MAT 25A - Class Notes

Kevin Lamb

November 11, 2016

Contents

0 A Note to My Students

These are the class notes for a 10-week introduction to real analysis. The first goal of this course is to introduce some of the basic properties of the real numbers including their sequences and standard topology. The second is to teach students the fundamentals of rigorous, mathematical proof.

The Math Department established this course as a sort of early look into the upper-division material found within the Mathematics major. As such, the prerequisites as few: Derivative and integral Calculus. This is a double-edged sword: The student does not need a complex mathematical background to understand the material, but there are still essential topics that the student must learn in order to achieve our second goal. That is, we need to build the student's familiarity with mathematical rigor and methods of proof. While necessary and useful in their own right, their detailed study is the focus of another course. Hence, we need to fit an entire course's worth of material into about five or so lectures. The student should be aware that they are only receiving a bare-bones education in this material, but this is all they need in order to succeed in the current course.

In regards to the second goal, the student should also be aware that clear and effective writing is the key to communicating mathematics with others. The history of Analysis is rife with examples of this point. The loose and "hand-wavy" nature of Calculus a la Leibniz and Newton is what inspired mathematicians like Euler, Cauchy, and Weierstrass to develop a rigorous foundation for the techniques that proved so useful and accurate in practice. Introductory analysis is a natural setting within which students can become acquainted with conveying their ideas in an eloquent and convincing manner.

On a more personal note to my students: Learning math will be difficult. You will fail **a lot** before you succeed. It is at this point in your academic career where I hope you take to heart the ubiquitous idea that *making mistakes is a critical part of learning.* You should never feel unintelligent or belittle yourself for making a mistake. I hope you all make lots of them and learn from them. When I was learning this stuff years ago, the hardest lesson for me to learn was humility - that I didn't always understand things the first time I saw them even though there were others around we who seemed so much better at it than me. For you, now is the time to recognize that this is where math gets harder and and to realize that this is an opportunity to grow. There's always room for improvement.

As the instructor, my only expectation of you is for you to put an honest effort into learning this material. It doesn't have to be your favorite subject. You don't have to master it while you're in my class. I just want to you to push. Push past your mistakes and frustration. Push yourself to a high standard. Push to be the best you can be. Know that we're doing this together and that I'm always here to help.

Finally, the best way to learn math is to do math. That being said...

Let's do some math!

1 Introduction to Proof

The notes from this section are drawn from the first chapter of the popular text *A Transition to Advanced Mathematics.* The same ideas may be found in the appendix of Bartle and Sherbert's *Introduction to Real Analysis.* The editions of the books aren't important since the material we're going to use hasn't changed for hundreds, if not thousands, of years.

1.1 Propositions and Connectives

In order to talk about math in a rigorous way, we need to establish how we can convince ourselves of truth. What is a fact and what isn't? This is where the casualty of the English language puts us at a disadvantage. We are oftentimes lazy with the meaning of our speech. Statements like

"I live in Davis" and "$x = y$"

are true under certain conditions and false under others. The first doesn't specify a time period for which it is true. The second depends on the values of x and y and, even less remembered, on what kinds of objects x and y are.

To remedy this issue of ambiguity, we define the statements whose truth are unconditional upon any other outside conditions.

Definition 1.1. A *proposition* is a statement with exactly one truth value. The *truth value* of a proposition is either the affirmative *true* or the negative *false.*

Example 1.1. The follow statements are all propositions:

1. "$3 > \pi$." (This is always false.)
2. "In our solar system, the closest planet to the sun is Mercury." (This is always true.)
3. "In 2016, Kevin's office is located in the Mathematical Sciences Building on the UC Davis campus." (It's true, but notice how much more specific I had to be.)
4. "$\sqrt{2}$ is either a rational number or an irrational number." (This is true but doesn't give us any information about the rationality of $\sqrt{2}$.)
5. "If you pass the final, you will receive a passing grade." (False if you didn't do any of the homework all quarter - so this isn't always true.)
6. "It is the case that $0 = 1$ or Kevin passes all of his students every time he teaches." (We'll let you ponder the truth value of this statement a little longer.)

The following statements are all not propositions:

1. "Either *The Room* is the best movie ever, or the Pope isn't Catholic." (This is an opinion.)
2. "$x = 5$." (This depends on the value of x.)

