to a} f(x) = L$.

Now we compute the limits of a couple of simple functions. These, combined with the "new-from-old" limit theorem we prove next, will allow us to easily compute the limits of all rational functions, i.e., of all quotients of polynomials.

Proposition. Let $a, c \in F$. Then

1. $\lim\limits_{x\to a} c = c$.
2. $\lim\limits_{x\to a} x = a$.

Proof.

1 In this case, our function is $f(x) = c$ for all x, and we are claiming the limit is c. Given $\varepsilon > 0$,

$$|f(x) - c| = |c - c| = 0 < \varepsilon$$

for *all* x. This means the ε-δ definition of the limit is satisfied for any choice of $\delta > 0$, reflecting that fact that it is pretty easy to make $f(x)$ close to c!

2 Here, $f(x) = x$ for all x, and we claim the limit is a. Given $\varepsilon > 0$, let $\delta = \varepsilon$. Then, if we assume $0 < |x - a| < \delta = \varepsilon$, it follows that

$$|f(x) - a| = |x - a| < \varepsilon.$$

□

Here is the function-version of a "new-from-old" limit theorem, analogous to the one given earlier for sequences. It allows us to determine limits of functions that are built from simpler functions.

Theorem. Suppose that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Then,

1. $\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = L + M$,
2. $\lim_{x\to a} f(x)g(x) = (\lim_{x\to a} f(x))\,(\lim_{x\to a} g(x)) = LM$,
3. $$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{L}{M},$$
   provided $\lim_{x\to a} g(x) \neq 0$.

Proof. The proof of this theorem is almost identical to the proof of the limit theorems for sequences given earlier.

1 Given $\varepsilon > 0$ there exists $\delta > 0$ such that $0 < |x - a| < \delta$ implies

$$|f(x) - L| < \frac{\varepsilon}{2} \quad \text{and} \quad |g(x) - M| < \frac{\varepsilon}{2}.$$

(Again, as in the earlier proof for sequences, at first we might find a δ_1 for f and a δ_2 for g, then let $\delta := \min\{\delta_1, \delta_2\}$ to work simultaneously for both f and g.) It then follows that for $0 < |x - a| < \delta$,

$$\begin{aligned} |(f(x) + g(x)) - (L + M)| &= |(f(x) - L) + (g(x) - M)| \\ &\leq |f(x) - L| + |g(x) - M| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

We leave the proofs of parts 2 and 3 these exercises. □

Example. Claim: $\lim_{x\to 5} x^2 = 25$.

Proof. From the earlier Proposition, we have $\lim_{x\to 5} x = 5$. Using our limit theorem,

$$\lim_{x\to 5} x^2 = \lim_{x\to a}(x \cdot x) = \left(\lim_{x\to 5} x\right)\left(\lim_{x\to 5} x\right) = 5 \cdot 5 = 25.$$

Compare how easy this proof was compared to the earlier ε-δ proof in which we chose $\delta = \min\{1, \varepsilon/11\}$.

Example. Claim: $\lim_{x\to 1} \dfrac{x^2+5}{x^3-3x+1} = -6$.

Proof.

$$\begin{aligned}
\lim_{x\to 1} \frac{x^2+5}{x^3-3x+1} &= \frac{\lim_{x\to 1}(x^2+5)}{\lim_{x\to 1}(x^3-3x+1)} \\
&= \frac{(\lim_{x\to 1} x)^2 + \lim_{x\to 1} 5}{(\lim_{x\to 1} x)^3 + \lim_{x\to 1}(-3)\lim_{x\to 1} x + \lim_{x\to 1} 1} \\
&= \frac{1^2+5}{1^3 - 3\cdot 1 + 1} \\
&= -6.
\end{aligned}$$

□

Continuity.

In the previous example, we found the limit of $f(x) = \frac{x^2+5}{x^3-3x+1}$ at $x = 1$ to simply be $f(1)$. A similar proof would show that for any $a \in F$, we have $\lim_{x\to a} f(x) = f(a)$ (for this particular function f). This function is an example of a *continuous* function, where we can find the value of the limit simply by evaluating the function at the limit point. If all functions were continuous, there would be little need for the notion of a limit. Limits are more interesting when you need to determine the value of a function at a point where the function is undefined,

Definition. Let $A \subseteq F$ and $f\colon A \to F$. Then f *is continuous at* $a \in A$ if for all $\varepsilon > 0$ there exist $\delta > 0$ such that $|x-a| < \delta$ implies $|f(x)-f(a)| < \varepsilon$. We say f is *continuous* or *continous on* A if f is continuous at every point of A.

Remarks.

1. In comparing the definition of limits with that of continuity, note that in the latter, we have the condition $|x-a| < \delta$ rather than $0 < |x-a| < \delta$ and the condition $|f(x)-f(a)| < \varepsilon$ rather than $|f(x)-L| < \varepsilon$. This says that if every point of A is a limit point of A, for instance if A is an open or closed ball, then f is continuous on A if and only if
$$\lim_{x\to a} f(x) = f(\lim_{x\to a} x) = f(a)$$

for all $a \in A$. (Thus, we sometimes say the continuous functions are those that commute with limits.) In other words, to find the limit of $f(x)$ as x approaches a, we simply evaluate f at a. For instance, every polynomial or quotient of polynomials, every trig function, the exponential function, logarithms, the square root function, and the absolute value function are continuous wherever they defined.

On the other hand, if $a \in A$ is not a limit point of A, the definition of continuity implies that f is automatically continuous at a. We leave this as an exercise. A close look at the definition of limits reveals that limits are not defined for points that are not limit points of the domain.

2. Using the limit theorems, is easy to show that sums, products and quotients (where defined) of continuous functions are continuous.

3. A straightforward argument from the definitions shows that compositions of continuous functions are continuous.

Derivatives.

Definition. Let $A \subseteq F$, and let $a \in A$ be a limit point of A. Let $f\colon A \to F$. Then the *derivative of f at a* is

$$\frac{df}{dx}(a) := f'(a) := \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

provided this limit exists. If the limit exists, we say f is *differentiable* at a.

Remark. Equivalently, we could define the derivative at a to be

$$f'(a) := \lim_{x\to a} \frac{f(x) - f(a)}{x - a}.$$

Example. Let $f\colon \mathbb{C} \to \mathbb{C}$ be defined by $f(z) = z^2$. Then $f'(i) = 2i$.

Proof. Calculate:

$$\begin{aligned} f'(i) &= \lim_{h\to 0} \frac{f(i+h) - f(i)}{h} \\ &= \lim_{h\to 0} \frac{(i+h)^2 - i^2}{h} \\ &= \lim_{h\to 0} \frac{(i^2 + 2ih + h^2) - i^2}{h} \\ &= \lim_{h\to 0} \frac{2ih + h^2}{h} \end{aligned}$$

$$= \lim_{h\to 0}(2i + h)$$

$$= 2i.$$

□

The usual properties of derivatives hold over $\mathbb{C}$ as well as $\mathbb{R}$:

$$(z^n)' = nz^{n-1}, \quad (f+g)' = f' + g', \quad (cf)' = c(f') \text{ for a constant } c,$$

including the product and quotient rules:

$$(fg)' = f'g + fg', \qquad \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2},$$

and the chain rule:

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

Proposition. Differentiable functions are continuous. Suppose $f\colon A \to F$ is differentiable, and let a be a limit point of A. Then f is continuous at a.

Proof. Using our limit theorems, we have

$$0 = f'(a) \cdot 0$$

$$= \left(\lim_{x\to a} \frac{f(x) - f(a)}{x - a}\right)\left(\lim_{x\to a}(x - a)\right)$$

$$= \lim_{x\to a}\left(\frac{f(x) - f(a)}{x - a} \cdot (x - a)\right)$$

$$= \lim_{x\to a}\left(f(x) - f(a)\right).$$

Thus, $\lim_{x\to a}(f(x) - f(a)) = 0$. We now us the facts that $\lim_{x\to a}(f(x) - f(a))$ and $\lim_{x\to a} f(a)$ exist along with the sum formula for limits to see that $\lim_{x\to a} f(x)$ exists and

$$\lim_{x\to a} f(x) = \lim_{x\to a}(f(x) - f(a)) + \lim_{x\to a} f(a) = 0 + f(a) = f(a).$$

So $\lim_{x\to a} f(x) = f(a)$, as required. □

Proposition (product rule). Suppose that f and g are differentiable as some point $a \in F$. Then so is their product, fg, and

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a).$$

Proof. We have

$$
\begin{aligned}
(fg)'(x) &= \lim_{h\to 0} \frac{(fg)(a+h)-(fg)(a)}{h} && \text{(def. of the derivative)}\\
&= \lim_{h\to 0} \frac{f(a+h)g(a+h)-f(a)g(a)}{h} && \text{(def. of } fg\text{)}\\
&= \lim_{h\to 0} \frac{f(a+h)g(a+h)-f(a)g(a+h)+f(a)g(a+h)-f(a)g(a)}{h} && \text{(tricky!)}\\
&= \lim_{h\to 0} \frac{(f(a+h)-f(a))g(a+h)+f(a)(g(a+h)-g(a))}{h}\\
&= \lim_{h\to 0} \frac{(f(a+h)-f(a))g(a+h)}{h} + \lim_{h\to a} \frac{f(a)(g(a+h)-g(a))}{h}\\
&= \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \lim_{h\to a} g(a+h) + f(a) \lim_{h\to a} \frac{g(a+h)-g(a)}{h}\\
&= f'(a)g(a)+f(a)g'(a).
\end{aligned}
$$

The fact that $\lim_{h\to 0} g(a+h) = g(\lim_{h\to a}(a+h)) = g(a)$ follows from the fact that g, being differentiable at a, is continuous at a. □

Corollary (quotient rule). Suppose that f and g are differentiable as some point $a \in F$ and that $g(a) \neq 0$. Then the quotient, f/g is differentiable at a, and

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a)-f(a)g'(a)}{g(a)^2}.$$

Proof. Apply the product rule and the chain rule:

$$
\begin{aligned}
\left(\frac{f}{g}\right)'(a) &= \left(f\cdot\frac{1}{g}\right)'(a)\\
&= f'(a)\frac{1}{g(a)} + f(a)\left(\frac{1}{g}\right)'(a)\\
&= f'(a)\frac{1}{g(a)} + f(a)\left(-\frac{g'(a)}{g(a)^2}\right)\\
&= \frac{f'(a)g(a)-f(a)g'(a)}{g(a)^2}.
\end{aligned}
$$

□

Week 12, Monday: Power series I

Power series I

(Supplemental reading: Section 9.3 in Swanson.)

Definition. Let $\{a_n\}$ be a sequence of complex numbers. The series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

is a (complex) *power series* with n-th coefficient a_n.

Remarks.