3. "What time does class start?" (This is a question... Questions have no truth value, but their answers do.)

4. "It will rain." (This statement admits more than one truth value.)

This does not completely solve our problem, however. There are propositions whose very truth value may or may not exist. For example, the statement

"This statement is false"

is an example of a *paradox*. Either the statement is true or it is false. The statement must be false if we assign it the truth value `true`, yet it must be true if we assign it the truth value `false`. At the beginning of the 20^{th} century, Kurt Gödel showed in his famous Incompleteness Theorem that such paradoxes are an inevitability of any *axiomatic set theory*. That is, in any way that we develop a system of logic there will always be statements whose truth values are *undecidable* - neither `true` nor `false`. The theory will still be consistent for either assignment of truth value. Therefore, in order to still be able to create a viable logic system, we choose the truth value that makes the most sense to us within our own observation and experience. We will see such undecidable statements again, but for now we can avoid paradoxes and still get a firm grasp of mathematical rigor.

For propositions, their truth value is set in stone. Therefore, we can talk about the case that encompass every other scenario apart from the original proposition. That is,

Definition 1.2. The *negation* (also called a *denial*) of a proposition P is denoted $\sim P$, and its truth value is the opposite of that of P.

Example 1.2. The following are, in order, the negations to the propositions in the previous example:

1. "$3 \leq \pi$."

2. "In our solar system, the planet closest to the sun is not Mercury."

3. "In 2016, Kevin's office is located somewhere other than the Mathematical Sciences building on the UC Davis campus."

4. "$\sqrt{2}$ is neither rational nor irrational."

5. "You passed the final but you still failed the class."

6. "It is the case that $0 \neq 1$ and Kevin failed at least one student at least once when he taught previously."

This definition now assumes that if a proposition is not `true`, then it must be false and vice versa. This is the *law of excluded middle*: A proposition is either `true`or it is `false`, not both or neither. Suffice it to say that it is possible to create logic systems where the law of excluded middle is violated, and yet the theory is still consistent; however, such theories are rather untenable and obnoxious - hence they are avoided like the plague.

We can now create new propositions from ones that we already have:

Definition 1.3. Let P and Q be propositions.

The *conjunction* of P and Q, denoted $P \wedge Q$, is the proposition "P and Q." It is `true` when both P and Q are true.

The *disjunction* of P and Q, denoted $P \vee Q$, is the proposition "P or Q." It is `true` when **at least one** of P or Q is true.

Example 1.3. These examples combine those from the first example:

1. "$3 > \pi$ and, in our solar system, the closest planet to the sun is Mercury." (False, since the P proposition is false.)

2. "$3 > \pi$ or, in our solar system, the closest planet to the sun is Mercury." (True, since the Q proposition is true. Observe that this use of "or" is the *inclusive or* rather than the *exclusive or*.)

3. Example 1.1 (6) is a disjunction with P="$0 = 1$." and Q ="Kevin passes all of his students every time he teaches." The conjunction is false since P is false.

This may seem a bit contradictory: The truth value of a conjunction and a disjunction depends on the truth values of their constituents. To convince you that these types of statements are indeed still propositions, consider the following analogy:

> "Assume $x = 5$ and $y = 2$. Then $x + y = 7$ or $x - y = 3$. Also, $x + y = 3$ and $2x = 10$."

This pair of statements are indeed propositions since, although they depend on variables x and y, their values are already prescribed. That is, we can't change the values of x and y. The same holds for the truth values of P and Q in the definitions of conjunction and disjunction. They are propositions, hence their truth values cannot be changed. Therefore, the statements $P \wedge Q$ and $P \vee Q$ have a definite truth value as per theirs definitions - hence they must be propositions.

Perhaps our concerns above are moot: What if the value of a proposition is independent of the truth values of its constituents?

Definition 1.4. A *tautology* is a proposition whose truth value is always `true` independent of the truth values of its constituents.

Likewise, a *contradiction* is a proposition whose truth value is always `false` independent of the truth values of its constituents.

Example 1.4. The traditional examples stem from the law of excluded middle, and we use a *truth table* to validate our claims.

Let P be a proposition. The proposition $P \vee P$ is a tautology:

P	$\sim P$	$P \vee \sim P$	$P \wedge \sim P$
T	F	T	F
F	T	T	F

Because the third column is composed solely of `true`values, the proposition there is a tautology. Because the fourth column is composed solely of `false`values, the proposition there is a contradiction.