1. A power series may be thought of as a family of ordinary series of the type we've just studied. We get one series for each point $z \in \mathbb{C}$. For instance, consider the series $\sum_{n=0}^{\infty} \frac{z^n}{n!}$. Letting $z = 1$ gives the series $\sum_{n=0}^{\infty} \frac{1}{n!}$, and letting $z = 2 + 3i$ gives $\sum_{n=0}^{\infty} \frac{(2+3i)^n}{n!}$.
2. The *n-th term* of the series is $a_n z^n$, and the *n-th coefficient* is a_n.
3. Suppose that $f(z) = \sum_{n=0}^{\infty} a_n z^n$ converges for all $D \subseteq \mathbb{C}$. Then f defines a function $f\colon D \to \mathbb{C}$.

Theorem (Main theorem for power series.) Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a complex power series. Then one of the following holds:

1. $f(z)$ converges only when $z = 0$.
2. $f(z)$ converges for all $z \in \mathbb{C}$.
3. There exists a real number $R > 0$ such that $f(z)$ converges absolutely when $|z| < R$ and diverges for $|z| > R$.

Definition. The number R defined above is called the *radius of convergence* for the series. We say $R = 0$ in case 1 and $R = \infty$ in case 2 of the theorem.

We will prove this theorem next time. To find the radius of convergence, one usually uses the ratio test.

Example. Find the radius of convergence for $f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$.

Solution. The ratio tests applies *after taking absolute values* of the terms:

$$\left|\frac{z^{n+1}}{(n+1)!}\right| \Big/ \left|\frac{z^n}{n!}\right| = \frac{n!}{(n+1)!} \cdot \frac{|z|^{n+1}}{|z|^n} = \frac{|z|}{n+1} \to 0,$$

as $n \to \infty$. So the ratio test says that the series converges absolutely for all $z \in \mathbb{C}$. Therefore, the radius of convergence is ∞.

Example. Find the radius of convergence of $f(z) = \sum_{n=0}^{\infty} \frac{1}{5^n} z^n$.

Solution. One could use the geometric series test here, but we'll use the ratio test again:

$$\left|\frac{1}{5^{n+1}} z^{n+1}\right| \Big/ \left|\frac{1}{5^n} z^n\right| = \frac{5^n}{5^{n+1}} \frac{|z|^{n+1}}{|z|^n} = \frac{|z|}{5} \to \frac{|z|}{5},$$

as $n \to \infty$. The ratio test says the series converges if

$$\frac{|z|}{5} < 1$$

and diverges if $\frac{|z|}{5} > 1$. We have

$$\frac{|z|}{5} < 1 \quad \Rightarrow \quad |z| < 5.$$

So the series converges absolutely for $|z| < 5$ and diverges for $|z| > 5$, i.e., the radius of convergence is 5. (On the boundary, where $|z| = 5$, we know the series diverges by the geometric series test: $\sum_{n=0}^{\infty} \left(\frac{z}{5}\right)^n$ diverges if $|z/5| \geq 1$.)

Example. Find the radius of convergence of the power series $\displaystyle\sum_{n=0}^{\infty} \frac{(3n)!}{n!(2n)!} z^n$.

Solution. Apply the ratio test:

$$\begin{aligned}
\left(\frac{(3(n+1))!}{(n+1)!(2(n+1))!}\right)|z|^{n+1} \Big/ \left(\frac{(3n)!}{n!(2n)!}\right)|z|^n &= \frac{(3(n+1))!}{(3n)!} \frac{n!}{(n+1)!} \frac{(2n)!}{(2(n+1))!}|z| \\
&= \frac{(3n+3)(3n+2)(3n+1)}{(n+1)(2n+2)(2n+1)}|z| \\
&\longrightarrow \frac{27}{4}|z|,
\end{aligned}$$

as $n \to \infty$. By the ratio test, we get convergence if

$$\frac{27}{4}|z| < 1,$$

or in other words,

$$|z| < \frac{4}{27}.$$

So the radius of convergence is $\frac{4}{27}$.

A version of the ratio test for power series. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$. Apply the ratio test for series (remembering to take absolute values since the ratio test requires positive terms):

$$\frac{|a_{n+1}||z|^{n+1}}{|a_n||z|^n} = \frac{|a_{n+1}|}{|a_n|}|z| \to r|z|,$$

where

$$r = \lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} \geq 0$$

which we are supposing exists. The ratio test then says we get convergence if

$$r|z| < 1.$$

If $r = 0$, the power series convergences for all z, and the radius of convergence is ∞. Otherwise, the power series converges for z satisfying

$$|z| < \frac{1}{r},$$

and the radius of convergence is $1/r$. However, note that

$$\frac{1}{r} = \frac{1}{\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|}} = \lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|}.$$

If $r = 0$, we'll have $\lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|} = \infty$. So we get the following result:

Proposition (Ratio test for power series). Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a complex power series with (eventually) nonzero coefficients, and suppose that

$$\lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|} = R \in \mathbb{R} \cup \{\infty\}.$$

Then R is the radius of convergence of f.

WARNING: Note that the ratio test for an ordinary series of positive terms $\sum_{n=0}^{\infty} a_n$ requires checking whether

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} < 1.$$

Compare this with the limit in the proposition. There are two differences: (i) We take absolute values since we are not assuming the terms in the power series are positive, and (ii) The order in which a_n and a_{n+1} appear in the numerator and denominator switch.

Example. Repeating the previous example using the power series ratio test rather than the ordinary ratio test we would do the following calculation:

$$\left(\frac{(3n)!}{n!(2n)!}\right) \Big/ \left(\frac{(3(n+1))!}{(n+1)!(2(n+1))!}\right) = \frac{(3n)!}{(3(n+1))!}\frac{(n+1)!}{n!}\frac{(2(n+1))!}{(2n)!}$$

$$= \frac{(n+1)(2n+2)(2n+1)}{(3n+3)(3n+2)(3n+1)}$$

$$\longrightarrow \frac{4}{27},$$

as $n \to \infty$. Using the power series ratio test is quicker than using the usual ratio test since we leave out factors of $|z|$ and we don't have to solve for $|z|$ at the end.

Example. Here is an example from the Math 112 notes by Ray Mayer (Math 112 Notes). Consider the power series

$$f(z) = \sum_{n=0}^{\infty} \frac{z^{n^2}}{n^2}.$$

What is its radius of convergence? The ratio test and power series ratio test do not directly apply to $f(z)$ since its k-th coefficient is 0 unless $k = n^2$ for some integer n (and, hence, we cannot divide by it). In this case, define $a_n := \frac{z^{n^2}}{n^2}$, and then

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} \frac{z^{n^2}}{n^2}.$$

We can then apply the ordinary ratio test to $\sum_{n=0}^{\infty} a_n$, which is a series of nonzero terms:

$$\frac{|a_{n+1}|}{|a_n|} = \left|\frac{z^{(n+1)^2}}{(n+1)^2}\right| \Big/ \left|\frac{z^{n^2}}{n^2}\right| = \frac{n^2}{(n+1)^2}\frac{|z|^{n^2+2n+1}}{|z|^{n^2}} = \frac{n^2}{(n+1)^2}|z|^{2n+1}.$$

We have

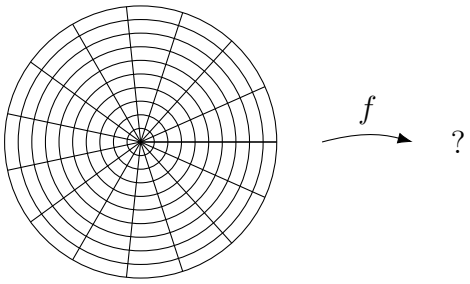
$$\lim_{n\to\infty} \frac{n^2}{(n+1)^2}|z|^{2n+1} = \begin{cases} 0 & \text{if } |z| < 1 \\ \infty & \text{if } |z| > 1. \end{cases}$$

It follows that the radius of convergence for the original series is $R = 1$. When $|z| = 1$, the series converges absolutely since $\sum_{n=0}^{\infty} \frac{1}{n^2}$ converges by the p-series test.

Let $D = \{z \in \mathbb{C} : |z| \leq 1\}$ be the closed unit disk, and consider the function

$$f \colon D \to \mathbb{C}$$
$$z \mapsto \sum_{n=0}^{\infty} \frac{z^{n^2}}{n^2}.$$

To picture f draw the following picture in the plane, centered at the origin, and look at its image after applying f:



The image is pictured below:

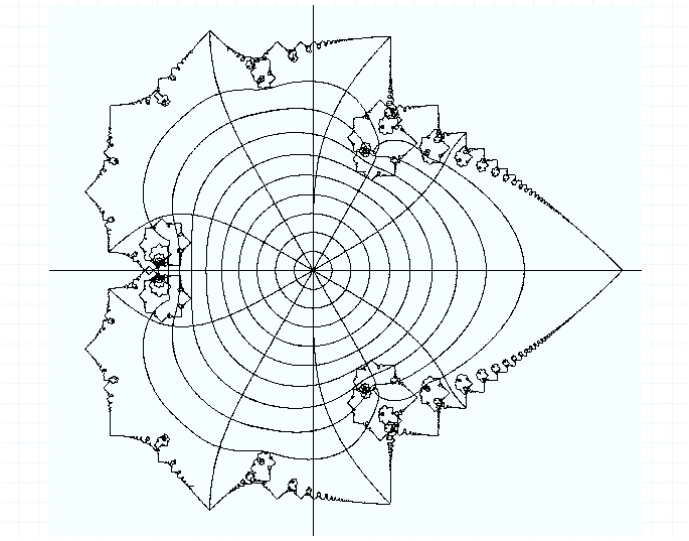


Image of concentric circles and radial lines under $f(z) = \sum_{n=0}^{\infty} \frac{z^{n^2}}{n^2}$.

The fractal-like boundary is the image of the boundary of the disc. We see that when z is small, the function looks almost like the identity function—sending circles in the domain to near-circles in the codomain. As z approaches the boundary of the disc, f gets more and more "confused".

Week 12, Wednesday: Power series II

Power series II

(Supplemental reading: Sections 9.3 and 9.4 in Swanson.)

Our goal is to prove the following:

Theorem (Existence of radius of convergence.) Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a real or complex power series. Then one of the following holds:

(a) $f(z)$ converges only when $z = 0$.

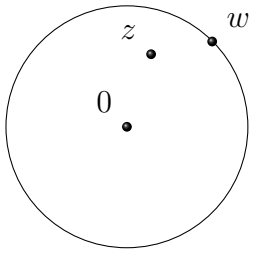
(b) $f(z)$ converges for all $z \in \mathbb{C}$.

(c) There exists a positive real number R such that $f(z)$ converges absolutely for $|z| < R$ and diverges for $|z| > R$.

Definition. The R in the above theorem is the *radius of convergence* of $f(z)$. We define $R = 0$ if $f(z)$ converges only when $z = 0$, and we take $R := \infty$ if $f(z)$ converges for all $z \in \mathbb{C}$.

We will need the following lemma:

Lemma. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a complex power series and suppose that $f(w)$ converges for some $w \in \mathbb{C} \setminus \{0\}$. Then $f(z)$ converges absolutely for all $z \in \mathbb{C}$ such that $|z| < |w|$:



Proof. Recall that if a series converges then the limit of its sequence of terms must go to zero and that convergent sequences are bounded. Therefore,

$$\sum_{n=0}^{\infty} a_n w^n \text{ convergent} \quad \Rightarrow \quad \lim_{n\to\infty} a_n w^n = 0$$

$$\Rightarrow \quad \{a_n w^n\} \text{ bounded}$$

$$\Rightarrow \quad \exists M \in \mathbb{R}_{\geq 0} \text{ such that } |a_n w^n| \leq M \text{ for all } n.$$

Now suppose that $|z| < |w|$. We will use the comparison test to show that $\sum_{n=0}^{\infty} a_n z^n$ is absolutely convergent. First note that

$$0 \leq |a_n z^n| = \left| a_n w^n \left(\frac{z}{w}\right)^n \right| = |a_n w^n| \left|\frac{z}{w}\right| \leq M \left|\frac{z}{w}\right|^n. \tag{32.1}$$

Define

$$r := \left|\frac{z}{w}\right|.$$

Then $|r| < 1$, so the following geometric series converges:

$$\sum_{n=0}^{\infty} M r^n = M \frac{1}{1-r}.$$

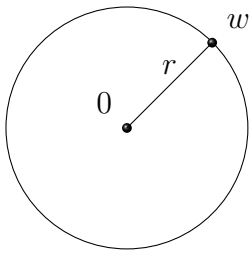
Hence, by (32.1) and the comparison theorem, $\sum_{n=0}^{\infty} a_n z^n$ is absolutely convergent. □

Corollary. If $\sum_{n=0}^{\infty} a_n w^n$ diverges and $|z| > |w|$, then $\sum_{n=0}^{\infty} a_n z^n$ diverges.

Proof. If $\sum_{n=0}^{\infty} a_n z^n$ converged, then by the lemma, with the letters z and w swapped, $\sum_{n=0}^{\infty} a_n w^n$ would converge, too. □

Proof of the main theorem. Suppose that (a) and (b) do not hold. Then there exist various points $w \neq 0$ at which the series converges. We first collect the lengths of these w in a set:

$$S := \{r \in \mathbb{R} : \exists w \in \mathbb{C} \text{ s.t. } |w| = r \text{ and } f(w) = \textstyle\sum_{n=0}^{\infty} a_n w^n \text{ converges}\} :$$



Then S is a set of real numbers, and we'd like to show that S has a supremum. First, since $f(0)$ converges, $|0| = 0 \in S$. So S is nonempty. Next, we claim that S is bounded above. To see this, note that since (b) does not hold, there exists some $\tilde{w} \in \mathbb{C}$ such that $f(w)$ diverges. By the lemma, it follows that $f(z)$ diverges for all z such that $|z| > |\tilde{w}|$. Thus, S is bounded above by $|\tilde{w}|$. Therefore, by completeness of the real numbers, $\sup(S)$ exists. Define $R := \sup(S)$. Our next goal is to show that R is the radius of convergence for f.

On the one hand, if $|z| < R$, then $|z|$ is strictly less than the least upper bound for S. Hence, $|z|$ is not an upper bound for S. This means there exists $r \in S$ such that $|z| < r$. By definition of S, we have $r = |w|$ for some $w \in \mathbb{C}$ such that $f(w)$ converges. The lemma then implies that $\sum_{n=0}^{\infty} |a_n z^n|$ converges, i.e., $f(z)$ converges absolutely.

On the other hand, if $|z| > R$, then $|z|$ is greater than the least upper bound for S, which means that $|z| \notin S$, which in turn means that $f(z)$ does not converge. □

Power series have a remarkable property with respect to differentiation. We state the result but do not include a proof:

Theorem (Differentiation of power series). Suppose that $f(z) = \sum_{n=0}^{\infty} a_n z^n$ has radius of convergence R. Then f is differentiable for all z in the open ball of radius R centered at the origin

$$B(0; R) := \{w \in \mathbb{C} : |w| < R\},$$

and

$$f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}.$$

Further, the radius of convergence for $f'(z)$ is R, the same as for f.

If f' has radius of convergence R, just like f, then we can apply the Theorem to f', to conclude that f'' also has radius of convergence R. We then apply the Theorem to f'' to conclude f''' has radius of convergence R, and so on.

Corollary. Inside its radius of convergence, a power series is infinitely differentiable.

Remark. Differentiability of a function is a measure of its *smoothness*. For instance, the graph of a differentiable function will not be pointy (as is, for example, that of the function $x \mapsto |x|$). If the second derivative exists, then the function is even more smooth. Thinking of the derivative as speed, this would mean that the speed does not change abruptly. If the third derivative exists, then the acceleration does not change abruptly, etc. By this measure, the smoothest functions are those that are infinitely differentiable.

Example. Most real functions, even differentiable ones, are not infinitely differentiable. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^{\frac{4}{3}}$. It is differentiable with derivative $f'(x) = \frac{4}{3} x^{\frac{1}{3}}$. However, $f''(0)$ does not exist. One conclusion we can draw from this is that f cannot be written as a power series with some positive radius of convergence.

Example. Consider the power series

$$f(z) = \sum_{n=0}^{\infty} z^n.$$

By our knowledge of geometric series (or by a quick calculation using the ratio test), we know that f has radius of convergence $R = 1$. So by the above theorem,

$$f'(z) = \sum_{n=0}^{\infty} nz^{n-1}$$

has radius of convergence 1. On the other hand, since f is a geometric series,

$$f(z) = \frac{1}{1-z}$$

for $|z| < 1$. Taking the derivative of both sides of this equation, we get

$$\sum_{n=0}^{\infty} nz^{n-1} = \frac{1}{(1-z)^2}$$

for $|z| < 1$. For instance, letting $z = 1/2$, we get

$$1 + \frac{2}{2} + \frac{3}{4} + \frac{4}{8} + \frac{5}{16} + \cdots = \frac{1}{(1-1/2)^2} = 4.$$

So far, we have been considering power series "centered at 0". The balls in which they converge are centered at the origin. To provide more versatility, we make the following

Definition. Let $\{a_n\}$ be a sequence of complex numbers, and let $c \in \mathbb{C}$. The series

$$f(z) = \sum_{n=0}^{\infty} a_n(z-c)^n$$

is a (complex) *power series centered at c.*

Remark. All of our previous results now apply to this more general context. A power series centered at c will have a radius of convergence R, and the series will be infinitely differentiable in the ball of radius R centered at c. To compute R, one may use the power series ratio test without modification, or one may use the ordinary ratio test, substituting $|z-c|$ for $|z|$.

Example. Consider the power series (centered at 3):

$$f(z) = \sum_{n=0}^{\infty} n^2(z-3)^n.$$

Its radius of convergence may be computed using the power series ratio test:

$$R = \lim_{n\to\infty} \frac{n^2}{(n+1)^2} = 1.$$

Alternatively, we can use the ordinary ratio test:

$$\lim_{n\to\infty} \frac{(n+1)^2|z-3|^{n+1}}{n^2|z-3|^n} = \lim_{n\to\infty} \frac{(n+1)^2}{n^2}|z-3| = |z-3|.$$

We have $|z-3| < 1$ if and only if z is in the ball of radius 1 about the point 3.

Week 12, Friday: Taylor series I

(Supplemental reading: Sections 6.5 and 9.7 in Swanson.)

Taylor series I

The purpose of Taylor series is to approximate functions by polynomials.

Suppose that $f\colon D \to \mathbb{C}$ for some open set $D \subseteq \mathbb{C}$, and let $a \in D$. Suppose that the k-th derivative $f^{(k)}(a)$ exists for all $k \geq 0$.

Recall that by definition,

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Therefore, if h is small, we have

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}.$$

Solving for $f(a+h)$ gives

$$f(a+h) \approx f(a) + f'(a)h.$$

for h close to 0. Finally, define $z = a + h$ and substitute to get

$$f(z) \approx f(a) + f'(a)(z-a)$$

for z close to a. Geometrically, we have approximated f near z using a linear function (which is the equation for its tangent line). We call this a *first-order* approximation of f. The idea behind a Taylor polynomial or Taylor series is to get higher-order and more accurate approximations of f.

Here is a heuristic that is useful in understanding where Taylor series come from. Suppose you know that $f(z)$ is given by a power series

$$f(z) = a_0 + a_1(z-a) + a_2(z-a)^2 + a_3(z-a)^3 + a_4(z-a)^4 + a_5(z-a)^5 + \cdots$$

Then it follows that

$$f'(z) = a_1 + 2a_2(z-a) + 3a_3(z-a)^2 + 4a_4(z-a)^3 + 5a_5(z-a)^4 + \cdots$$

$$\begin{aligned}
f''(z) &= 2\cdot 1\, a_2 + 3\cdot 2\, a_3(z-a) + 4\cdot 3\, a_4(z-a)^2 + 5\cdot 4\, a_5(z-a)^3 + \cdots \\
f^{(3)}(z) &= 3\cdot 2\cdot 1\, a_3 + 4\cdot 3\cdot 2\, a_4(z-a) + 5\cdot 4\cdot 3\, a_5(z-a)^2 + \cdots \\
f^{(4)}(z) &= 4\cdot 3\cdot 2\cdot 1\, a_4 + 5\cdot 4\cdot 3\cdot 2\, a_5(z-a) + \cdots \\
&\vdots
\end{aligned}$$

Next, evaluate these derivatives at $z = a$, at which point, all terms of the form $(z-a)^k$ vanish, and solve for the a_i:

$$\begin{aligned}
f'(a) = a_1 \quad &\Rightarrow \quad a_1 = f'(a) \\
f''(a) = 2\cdot 1\, a_2 \quad &\Rightarrow \quad a_2 = \frac{f''(a)}{2!} \\
f^{(3)}(a) = 3\cdot 2\cdot 1\, a_3 \quad &\Rightarrow \quad a_3 = \frac{f^{(3)}(a)}{3!} \\
f^{(3)}(a) = 3\cdot 2\cdot 1\, a_3 \quad &\Rightarrow \quad a_3 = \frac{f^{(3)}(a)}{3!} \\
f^{(4)}(a) = 4\cdot 3\cdot 2\cdot 1\, a_4 \quad &\Rightarrow \quad a_3 = \frac{f^{(4)}(a)}{4!} \\
&\vdots
\end{aligned}$$

Thus, if f is a power series, we can determine the coefficients:

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n.$$

Note: By definition, $f^{(0)}(a) = f(a)$, and $f^{(n)}$ is the n-th derivative for $n \geq 1$.

Definition. The *k-th order Taylor polynomial* for f centered at a is

$$\sum_{n=0}^{k} \frac{f^{(n)}(a)}{n!}(z-a)^n,$$

and the *Taylor series* for f centered at a is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n.$$

Example. The first-order Taylor polynomial of f is

$$f(a) + f'(a)(z-a),$$

which is the approximation we derived earlier using the definition of the derivative of f.