Observe that I have organized this table with a double vertical line. This separates the values that are variable (on the left) from those that depend on those values (on the right). This becomes all the more important when multiple variable components are added to the mix.

Now that we know when a proposition is always true or always false, independent of the truth values of its constituents, we can define a notion of equivalence of propositions:

Definition 1.5. Two propositions P and Q that depend on the same component propositions are *equivalent*, denoted $P \Leftrightarrow Q$, exactly when they have the same truth values for the same combination of truth values of those constituents. Otherwise, they are said to be *inequivalent*.

Example 1.5. A standard example illustrates how to negate the conjunction and disjunction operations. That is, we show *DeMorgan's Laws* that

> For propositions P and Q, we have $[\sim (P \wedge Q)] \Leftrightarrow [(\sim P) \vee (\sim Q)]$ and $[\sim (P \vee Q)] \Leftrightarrow [(\sim P) \wedge (\sim Q)]$.

We verify the first of these with a truth table:

P	Q	$\sim P$	$\sim Q$	$P \wedge Q$	$\sim (P \wedge Q)$	$(\sim P) \vee (\sim Q)$
T	T	F	F	T	F	F
T	F	F	T	F	T	T
F	T	T	F	F	T	T
F	F	T	T	F	T	T

(Notice that I've grouped my truth values for P and Q in such a way that it is obvious that I've exhausted all four possibilities.) Observe that the final two columns have identical truth values in each row (for each combination of truth values for P and Q). Therefore, the two statements are equivalent. The reader should create the truth table for the other negation explained above.

1.2 Conditionals and Biconditionals

We can now form more complex propositions. Some of the most common types of propositions aren't necessarily those that we've already seen as either stand-alone propositions or conjunctions and disjunctions. Quite the contrary, in math we give ourselves a set of conditions that guarantees a certain outcome. From Calculus, some examples include

> "If a function is differentiable at some point, then it is also continuous at that point"

and

> "If the terms of a series don't approach zero, then the series must diverge."

These kinds of propositions are defined as follows:

Definition 1.6. Let P and Q be propositions. The *conditional* proposition $P \Rightarrow Q$ is the proposition "If P, then Q" or, equivalently, "P implies Q." The "if" statement is called the *antecedent*, and the "then" statement is called the *consequent*. This proposition is `true` exactly when P is false or Q is true.

The truth table for $P \Rightarrow Q$ is given below for clarity:

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

You can think of the truth of $P \Rightarrow Q$ as a promis: I guarantee you that Q happens whenever P happens. This is false only when P happens but I don't follow through to make Q happen as well. Also notice that, by a combination of its definition and an application of DeMorgan's Law, the negation of the implication is given as

$$[\sim (P \Rightarrow Q)] \Leftrightarrow [(\sim P) \wedge Q].$$

The student should verify this immediately using a truth table.

Example 1.6. As a rudimentary example, consider the proposition

"If it rains tomorrow, then we won't have a picnic outside."

The only way for me to not follow through on the promise is if it does indeed rain the next day and yet we still go have our picnic outside in the rain (I know, I know... Worst... Teacher... EVER...). Notice what this implies: I'm still keeping my promise to you no matter what happens after we determine it isn't raining. That is, **if the antecedent is false, then the conditional is true regardless of the truth value of the consequent**. Thus, the only time we need to worry about the conditional statement failing is when the antecedent is true.

We can now define some related conditionals:

Definition 1.7. Let P and Q be propositions.
The *converse* of $P \Rightarrow Q$ is the conditional $Q \Rightarrow P$.
The *inverse* of $P \Rightarrow Q$ is the conditional $(\sim P) \Rightarrow (\sim Q)$.
The *contrapositive* of $P \Rightarrow Q$ is the conditional $(\sim Q) \Rightarrow (\sim P)$.

The inverse isn't usually a useful proposition to consider as it doesn't have a useful relation to $P \Rightarrow Q$. The other two, however, are **extremely** useful for two different ideas. Let's first consider why we like the contrapositive so much.