Example. Consider the polynomial

$$f(z) = 3 - 2z + 3z^2 + z^3$$

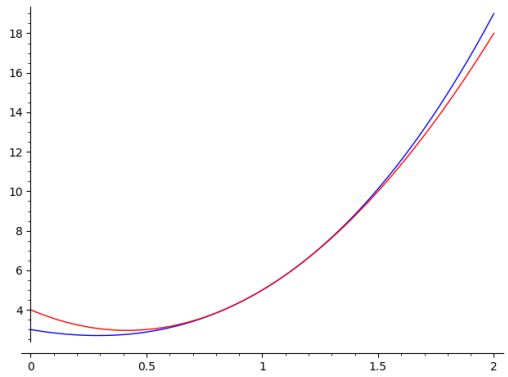
To find the Taylor series centered at $a = 1$, we compute the relevant derivatives:

$$f'(z) = -2 + 6z + 3z^2, \quad f''(z) = 6 + 6z, \quad f^{(3)}(z) = 6,$$

and all higher-order derivatives are 0. The second-order Taylor polynomial of f centered at $a = 1$ is

$$\begin{aligned} T_2(z) &:= f(1) + f'(1)(z-1) + \frac{f''(1)}{2!}(z-1)^2 \\ &= 5 + 7(z-1) + 6(z-1)^2. \end{aligned}$$

Here is a plot of f and T_2 near $z = 1$ (T_2 is in red, and note that the two axes have different scales):



Close to $z = 1$, the second-order Taylor polynomial is a great approximation for f. The third-order Taylor polynomial is

$$\begin{aligned} T_3(z) &:= f(1) + f'(1)(z-1) + \frac{f''(1)}{2!}(z-1)^2 + \frac{f^{(3)}(1)}{3!}(z-1)^3 \\ &= 5 + 7(z-1) + 6(z-1)^2 + (z-1)^3. \end{aligned}$$

Since all derivatives of f of order 4 or higher are 0, the fourth- and higher-order Taylor polynomials are all equal to this third-order Taylor polynomial (as is the Taylor series). In fact, this third-order Taylor polynomial is as good an approximation to f as you could ever want: you can check for yourself that if you multiply out T_3, you actually

get f, i.e., $f(z) = T_3(z)$ for all z. This is what you'd expect from the best third-order approximation of a polynomial of degree 3.

Proposition. Let f be a polynomial of degree d, and let T_k denote the Taylor polynomial for f of order k centered at any point. Then $T_k(z) = f(z)$ for all z for all $k \geq d$.

Example. Let $f(z) = \cos(z)$. The derivatives of f are

$$f'(z) = -\sin(z), \quad f''(z) = -\cos(z), \quad f^{(3)}(z) = \sin(z), \quad f^{(4)}(z) = \cos(z), \ldots$$

From this point on, the higher-order derivative cycle among those we've just calculated. Let's compute the Taylor series at $z = 0$. We have

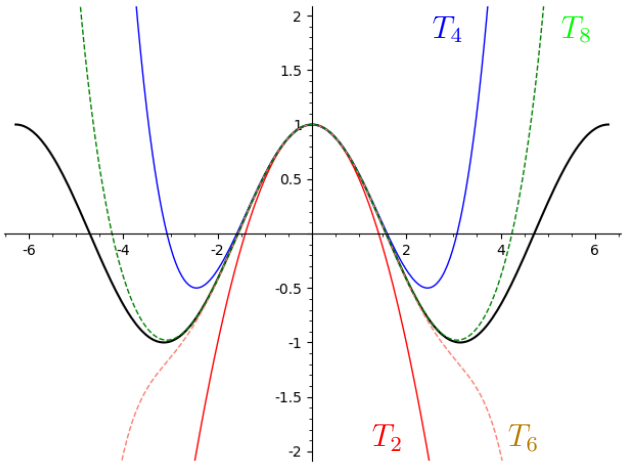
$$f(0) = 1,\ f'(0) = 0,\ f''(0) = -1,\ f^{(3)}(0) = 0,\ f^{(4)}(0) = f(0) = 1,$$

and these values keep cycling. This means that the Taylor series for $\cos(z)$ is

$$T := \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$$

$$= 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \ldots$$

Here is a plot of $f(z) = \cos(z)$ (in black), $T_2(z)$ (in red), $T_4(z)$ (in blue), $T_6(z)$ (salmon and dashed), and $T_8(z)$ (green and dashed):



You should notice that the approximations near $z = 0$ get successively better.

The following theorems help to express that fact that Taylor polynomials and Taylor series are good approximations for functions.

Theorem (Taylor's theorem over $\mathbb{C}$). Let $A \subseteq \mathbb{C}$ and let $f\colon A \to \mathbb{C}$. Suppose that $B(a;r) \subseteq A$ and that f has derivatives of orders $1, \ldots, n$ on $B(a;r)$ where $n \geq 1$. Let $T_{n,a}$ be the Taylor polynomial of order n for f centered at a. Then for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $z \in B(a;\delta)$, then

$$|f(z) - T_{n,a}(z)| < \varepsilon.$$

Theorem. (Taylor's theorem over $\mathbb{R}$) Let $A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$. Suppose that A contains an open interval containing the closed interval $[u,v]$. Also suppose that all the derivatives up to order n exist and are continuous on $[u,v]$ and the $(n+1)$-th derivative exists on (u,v). Let $a \in A$, and let $T_{n,a}$ be the Taylor polynomial of order n for f centered at a. Then for all $x \in [u,v]$, there exists a number c strictly between x and a such that

$$f(x) = T_{n,a}(x) + \frac{1}{(n+1)!} f^{(n+1)}(c)(x-a)^{n+1}.$$

How to interpret this last theorem. The term

$$\frac{1}{(n+1)!} f^{(n+1)}(c)(x-a)^{n+1} = f(x) - T_{n,a}(x)$$

is the error in approximating f by its n-th order Taylor polynomial In particular, if $|f^{n+1}|$ is bounded by M on the interval $[u,v]$, then the error is bounded:

$$\left| \frac{1}{(n+1)!} f^{(n+1)}(c)(x-a)^{n+1} \right| \leq \frac{M}{(n+1)!} |x-a|^{n+1}.$$

Note that if x is close to a then the factor $\frac{1}{(n+1)!} \cdot (x-a)^{n+1}$ will be very small. (For instance, for any constant α, we have $\lim_{n\to\infty} \frac{\alpha^{n+1}}{(n+1)!} = 0$.)

Week 13, Monday: Taylor series II

(Supplemental reading: Sections 6.5 and 9.7–9 in Swanson.)

Taylor series II

Convergence of Taylor series. Given a function $f(z)$ whose derivatives of all orders exist at a point a, it makes sense to compute the Taylor series of $f(z)$ centered at a:

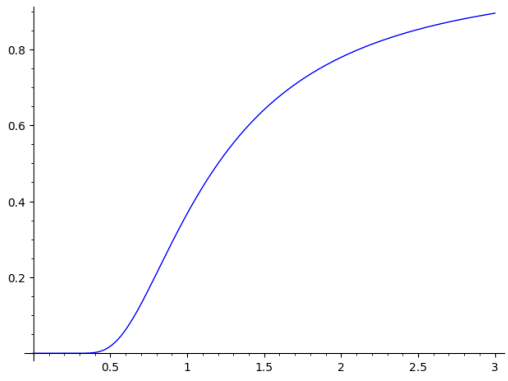
$$T(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n.$$

Then $T(z)$ is a power series in $z-a$, and so it has a radius of convergence R. The question arises: at points z inside the disk of convergence (centered at a), i.e., where $|z-a| < R$, is it true that $f(z) = T(z)$. Over the complex numbers, the answer is "yes", but over the real numbers things are more complicated:

Example. Consider the function

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ e^{-1/x^2} & \text{if } x > 0. \end{cases}$$

Its graph looks like this:



It turns out that the derivatives of f of all orders exist at 0, and they are all 0. Hence, its Taylor polynomial is 0. However, $f(x) > 0$ for all $x > 0$. So f is not equal to its Taylor polynomial centered at 0 on any interval about 0.

In the last lecture, we saw that over the reals, the error term in approximating the function f with by its n-th Taylor polynomial is

$$|f(x) - T_n(x)| = \left|\frac{f^{(n+1)}(c)}{(n+1)!}\right| |x-a|^{n+1} \tag{34.1}$$

where c is a point between x and a. Suppose that $f^{(n+1)}$ is bounded in some interval containing x and a by a constant, independent of n. In other words, suppose there is a constant B such that for all n,

$$\left|f^{(n+1)}(c)\right| \leq B$$

for all c in some interval containing x and a. In that case, we have

$$0 \leq \lim_{n\to\infty} |f(x) - T_n(x)| = \lim_{n\to\infty} \left|\frac{f^{(n+1)}(c)}{(n+1)!}\right| |x-a|^{n+1} \leq \lim_{n\to\infty} \frac{B}{(n+1)!}|x-a|^{n+1} = 0.^{1}$$

It is easy to check that $x \mapsto |x|$ is a continuous function and, hence, commutes with limits. This means that

$$0 = \lim_{n\to\infty} |f(x) - T_n(x)| = \left|\lim_{n\to\infty} f(x) - T_n(x)\right| = |f(x) - T(x)|.$$

Therefore, in this case, the real function f will converge to its Taylor series.

Example. The real Taylor series for $\cos(x)$ centered at 0 is

$$T(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots.$$

An easy application of the ratio test shows that its radius of convergence is $R = \infty$. Then at any point c we have that $\cos^{(n)}(c)$ is either $\pm\cos(c)$ or $\pm\sin(c)$, which are all bounded in absolute value by $B = 1$. Therefore, $\cos(x) = T(x)$ for all $x \in \mathbb{R}$.

Example. For a more problematic example, consider the Taylor series for $f(x) = \ln(x)$ centered at $x = 1$. We have

$$f'(x) = \frac{1}{x},\ f''(x) = -\frac{1}{x^2},\ f^{(3)}(x) = \frac{2\cdot 1}{x^3},\ f^{(4)}(x) = -\frac{3\cdot 2\cdot 1}{x^4}, f^{(5)}(x) = \frac{4\cdot 3\cdot 2\cdot 1}{x^5},$$

and so on: $f^{(n)}(x) = (-1)^{n-1}\frac{(n-1)!}{x^n}$. In general, for $n > 0$, the n-th coefficient for the Taylor series centered at 1 is

$$\frac{f^{(n)}(1)}{n!} = (-1)^{n-1}\frac{(n-1)!}{n!} = \frac{(-1)^{n-1}}{n}.$$

[1] For any real or complex number α, we have $\lim_{n\to\infty} \frac{\alpha^n}{n!} = 0$. A roundabout way to see this is to note that $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ has radius of convergence $R = \infty$. Hence, for all z, the sequence of terms of the series converges to 0.