Example 1.7. For propositions P and Q, we claim that the conditional $P \Rightarrow Q$ is equivalent to its contrapositive $(\sim Q) \Rightarrow (\sim P)$. To see this, consider the truth table

P	Q	$\sim P$	$\sim Q$	$P \Rightarrow Q$	$(\sim Q) \Rightarrow (\sim P)$
T	T	F	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Indeed, this shows that the final two columns are equivalent propositions.

The final logical notion that we wish to describe is just a reiteration of the notion of equivalence defined in the previous section.

Definition 1.8. Let P and Q be propositions. The *biconditional* $P \Leftrightarrow Q$ is the proposition "P if and only if Q." This is true exactly when P and Q have the same truth values. We sometimes abbreviate "if and only if" by "iff."

The truth table for this proposition looks like this:

P	Q	$P \Leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

Example 1.8. For propositions P and Q, we claim that $P \Leftrightarrow Q$ is equivalent to the conjunction $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$. Again, we verify this by a truth table:

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$P \Leftrightarrow Q$	$(P \Rightarrow Q) \wedge (Q \Rightarrow P)$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	T

Sure enough, the final two columns agree so that they are equivalent propositions. In fact, this is usually how we prove an equivalence in practice. We will see plenty of examples of this later.

Remark 1.1. One phrase that comes up quite frequently in practice is "give a necessary and sufficient condition for a proposition S."

We first comment that a *necessary condition* for a proposition S is a proposition P such that $S \Rightarrow P$ is true. That is, if P doesn't happen, then S can't happen - it is **necessary** that P be satisfied in order for S to be satisfied.

We now comment that a *sufficient condition* for a proposition S is a proposition P such that $P \Rightarrow S$ is true. That is, if P is satisfied, then we guarantee that S happens as well - it is **sufficient** that P be satisfied in order to satsify S.

Therefore, a *necessary and sufficient condition* for a proposition S is a proposition P that is both a necessary condition for S and a sufficient condition for S (duh...). Observe that by our last example, we would have both $S \Rightarrow P$ and $P \Rightarrow S$ so that $P \Leftrightarrow S$. Hence, a necessary and sufficient condition for S is just a proposition that is equivalent to S (and, hopefully, easier to work with in practice).

1.3 Quantifiers

We now revisit the fact that the statement "$x = 5$" isn't a proposition. Recall that this was because we didn't know what x was in order for us to evaluate the equivalence. In order to turn this statement into a proposition, we have to both allow ourselves to assign values to x without assigning any fixed value to x. Here's one way to do this:

Example 1.9. If we consider x to be abale to take on natural number quantities, then we know there is **some** value of x that satisfies this statement "$x = 5$." However, it is quite evident that not **every** natural number satisfies this proposition.

Now also consider the statement "$n < n + 1$. Again, we consider n to be able to take on natural number quantities. Not only does there exist **some** number that satisfies this statement, this statement is true for **all** natural numbers n.

However, consider that if x and n were instead only allowed to take on colors, then neither of these statements ever holds true - yet they are then both still propositions.

To make this point rigorous, we create the following defintions.

Definition 1.9. Let $P(x)$ be a statement whose truth value is dependent on the value of the object x.

A *universe* is a collection U of states that an object x is allowed to occupy.

The statement $(\forall x)\ P(x)$ is the proposition "For all x, $P(x)$." This is true exactly when every x in U makes $P(x)$ true. The symbol $\forall$ is called the *universal quantifier.*

The statement $(\exists x)\ P(x)$ is the proposition "There exists an x such that $P(x)$." This is true exactly when the number of x in U that make $P(x)$ true is nonzero. The symbol $\exists$ is called the *existential quantifier.*

The statement $(\exists! x)\ P(x)$ is the proposition "There exists a unique x such that $P(x)$." This is true exactly when the number of x in U that make $P(x)$ true is one. The symbol $\exists!$ is called the *unique existential quantifier.*

Let $Q(x)$ be another statement whose truth value depends on x. We say that $P(x)$ and $Q(x)$ are *equivalent* if they are true on the same values of x.

Example 1.10. Consider the following propositions:

1. "$(U = \mathbb{N})(\exists x)(x = 5)$." (This is true, since there is some x in the naturals where $x = 5$ is true.)

2. "$(U = \mathbb{N})(\forall x)(x = 5)$." (This is false, since this is false for $x = 1$ - hence, it's not true for all x.)