Thus, the Taylor series for f centered at 1 is

$$\begin{aligned} T(x) &:= \sum_{n=0}^{\infty} \frac{f^{(n)}(1)}{n!}(x-1)^n \\ &= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(n-1)!}{n!}(x-1)^n \\ &= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}(x-1)^n \\ &= (x-1) - \frac{(x-1)}{2} + \frac{(x-1)^2}{3} - \frac{(x-1)^4}{4} + \cdots . \end{aligned}$$

We can use the power series ratio test find the radius of convergence:

$$\lim_{n\to\infty} \left| \frac{(-1)^{n-1}}{n} \right| \Big/ \left| \frac{(-1)^n}{n+1} \right| = \lim_{n\to\infty} \frac{n+1}{n} = 1.$$

Hence, the series converges (absolutely) for $|x-1| < 1$ and diverges for $|x-1| > 1$. Thus, in the open interval of radius 1 centered at 1, i.e., in $(0, 2)$, the series converges absolutely. What happens on the boundary? At $x = 2$, we get the alternating harmonic series

$$T(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots,$$

which converges by the alternating series test. At $x = 0$, we have

$$T(0) = -1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots,$$

which diverges since its partial sums differ from those of the harmonic series by a constant.

Does the Taylor series for $\ln(x)$ converge to $\ln(x)$ on $(0, 2)$? It turns out that it does, but the argument depends on a different formulation for the error term when approximating a function by a Taylor polynomial. There is a further interesting question: what happens when $x = 2$, on the boundary of the interval of convergence? We have seen above that $T(1)$ is the alternating harmonic series. Is it true that $\ln(2) = T(2)$ The answer is, again, yes. The details appear in the appendix at the end of this lecture for those who are interested.

The complex exponential function

The usual exponential, cosine, and sine functions generalize to functions on the whole complex plane. Moreover, the generalization leads to a hidden relation among all three functions (Euler's formula).

Definition. For $z \in \mathbb{C}$ define the power series

$$E(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots$$

$$C(z) := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots$$

$$S(z) := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots .$$

Exercise. Here are a few straightforward exercises concerning these functions:

1. Use the ratio test to show that $E(z)$, $C(z)$, and $S(z)$ converge for all $z \in \mathbb{C}$.
2. Show that $C(-z) = C(z)$ and $S(-z) = -S(z)$.
3. Show that
$$E'(z) = E(z), \quad C'(z) = -S(z), \quad \text{and} \quad S'(z) = C(z).$$
4. Show that $C^2(z) + S^2(z) = 1$. (Hint: use derivatives.)
5. For $z \in \mathbb{R}$, a real number, show that these functions are the Taylor series for e^z, $\cos(z)$, and $\sin(z)$.

Proposition. For $x \in \mathbb{R}$,

$$E(x) = e^x, \quad C(x) = \cos(x), \quad \text{and} \quad S(x) = \sin(x).$$

Proof. It suffices to show that e^x, $\cos(x)$, and $\sin(x)$ equal their Taylor series. That follows from the error term for the approximation of a function by a Taylor polynomial which we have discussed previously. For instance, let $f(x) = e^x$, and let $T_n(x)$ be the n-th order Taylor polynomial for e^x. Given $x \in \mathbb{R}$, there exists c between x and 0 such that

$$e^x - T_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1} = \frac{e^c x^{n+1}}{(n+1)!}.$$

Therefore,

$$e^x - T(x) = \lim_{n\to\infty} (e^x - T_n(x)) = e^c \lim_{n\to\infty} \frac{x^{n+1}}{(n+1)!} = 0.^{2}$$

We gave the argument for $\cos(x)$ in an earlier example, and the example for $\sin(x)$ is similar. □

[2] See the previous footnote for an proof that $\lim_{n\to\infty} \frac{x^n}{n!} = 0$.

Thus, for instance, we know that $E(\sqrt{2}) = e^{\sqrt{2}}$. What can we say about something like e^{2+3i}? What about $\cos(2+3i)$? It turns out that up to this point, we have no definition of what it means take the exponential, cosine, or sine of a complex number. So we are free to decide what these mean, as we do in the following definition:

Definition. For $z \in \mathbb{C}$ let

$$\exp(z) = E(z), \quad \cos(z) = C(z), \quad \sin(z) = S(z).$$

Proposition (Euler's formula). For all $z \in \mathbb{C}$, we have

$$e^{iz} = \cos(z) + i\sin(z).$$

Proof. We have

$$\begin{aligned}
\exp(iz) &= 1 + (iz) + \frac{(iz)^2}{2!} + \frac{(iz)^3}{3!} + \frac{(iz)^4}{4!} + \frac{(iz)^5}{5!} + \frac{(iz)^6}{6!} + \dots \\
&= 1 + (iz) + \frac{i^2z^2}{2!} + \frac{i^3z^3}{3!} + \frac{i^4z^4}{4!} + \frac{i^5z^5}{5!} + \frac{i^6z^6}{6!} + \dots \\
&= 1 + iz - \frac{z^2}{2!} - i\frac{z^3}{3!} + \frac{z^4}{4!} + i\frac{z^5}{5!} - \frac{z^6}{6!} + \dots \\
&= \left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots\right) + i\left(z - \frac{z^3}{3!} + \frac{z^5}{5!} + \dots\right) \\
&= \cos(z) + i\sin(z).
\end{aligned}$$

□

Thus, for instance, we may parametrize the unit circle centered at the origin by

$$\begin{aligned}
\mathbb{R} &\to \mathbb{C} \\
t &\mapsto e^{it}.
\end{aligned}$$

Corollary. For $z \in \mathbb{C}$

$$\cos(z) = \frac{\exp(iz) + \exp(-iz)}{2} \quad \text{and} \quad \sin(z) = \frac{\exp(iz) - \exp(-iz)}{2i}.$$

Proof. Exercise. □

One thing that is not immediately clear from our definition of e^z is that it obeys the usual exponent laws for all complex numbers! We show that now.

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

1. $\exp(-z) = \dfrac{1}{\exp(z)}$.

2. $\exp(w+z) = \exp(w)\exp(z)$

Proof. For part 1, fix $w \in \mathbb{C}$ and define the function $f(z) = \exp(w+z)\exp(-z)$. Apply the product and chain rules for differentiation with respect to z to find

$$\begin{aligned} f'(z) &= \exp'(w+z)\exp(-z) + \exp(w+z)\exp'(-z) \\ &= \exp(w+z)\exp(-z) - \exp(w+z)\exp(-z) \\ &= 0. \end{aligned}$$

It follows that $f(z) = c$ for some constant c. To find the constant, we just need to evaluate f at any point, say $z = 0$. We find $c = f(0) = \exp(w+0)\exp(0) = \exp(w)$. Hence, $f(z) = f(w)$ for all $z \in \mathbb{C}$:

$$\exp(z+w)\exp(-z) = \exp(w). \tag{34.2}$$

Equation (34.2) holds for all $w, z \in \mathbb{C}$. In particular, setting $w = 0$, we find

$$\exp(z)\exp(-z) = \exp(0) = 1.$$

This proves part 1. Part 2 then immediately follows from 1 and (34.2). □

Remark. Given $z = x + iy \in \mathbb{C}$ where $x, y \in \mathbb{R}$, we have

$$e^z = e^{x+iy} = e^x e^{iy} = e^x(\cos(y) + i\sin(y)).$$

Note that on the right, we have the polar form for the complex number e^z. Hence,

$$|e^z| = e^x \quad \text{and} \quad \arg(e^z) = y.$$

In general, exponential notation arises naturally when writing a complex number in polar form:

$$w = r(\cos(\theta) + \sin(\theta)) = re^{i\theta}.$$

The geometry of multiplication of complex numbers—that lengths multiply and angles add—takes the form

$$(re^{i\theta})(se^{i\psi}) = rse^{i\theta+i\psi} = rse^{i(\theta+\psi)}.$$

Appendix

We would like to show that $\ln(x)$ converges to its Taylor series centered at 1 on the interval $(0, 2]$. Earlier, we calculated the derivatives

$$\ln^{(n+1)}(c) = (-1)^n \frac{n!}{c^{n+1}}$$

and found the Taylor series to be

$$T(x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(x-1)^n}{n} = (x-1) - \frac{(x-1)}{2} + \frac{(x-1)^2}{3} - \frac{(x-1)^4}{4} + \cdots,$$

which converges on $(0, 2]$.

To show $\ln(x) = T(x)$ on this interval takes several steps. First note that the case $x = 1$ is trivial: $\ln(1) = 0 = T(1)$. The case of the endpoint $x = 2$ will be handled at the end. We first consider the two cases $0 < x < 1$ and $1 < x < 2$ separately. For the latter, we will need to use an alternative form of the error term.

Let $T_n(x)$ denote the n-th Taylor polynomial for $\ln(x)$. For the case where $0 < x < 1$, we use Equation (34.1) to get

$$|\ln(x) - T_n(x)| = \frac{1}{(n+1)!} \frac{n!}{c^{n+1}} |x-1|^{n+1} = \frac{1}{n+1} \cdot \left(\frac{1-x}{c}\right)^{n+1}.$$

where $0 < x < c < 1$. In this case, $1 - x < c$, and therefore

$$\frac{1}{n+1} \cdot \left(\frac{1-x}{c}\right)^{n+1} < \frac{1}{n+1} \longrightarrow 0,$$

as $n \to \infty$. Hence, we see that $\ln(x) = T(x)$ on $(0, 1)$.

For the next case, we need a different form for the remainder.

Theorem. Let $A \subseteq \mathbb{R}$ and let $f \colon A \to \mathbb{R}$. Suppose that A contains an open interval containing the closed interval $[u, v]$. Also suppose that all the derivatives up to order n exist and are continuous on $[u, v]$ and the $(n+1)$-th derivative exists on (u, v). Let $a \in A$, and let $T_{n,a}$ be the Taylor polynomial of order n for f centered at a. Then for all $x \in [u, v]$, there exists a number c strictly between x and a such that

$$f(x) = T_{n,a}(x) + \frac{1}{n!} f^{(n+1)}(c)(x-a)(x-c)^n. \tag{34.3}$$

Proof. See Swanson, Theorem 6.5.5. □

We apply this theorem to $\ln(x)$ in the case $1 < x < 2$ to see

$$|\ln(x) - T_n(x)| = \frac{1}{n!} \frac{n!}{c^{n+1}} (x-1)(x-c)^n = \frac{(x-1)(x-c)^n}{c^{n+1}}$$

for some c such that $1 < c < x < 2$. Since $1 < c$ and $0 < x - c < x - 1 < 1$, it follows that

$$0 \leq \frac{(x-1)(x-c)^n}{c^{n+1}} \leq (x-1)^{n+1} \longrightarrow 0$$

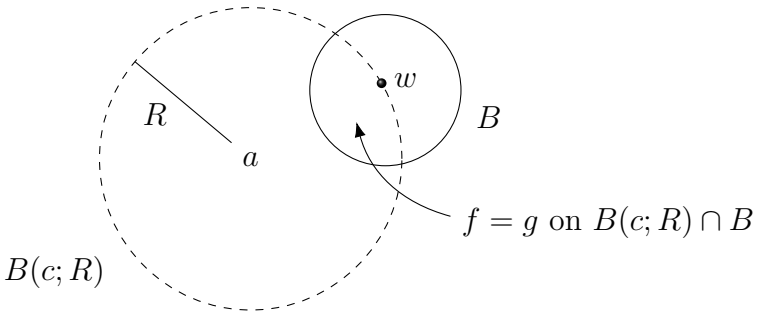
as $n \to \infty$.