3. "(U =negative reals)$(\exists x)(x = 5)$." (This is false, since no negative number is equal to a positive number.)

4. "$(U = \mathbb{R})(\forall x)(\exists! y)(x + y = 0)$." (This is true, since every real number has exactly one additive inverse - the value y is a function of x.)

5. "$(U = \mathbb{R})(\exists x)(\forall y)(x+y = 0)$." (This is false, since there is no x that is the additive inverse for every other real number y.)

Remark 1.2.

1. The universe in use changes from proposition to proposition; however, it is usually obvious from the context what set it should be, and we never mention it unless we need to specify which quantified variable comes from which universe.

2. In order to negate the proposition "$(\forall x)P(x)$," we can't merely negate $P(x)$. The quantifier must change as well.

   For example, the negation of (2) in Example 1.10 isn't just "$(\forall x)(x \neq 5)$. This can't be true since I can set $x = 5$. Since my original proposition was false, I need this negation to be true. Hence, I need to change the quantifier as well. The appropriate negation is "$(\exists x)(x \neq 5)$," which is now certainly true (choose $x = 4$).

   A similar effect takes place in the negation of "$(\exists x)P(x)$. Its negation is "$(\forall x) \sim P(x)$.

1.4 Various Proof Techniques

In this section, we briefly go over examples of some common proof techniques that one sees throughout his or her academic career.

1.4.1 Direct Proofs

There are two commons types of *direct proofs*, and both involved proving the proposition as it is stated. It is perhaps best to give examples of its various forms. While reading through this section, pay careful attention to both the format and style in which proofs are written.

Example 1.11. (Direct proof of $P \Rightarrow Q$) We first outline the method as follows:

Proof. `Assume` P `is true.`

$\vdots$

`Therefore,` Q `is true.`
`Conclusion:` $P \Rightarrow Q$ `is true.` □

Let's see this in action for the claim

"Let n be an integer. If n is even, then n^2 is even."

Proof. Suppose that n is an even integer. Then there must exist some integer k such that $n = 2k$. Now consider that

$$n^2 = (2k)^2 = 4k^2 = 2(2k^2).$$

Hence, there is another integer $m = 2k^2$ where we can write $n^2 = 2m$. Therefore, n^2 is also an even integer. □

I didn't have to change the form of the proposition to proceed, so this does indeed constitute a direct proof of the proposition.

From a stylistic point of view, notice that my proof wasn't simply a laundry list of successive equations. Rather, it was a compilation of reasons that strung together to create a logical argument (much like a short response essay). Also notice that when I did need to list a series of computations, I dropped them into the next line. The reason for this is twofold: (1) It helps the reader to step through my thought process, and (2) it draws attention to an important piece of my argument by setting it aside from the smaller details. My proof also begins with the word "Proof:" at the top and ends with a symbol to indicate the stopping point. Some people like to end their proofs with a "Q.E.D.", which is also perfectly acceptable ("Q.E.D." stands for "quod erat demonstratum", which in English roughly translates as "quite enough done").

Example 1.12. (Direct proof of $(\forall x)P(x)$) We outline this proof as follows:

Proof. `Pick an arbitrary` x `in the universe.`

$\vdots$

`Show that` $P(x)$ `is true.`
`Since` x `was an arbitrary element of the universe,`
`conclude that` $(\forall x)P(x)$ `is true.`

□

Let's see an example by proving

"For every natural number n, $4n^2 - 6.8n + 2.88 > 0$."

Proof. Let n be a natural number. Observe that $4n^2 - 6.8n + 2.88 = 4(n - .8)(n - .9)$. Since $n \geq 1$, each term in the product is positive. Therefore, the product itself is positive as well. Hence, the inequality holds for each natural number n. □

Example 1.13. (Direct proof of "$(\exists x)P(x)$") The method is just as easy as this:

Proof. `Just find an` x `that makes` $P(x)$ `true.` □

Proofs of this type are typically called *constructive proofs* since one usually has to construct/create an example that satisfies the proposition.

Here's a quick example:

"There is some real number such that $\sin x = x$."

Proof. Choose $x = 0$. □

And yes, that constitutes a complete proof. If you want, you can (and probably should, in general) explain why it constitutes an example.

Notice that this would be much trickier if I had asked you to find a **positive** real number x satisfiying $\sin x = x$. We will see later how to do this rigorously.