Having show that $\ln(x)$ equals its Taylor series on $(0, 2)$, we now consider the point $x = 2$. For that, we apply the following

Theorem. Let $f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$ be a real or complex power series with radius of convergence $R > 0$. Let c be a point on the boundary of the ball of convergence, i.e., such that $|c| = R$. Let B be any open ball centered at c, and let g be a continuous function with domain B such that $f(z) = g(z)$ on $B(c; R) \cap B$. Then, $f(c) = g(c)$.

Proof. See Swanson, Theorem 9.6.2. □

Picture for the case of complex series:



Now apply this theorem with $g(x) = \ln(x)$ and $f(x) = T(x)$ to conclude

$$\ln(2) = T(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

Week 13, Wednesday: Two theorems from calculus

(Supplemental reading: Sections 5.2 and 5.3 in Swanson.)

Two theorems from calculus

The extreme value theorem. We present proofs of two basic theorems whose proofs are usually skipped in a first course in calculus: the extreme value theorem and the intermediate value theorem.

The extreme value theorem is of fundamental importance in the theory of optimization since it states conditions under which the existence of a maximal and minimal value for a function is guaranteed.

Theorem (extreme value theorem, EVT). If $f\colon [a,b] \to \mathbb{R}$ is continuous, then f achieves is maximum and minimum on $[a,b]$, i.e., there exists $x_{\max}, x_{\min} \in [a,b]$ such that

$$f(x_{\min}) \leq f(x) \leq f(x_{\max})$$

for all $x \in [a,b]$.

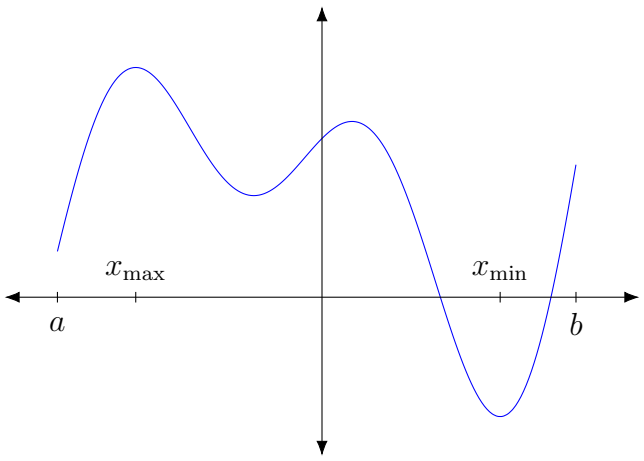


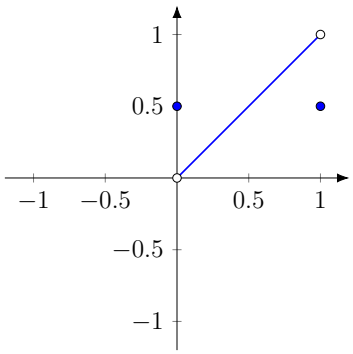
Figure 35.1: A continuous function on a closed interval $[a,b]$ with the points at which it achieves its (global) minimum and maximum labeled.

There are three important hypotheses in the EVT: (i) f is continuous, (ii) the interval is closed, and (iii) the interval is bounded.[1] The following three examples show that if you violate any one of these, the theorem no longer holds.

Example 1. Let $f\colon [0,1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x & \text{if } 0 < x < 1 \\ \frac{1}{2} & \text{if } x = 0 \text{ or } 1. \end{cases}$$

Then f is defined on a closed and bounded interval but is not continuous. It has no maximum or minimum:



Graph of f.

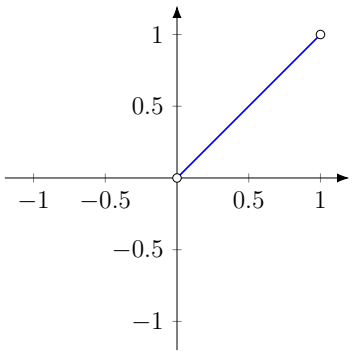
The image of this function is $(0,1)$. The function takes on values arbitrarily close to 0 and 1, but 0 and 1 are never achieved.

Example 2. The identity function

$$\begin{aligned} g\colon (0,1) &\to (0,1) \subset \mathbb{R} \\ x &\mapsto x \end{aligned}$$

is continuous and defined on a bounded interval, but the interval is not closed. The function g has no minimum or maximum:

[1] Note that (ii) does not imply (iii). For instance, the intervals $[0,\infty)$ and $(-\infty,\infty)$ are both closed—their complements are open—but not bounded.



Graph of g.

The function $x \mapsto 1/x$ with domain $(0, 1)$ is a similar example.

Example 3. The identity function

$$\begin{aligned} h\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x \end{aligned}$$

is continuous on the closed but unbounded interval $(\infty, \infty) = \mathbb{R}$. It has no minimum or maximum.

To prove the extreme value theorem we need a couple of preliminary results. First, recall that if $X \subseteq \mathbb{R}$ and $g\colon X \to \mathbb{R}$ is a function, then g is *bounded* if there exists a constant B such that $|g(x)| \leq B$ for all $x \in [a, b]$. Also, g is *continuous* at $c \in [a, b]$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $|x - c| < \delta$ implies $|g(x) - g(c)| < \varepsilon$.

Lemma. If $f\colon [a, b] \to \mathbb{R}$ is continuous, then f is locally bounded, i.e., given $c \in [a, b]$, there exists $\delta > 0$ such that f is bounded on $(c - \delta, c + \delta) \cap [a, b]$.[2]

Proof. Apply the definition of continuity at c with $\varepsilon = 1$. We find a $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon = 1$. Therefore, $f(c) - 1 < f(x) < f(c) + 1$ for $x \in (c - \delta, c + \delta) \cap [a, b]$. (So we could take $B = \max\{|f(c) - 1|, |f(c) + 1|\}$.) □

Proposition. If $f\colon [a, b] \to \mathbb{R}$ is continuous, then it is bounded.

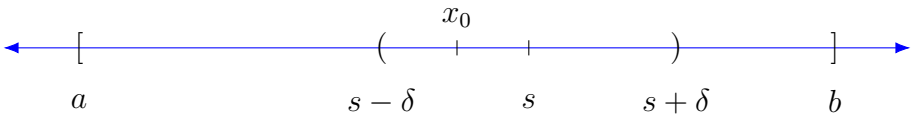
Proof. This proof is a sort of continuous version of induction. Define

$$A = \{x \in [a, b] : f \text{ is bounded on } [a, x]\}.$$

To see that the supremum of A exists, note that A is nonempty since $a \in A$, and A is bounded above by b. Hence, by completeness, $s := \sup(A)$ exists. Our next goal is to show $s = b$. We will prove this by contradiction. Since b is an upper bound for A,

[2] Thus, f is bounded in a "neighborhood" about each point $c \in [a, b]$.

and s is the least upper bound, we know that $s \leq b$. So suppose that $s < b$. By the Lemma, there exists $\delta > 0$ such that f is bounded on $(s-\delta, s+\delta)$ (by taking δ small enough, we may assume that $(s-\delta, s+\delta) \subset [a,b]$):



Since $s - \delta < s$, i.e., $s - \delta$ is less than the least upper bound for A, it follows that $s - \delta$ is not an upper bound for A. Therefore, there exists $x_0 \in A$ such that $s - \delta < x_0$. Since $x_0 \in A$, we have that f is bounded on $[a, x_0]$. But since f is also bounded on $(s-\delta, s+\delta)$, it then follows that it is bounded on

$$[a, x_0] \cup (s-\delta, s+\delta) = [a, s+\delta).$$

In particular, f is bounded on $[a, s+\delta/2]$. So $s + \delta/2 \in A$, contradicting the fact that $s = \sup(A)$. We conclude that, in fact, $s = b$.

Finally, we show that $b \in A$ by using an argument similar to the one we just used. We have established the fact that $b = \sup(A)$. By the Lemma, we know that there exists $\delta > 0$ such that f is bounded on $(b-\delta, b+\delta) \cap [a,b] = (b-\delta, b]$. Since b is the least upper bound for A and $b - \delta$ is less than b, it follows that $b - \delta$ is not an upper bound for A. So there exists $x_1 \in A$ such that $b - \delta < x_1$. Hence, f is bounded on

$$[a, x_1] \cup (b-\delta, b] = [a, b],$$

as required. □

We now know that a continuous function on a closed bounded interval is bounded. This is the key tool in proving the extreme value theorem.

Proof of the extreme value theorem. We just need to prove that $x_{\max}$ always exists. If we know that, then to prove that $x_{\min}$ exists, apply our result to the function $-f$ (which also continuous on $[a,b]$). The maximum we find for $-f$ will be a minimum for f.

Consider the set

$$\mathrm{im}(f) = f([a,b]) = \{f(x) : x \in [a,b]\}.$$

By our Proposition, f is bounded on $[a,b]$, which is equivalent to saying the set $\mathrm{im}(f)$ is a bounded set. Since $\mathrm{im}(f)$ is also nonempty, e.g., $f(a) \in \mathrm{im}(f)$, we can let $s := \sup(\mathrm{im}(f))$. Our proof is finished if we can show that $s \in \mathrm{im}(f)$ for in that case, there exists some $\tilde{x} \in [a,b]$ such that $f(\tilde{x}) = s$, and $f(\tilde{x}) \geq f(x)$ for all $x \in [a,b]$.

We prove $s \in \mathrm{im}(f)$ by contradiction. Suppose it is not, and define

$$g \colon [a,b] \to \mathbb{R}$$

$$x \mapsto \frac{1}{s - f(x)}.$$

Since, by supposition, $s \notin \text{im}(f)$, the function g is well-defined (we never divide by 0) and continuous. By the Proposition, g is therefore bounded. However, we will now show that g can't be bounded. Since $s = \sup(\text{im}(f))$, there are points $s' \in \text{im}(f)$ that are arbitrarily close to s, and writing $s' = f(x')$, we see $g(x') = \frac{1}{s-f(x')} = \frac{1}{s-s'}$ will be arbitrarily large. In detail, given any potential bound $B \in \mathbb{R}$, choose $\varepsilon > 0$ such that $\frac{1}{\varepsilon} > B$. Next, consider $s - \varepsilon$. Since it is less than the least upper bound s for $\text{im}(f)$ and $s \notin \text{im}(f)$, there exists $s' \in \text{im}(f)$ such that $s - \varepsilon < s' < s$. Since $s' \in \text{im}(f)$, there exists $x' \in [a, b]$ such that $f(x') = s'$. We then have

$$g(x') = \frac{1}{s - f(x')} = \frac{1}{s - s'} > \frac{1}{s - (s - \varepsilon)} = \frac{1}{\varepsilon} = B.$$

Hence, g has no upper bound B. □

The intermediate value theorem. The intermediate value theorem is intuitively obvious, but its proof requires familiarity with the completeness property of the real numbers.

Theorem (intermediate value theorem, IVT). Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function such that $f(a) < 0$ and $f(b) > 0$. Then there exists $s \in (a, b)$ such that $f(s) = 0$.

Proof. Define the set

$$A := \{x \in [a, b] : f \text{ is negative at all points in } [a, x]\}.$$

We have $a \in A$ since $f(a) < 0$. Thus, $A \neq \emptyset$. Since $A \subseteq [a, b]$, the set A is bounded above by b. By completeness of $\mathbb{R}$ is follows that $s := \sup(A)$ exists. We are done if we can show that $f(s) = 0$.

We first argue that $s \notin \{a, b\}$. First, since $f(a) < 0$ and f is continuous, there is a $\delta_a > 0$ such that f is negative on the interval $[a, a + \delta_a)$. This shows $s \neq a$. Next, since $f(b) > 0$ and f is continuous, there exists a $\delta_b > 0$ such that f is positive on $(b - \delta_b, b]$. Hence, $s \neq b$. Therefore, $s \in (a, b)$.

Now note that if $f(c) > 0$ for any point $c \in [a, b]$, then c is an upper bound for A. If it were not, then there would an $x \in A$ such that $c < x$. But then, by definition of A, the function f is negative on $[a, x]$ and since $c \in [a, x]$, that would mean $f(c) < 0$. In particular, since s is the least upper bound for A, the function f cannot be positive at any point strictly less than s. Thus, we have shown that

$$[a, s) \subseteq A.$$

For sake of contradiction, suppose that $f(s) < 0$. By continuity of f, there exists $\delta > 0$ such that f is negative on $(s-\delta, s+\delta) \cap [a,b]$. Since $s \in (a,b)$, we can take δ small enough to that $(s-\delta, s+\delta) \cap [a,b] = (s-\delta, s+\delta)$. But then f is negative on

$$[a,s) \cup (s+\delta/2] = [a, s+\delta/2],$$

which shows that $s + \delta/2 \in A$, contradicting the fact that s is an upper bound for A.

Next, for the sake of contradiction, suppose that $f(s) > 0$. By continuity of f it follows that f is positive in a small interval $(s-\eta, s+\eta)$. Since $s \notin \{a,b\}$, we can assume $(s-\eta, s+\eta) \subset A$. Then, for instance, letting $c := s - \eta/2$, we see that $f(c) > 0$. As discussed above, that means that c is an upper bound for A. But that's impossible since $c < s$ and s is the least upper bound for A.

We conclude that $f(s) = 0$, as desired. □

Week 13, Friday: Three theorems by Euler

Three theorems by Euler

Theorem 1. The number e is irrational.

Proof. Euler discovered that $e = \sum_{n=0}^{\infty} \frac{1}{n!}$, and used that to prove that e is irrational. To start, we can at least see that e is not an integer since

$$2 = 1 + \frac{1}{1!} < e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots < 1 + \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots\right) = 3.$$

Now, for the sake of contradiction, suppose that e is rational, and write

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

with $p, q \in \mathbb{Z}_{\geq 1}$. Since e is not an integer, $q > 1$. We have

$$e - \sum_{n=0}^{q} \frac{1}{n!} = \sum_{n=q+1}^{\infty} \frac{1}{n!} \quad \Longrightarrow \quad q!\left(e - \sum_{n=0}^{q} \frac{1}{n!}\right) = \sum_{n=q+1}^{\infty} \frac{q!}{n!}. \tag{36.1}$$

The expression on the left-hand side of (36.1),

$$q!\left(e - \sum_{n=0}^{q} \frac{1}{n!}\right),$$

is an integer since $e = \frac{p}{q}$ and $\frac{q!}{n!} \in \mathbb{Z}$ when $q \geq n$. We conclude that the right-hand side of (36.1) is also an integer:

$$\sum_{n=q+1}^{\infty} \frac{q!}{n!} \in \mathbb{Z}.$$

However,

$$\sum_{n=q+1}^{\infty} \frac{q!}{n!} = \frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \frac{q!}{(q+3)!} + \cdots$$

$$= \frac{1}{q+1} + \frac{1}{(q+2)(q+1)} + \frac{1}{(q+3)(q+2)(q+1)} + \cdots$$

$$< \frac{1}{q+1} + \frac{1}{(q+1)^2} + \frac{1}{(q+1)^3} + \cdots$$

$$= \frac{1}{q+1} \sum_{n=0}^{\infty} \left(\frac{1}{q+1} \right)^n$$

$$= \frac{1}{q+1} \cdot \frac{1}{1 - \frac{1}{(q+1)}} \qquad \text{(geometric series formula)}$$

$$= \frac{1}{q} < 1.$$

Contradicting the fact that this sum is an integer. □

Theorem 2. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

Proof. We first recall some basic algebra. If $P(z)$ is a polynomial of degree n with real or complex coefficients, then it will have n (not necessarily distinct) roots, $r_1, \ldots, r_n$.[1] We can then write

$$P(z) = k(z - r_1) \cdots (z - r_n)$$

for some constant k. If no $r_i = 0$, we can rewrite this expression as

$$P(z) = (-1)^n k r_1 \cdots r_n \left(1 - \frac{z}{r_1}\right) \cdots \left(1 - \frac{z}{r_n}\right).$$

Now, further assume that $P(0) = 1$. We then have

$$1 = P(0) = (-1)^n k r_1 \cdots r_n,$$

and, thus,

$$P(z) = \left(1 - \frac{z}{r_1}\right) \cdots \left(1 - \frac{z}{r_n}\right). \tag{36.2}$$

Consider $\sin(z) = z - \dfrac{z^3}{3!} + \dfrac{z^5}{5!} - \cdots$, and divide by z to define

$$P(z) = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \cdots \tag{36.3}$$

[1] By *roots* we mean $P(r_i) = 0$ for each i.

so that $P(z) = \dfrac{\sin(z)}{z}$ for $z \neq 0$, and at $z = 0$, we have.

$$P(0) = 1 = \lim_{z \to 0} \frac{\sin(z)}{z}.$$

Since $\sin(k\pi)$ is zero for all integers k, it follows that $P(k\pi) = 0$ for all nonzero integers $k = \pm 1, \pm 2, \ldots$ Therefore, skipping several technical details, we emulate (36.2) and write

$$\begin{aligned} P(z) &= \left(1 - \frac{z}{\pi}\right)\left(1 + \frac{z}{\pi}\right)\left(1 - \frac{z}{2\pi}\right)\left(1 + \frac{z}{2\pi}\right)\left(1 - \frac{z}{3\pi}\right)\left(1 + \frac{z}{3\pi}\right)\cdots \\ &= \left(1 - \frac{z^2}{\pi^2}\right)\left(1 - \frac{z^2}{4\pi^2}\right)\left(1 - \frac{z^2}{9\pi^2}\right)\cdots \end{aligned}$$

Going back to the definition (36.3) of P, we have shown that

$$1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \cdots = \left(1 - \frac{z^2}{\pi^2}\right)\left(1 - \frac{z^2}{4\pi^2}\right)\left(1 - \frac{z^2}{9\pi^2}\right)\cdots$$

Comparing the coefficient of z^2 on both sides, we find

$$-\frac{1}{6} = -\sum_{n=1}^{\infty} \frac{1}{n^2\pi},$$

and the result follows. □

Exercise. Show that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$.

What about $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3}$? This number was shown to be irrational in 1978. It is an open question whether there is a rational r such that the sum is equal to $r\pi^3$.

Theorem 3. There are infinitely many prime numbers.

Proof. Let $s \in \mathbb{C}$. Then

$$\begin{aligned} \prod_{p \text{ prime}}^{\infty} \left(\frac{1}{1 - p^{-s}}\right) &= \prod_{p \text{ prime}}^{\infty} \left(1 + p^{-s} + p^{-2s} + p^{-3s} + \cdots\right) \\ &= \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right)\cdots \end{aligned}$$

Patiently multiplying out the final expression, and using the fact that each positive integer can be uniquely factored into primes, you will find that

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}}^{\infty} \left(\frac{1}{1 - p^{-s}} \right)$$

If were only finitely many primes, the product on the right would be finite, and hence we could evaluate it as a complex number when $s = 1$. However,

$$\zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n},$$

which diverges. □

Note: The function $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{z^s}$ is known as the *Riemann zeta function.* Our previous result states that

$$\zeta(2) = \frac{\pi^2}{6}.$$

HANDOUTS

Mathematical writing

Audience

In all of the writing you do for a course based on this text, take your audience to be your fellow classmates. That will determine the amount of detail you need to include—do not skip important details, and do not include details that are not essential.

Sentences!

The most important rule is that your writing should consist solely of complete sentences. A sentence starts with a capital letter and ends with a period. If you have a long calculation, you might want to neatly display it, but it should still be part of a sentence; something like this:

Our result then follows from a calculation:

$$\begin{aligned} \text{blah} &= \text{blah} \\ &= \text{blah} \\ &\vdots \\ &= \text{blah.} \end{aligned}$$

Use of symbols

For better readability, do not start a sentence with a mathematical symbol (i.e., a mathematical symbol does not count as a capital letter):

No: f is injective by not surjective.

Yes: The function f is injective by not surjective.

In informal mathematical writing—the kind you might use on a blackboard or on scratch paper—it is common to use the symbols shown below:

$\Rightarrow$: "implies"
$\Leftrightarrow$: "if and only if"
$\forall$: "for all"
$\exists$: "there exists"

In your formal written work, e.g., homework assignments, do not use these. Instead, write out the words. It is slightly more work for you, but it makes life easier for your reader. Of course, you will need to use symbols in your writing—just do not use those listed in formal writing. On the other hand, here are some symbols from logic we encourage you to never use, even in informal work: "$\wedge$" for "and", "$\vee$" for "or", and "$\sim$" for "not". Again, it is easier on your read to use the word instead of these symbols.

Red herrings

Whenever you a finish a proof by contradiction and now have the ideas in front of you, ask yourself whether a straightforward proof (without contradiction) is at least as clear as the one you have given. If so, make the change. It makes for a less convoluted argument. Apply similar reasoning to any proof where you have replaced the statement of the result by its (logically equivalent) contrapositive (not Q implies P, rather than P implies Q).

More generally, review your proofs to see if you have included irrelevant information.

The "backwards" proof.

Theorem. Suppose $x \in \mathbb{R}$. Then $(x+1)^2 - (x-1)^2 = 4x$.

Incorrect proof. Calculate:

$$\begin{aligned} (x+1)^2 - (x-1)^2 &= 4x \\ (x^2+2x+1) - (x^2-2x+1) &= 4x \\ x^2+2x+1-x^2+2x-1 &= 4x \\ 4x &= 4x. \end{aligned}$$

□

The problem with this "proof" is in its first line: it seems to assert as true exactly what it is trying to prove—circular reasoning. To make the mistake even more clear, consider the following false statement, which uses the same reasoning:

Theorem. In $\mathbb{R}$, we have

$$1 = 0.$$

Incorrect proof. Calculate:

$$\begin{aligned} 1 &= 0 \\ 0 \cdot 1 &= 0 \cdot 0 \\ 0 &= 0. \end{aligned}$$

□

To fix the proof of the original theorem, one could just list the lines of the proof in reverse order—hence, the moniker "backwards"—starting with $4x = 4x$. However, notice another flaw with the proof: by just listing these lines, we break the rule that a proof consists of sentences. Here is the correct form, fixing both problems:

Theorem. Suppose $x \in \mathbb{R}$. Then $(x+1)^2 - (x-1)^2 = 4x$.

Proof. Calculate:

$$\begin{aligned}(x+1)^2-(x-1)^2 &= (x^2+2x+1)-(x^2-2x+1)\\ &= x^2+2x+1-x^2+2x-1\\ &= 4x.\end{aligned}$$

□

Miscellaneous

- To prove a statement is false, give a specific concrete counterexample. Try to find the simplest one. The counterexample is more convincing and easier on the reader than an abstract argument. Conversely, to prove a statement is true, an example, although sometimes helpful, is not a proof. You must show why the statement is true in all instances. Some examples:

  Statement: $f(n) = n + 2$ is even for all $n \in \mathbb{Z}$.
  Disproof: Note that $f(1) = 3$, which is not even.

  Statement: $f(n) = n + 2$ is divisible by 7 for all $n \in \mathbb{Z}$.
  False proof: We have $f(12) = 14$, which is divisible by 7.

- If you use the phrase "by definition" in your writing, make sure to be specific: by definition of what? For example, you might write "by definition of a Hausdorff space". Using the phrase "by definition" in isolation is usually ambiguous, and if your reader cannot determine which definition your are referring to, then it is no help at all—everything in mathematics follows from the definitions!

- When writing down a calculation, avoid crossing out terms (for example, when terms cancel in fractions or when they add up to zero). This type of bookkeeping is easy for the writer, who is crossing out sequentially, but is usually confusing for the reader, who sees all of the crosses at once.

TEX pointers

- When defining a function, use `\colon` rather than `:`, as in

  `f\colon X\to Y.`

  The latter symbol is regarded as an operator and is padded by unwanted spaces.

- Use the TEX versions of the trig functions, e.g., `\cos(t), \tan(t)`, etc., rather then plain `cos(t), tan(t)`, etc. The same goes for logarithms:`\log(t), \ln(t)`.

- Add a little space before a differential, e.g., `\int x\,dx`, which gives $\int x\,dx$ rather than `\int xdx`, which gives $\int xdx$.

Here is a link to a nice three-page article on mathematical writing by Francis Su:

Writing Mathematics Well

and a one-page summary:

Some Guidelines for Good Mathematical Writing

Proof templates

The "blah, blah, blah"s appearing in the template proofs below usually consists of two parts. The first part is automatic: you expand the statement of the theorem using relevant definitions so that the reader understands what you are trying to prove. The second is where the real math occurs. It often requires some insight and creativity.

By convention, the proofs of certain types of statements are expected to have a certain structure. Here, we provide examples of some of these that are important for us. Deviating from these structures risks confusing your reader and, thus, usually entail the addition of a few words of guidance in your proof.

Induction

Proposition. For all $n \geq 1$, we have

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof. We will prove this by induction. For the base case, $n = 1$, the result holds since in that case

$$1 + 2 + \cdots + n = 1 = \frac{1 \cdot (1+1)}{2}.$$

It follows that

$$\begin{aligned}
1 + 2 + \cdots + (n+1) &= (1 + 2 + \cdots + n) + (n+1) \\
&= \frac{n(n+1)}{2} + (n+1) && \text{(by the induction hypothesis)} \\
&= \frac{n(n+1) + 2(n+1)}{2} \\
&= \frac{(n+1)(n+2))}{2} \\
&= \frac{(n+1)((n+1)+1)}{2}.
\end{aligned}$$

So the result holds for the case $n + 1$, too. The proposition follows by induction. □

Rules.

1. Always start a proof by induction by telling your reader that you are giving a proof by induction.

2. Next, show that result holds for the smallest value of n in question—in this case, $n = 1$.

3. Note that we assume the result is true for some $n \geq 1$. If we said, instead: "assume the result holds for $n \geq 1$", this would mean we're assuming the result for all $n \geq 1$. But that would be circular: we'd be assuming what we are trying to prove. Instead, at the induction step, we are merely saying that *if* we did know the result for a particular value of n, we could prove that it follows for the next value of n.

4. End the proof with a $\square$. This tells the reader the proof is over.

Set containment

Definition. Let A and B be sets. Then A is a subset of B, denoted $A \subseteq B$ if $a \in A$ implies $a \in B$.

Definition. Let A and B be sets. Then these sets are equal, denoted $A = B$ if both $A \subseteq B$ and $B \subseteq A$.

Definition. Let A and B be sets. Then the *union* of A and B is

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

In words, $x \in A \cup B$ if x is in A or $x \in B$. The *intersection* of A and B is

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

In words, $x \in A \cap B$ if $x \in A$ and $x \in B$.

Theorem. Let A and B be sets defined by [`some condition`]. Then $A \subseteq B$.

Proof. Let $a \in A$. Then blah, blah, blah. Therefore, $a \in B$. It follows that $A \subseteq B$. $\square$

Theorem. Let A and B be sets defined by [`some condition`]. Then $A = B$.

Proof. Suppose $a \in A$. Then blah, blah, blah. Therefore $a \in B$, too. Thus, $A \subseteq B$.

To show the opposite containment, suppose that $b \in B$. Then blah, blah, blah. Thus, $b \in A$. It follows that $B \subseteq A$. $\square$

Equivalence relations

Theorem. Define a relation $\sim$ on a set A by blah, blah, blah. Then $\sim$ is an equivalence relation.

Proof. Let $a, b, c \in A$.

Reflexivity. We have $a \sim a$ since blah, blah, blah. Therefore, $\sim$ is reflexive.

Symmetry. Suppose that $a \sim b$. Then, blah, blah, blah. It follows that $b \sim a$. Therefore $\sim$ is symmetric.

Transitivity. Suppose that $a \sim b$ and $b \sim c$. Then blah, blah, blah. It follows that $a \sim c$. Therefore, $\sim$ is transitive.

Since $\sim$ is reflexive, symmetric, and transitive, it follows that $\sim$ is an equivalence relation. □

INJECTIVITY, SURJECTIVITY, BIJECTIVITY

Proposition. The function $f\colon A \to B$ is injective.

Proof. Let $x, y \in A$, and suppose that $f(x) = f(y)$. Then blah, blah, blah. It follows that $x = y$. Hence, f is injective. □

Proposition. The function $f\colon A \to B$ is surjective.

Proof. Let $b \in B$. Then blah, blah, blah. Thus, there exists $a \in A$ such that $f(a) = b$. Hence, f is surjective. □

Proposition. The function $f\colon A \to B$ is bijective.

Proof. (Alternative 1.) We first show that f is injective. [Follow the template above to show injectivity.] Next, we show f is surjective. [Follow the template above toe show surjectivity.] □

Proof. (Alternative 2) Define $g\colon B \to A$ as follows: blah, blah, blah. Note that $g \circ f = \mathrm{id}_A$ since blah, blah, blah. Next, note that $f \circ g = \mathrm{id}_B$ since blah, blah, blah. □

Axioms for the real numbers

Field axioms

Definition A *field* is a set F with two operations, $+ : F \times F \to F$ (addition) and $\cdot : F \times F \to F$ (multiplication), satisfying the following axioms:

A1. Addition is commutative. For all $x, y \in F$,

$$x + y = y + x.$$

A2. Addition is associative. For all $x, y, z \in F$,

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

F3. There is an additive identity. There is an element of F, usually denoted 0, such that for all $x \in F$,

$$x + 0 = x.$$

F4. There are additive inverses. For all $x \in F$, there is an element $y \in F$ such that

$$x + y = 0.$$

The element y is denoted $-x$. (Subtraction is then defined by $x - y := x + (-y)$ for all $x, y \in F$.)

M1. Multiplication is commutative. For all $x, y \in F$,

$$xy = yx.$$

M2. Multiplication is associative. For all $x, y, z \in F$,

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

M3. There is a multiplicative identity. There is an element, usually denoted 1, such that:

(a) $1 \neq 0$, and

(b) $1x = x$ for all $x \in F$.

M4. There are multiplicative inverses. For each *nonzero* $x \in F$, there is a $y \in F$ such that

$$xy = 1.$$

The element y is denoted $1/x$ or x^{-1}. (Division is then defined by $x \div y := xy^{-1}$ for nonzero y.)

D. Multiplication distributes over addition. For all $x, y, z \in F$,

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

Order axioms

Definition. An *ordered field* is a field F with a relation, denoted $<$, satisfying

O1. (Trichotomy) For all $x, y \in F$, exactly one of the following statements is true:

$$x < y, \quad y < x, \quad x = y.$$

O2. (Transitivity) The relation $<$ is transitive, i.e., for all $x, y, z \in F$,

$$x < y \quad \text{and} \quad y < z \quad \Longrightarrow \quad x < z.$$

O3. (Additive translation) For all $x, y, z \in F$,

$$x < y \quad \Longrightarrow \quad x + z < y + z$$

O4. (Multiplicative translation) For all $x, y, z \in F$,

$$x < y \quad \text{and} \quad 0 < z \quad \Longrightarrow \quad xz < yz.$$

Remark: of course, we write $x > y$ if $y < x$, and we write $x \leq y$ if either $x = y$ or $x < y$.

Completeness

Let S be a subset of an ordered field F. An element $M \in F$ is an *upper bound* for S if $M \geq s$ for all $s \in S$. An element $M \in F$ is the *least upper bound* or *supremum* for S if it is an upper bound and is less than or equal to every upper bound. In this case, we write $M = \operatorname{lub} S$ or $M = \sup S$. Similarly, an element $m \in F$ is a *lower bound* for S if $m \leq s$ for all $s \in S$. An element $m \in F$ is the *greatest lower bound* or *infimum* for S if it is a lower bound and is greater than or equal to every lower bound. In this case, we write $m = \operatorname{glb} S$ or $m = \inf S$.

Definition. An ordered field F is *complete* if every nonempty subset $S \subseteq F$ which has an upper bound has a least upper bound in F.

The real numbers

Theorem. *There exists a complete ordered field.*

It can also be shown that any two complete ordered fields are isomorphic, i.e, they are the same except for renaming elements. Thus, there is essentially one complete ordered field, and it is the set of real numbers, $\mathbb{R}$.