1.4.2 Indirect Proofs

These kinds of proofs really only boil down to two types again: Contraposition and contradiction.

Example 1.14. (Contraposition of $P \Rightarrow Q$) We outline this method as follows:

Proof. `Assume` $\sim Q$ `is true.`

$\vdots$

`Deduce that` $\sim P$ `is true.`
`Hence,` $(\sim Q) \Rightarrow (\sim P)$ `is true.`
`Equivalence allows us to conclude that` $P \Rightarrow Q$ `is true.` □

Again, we choose an example to demonstrate the method:

"Let n be an integer. If n^2 is even, then n is even."

The astute reader will realize that this is the converse of a previous proposition.

Proof. Suppose instead that n is not even. That is, assume that n is odd. Then there is some integer k such that $n = 2k + 1$. Then consider that

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Hence, there is another integer $m = 2k^2 + 2k$ such that $n^2 = 2m + 1$. Then n^2 is odd; that is, not even. Therefore, if n is odd, then so is n^2. This is equivalent to the proposition. □

Contraposition, like other methods of indirect proof, allows us to beat the proposition into a form that is easier for us to deal with that a direct proof. If we had to go directly here, we would have had to check the proposition by cases. This is usually not an option in general, though.

Example 1.15. (Proof of P by contradiction) We outline this method as well:

Proof. `Assume instead that` $\sim P$ `is true.`

$\vdots$

`Deduce` Q `is true.`

$\vdots$

`Yet also deduce that` $\sim Q$ `is also true.`
`Assuming` $\sim P$ `forces a contradiction.`
`Conclude` $\sim P$ `is false.`
`Therefore,` P `is true.` □

We exemplify this method with one of the most classic proofs of all:

"There is an infinite number of prime numbers."

Proof. Suppose to the contrary that there are only finitely many primes. List them as $p_1, p_2, \ldots, p_n$. Now consider the integer $P = p_1 p_2 \cdots p_n + 1$. Since P is an integer, it must have at least one of the p_i, say it's p_1 after potentially relabeling the list of primes.

Now notice that p_1 divides the difference $P - p_1 p_2 \cdots p_n$. But this difference is equal to 1 so that p_1 divides 1 as well. Therefore, $p_1 = 1$. Since 1 isn't prime, p_1 is both prime and not prime - this is a contradiction. Hence, our assumption about there being a finite number of primes must be false. Therefore, there is an infinite number of primes. □

This was first proven by Euclid in his work *Elements* circa 300 BCE.

When dealing with quantifiers, we need to recall that negating a quantified proposition requires the negation of the statement **and** the quantifier.

Example 1.16. (Proof of "$(\forall x)P(x)$" by contradiction) This is outlined simply as follows:

Proof. `Assume instead that` $(\exists x) \sim P(x)$.
`Let` t `be such an element where` $\sim P(t)$ `is true.`

$\vdots$

`Deduce` Q `is true (this doesn't necessarily depend on` t`).`

$\vdots$

`Also find` $\sim Q$ `to be true.`
`Observe the contradiction` $Q \wedge (\sim Q)$.
`Hence,` $(\exists x) \sim P(x)$ `is false.`
`Conclude` $(\forall x)P(x)$ `is true.` □

Let's try a harder example:

"For all $0 < x < \frac{\pi}{2}$, $\sin x + \cos x > 1$."

Proof. Suppose to the contrary that there exists some t in the interval $\left(0, \frac{\pi}{2}\right)$ such that $\sin t + \cos t \leq 1$.

Observe first that both $\sin x$ and $\cos x$ are positive for $0 < x < \frac{\pi}{2}$. Then we have

$$\begin{aligned} 0 < \sin t + \cos t &\leq 1 \\ 0 < (\sin t + \cos t)^2 &\leq 1^2 = 1 \\ 0 < (\sin^2 t + \cos^2 t + 2\sin t \cos t) &\leq 1 \\ 0 < 1 + 2\sin t \cos t &\leq 1 \\ -1 < 2\sin t \cos t &\leq 0. \end{aligned}$$

Recalling that we have already observed that both $\sin t$ and $\cos t$ are both positive, the final inequality shows instead that one of them must be no greater than zero. This apparent contradiction shows that our initial assumption about the existence of t was false. Therefore, $\sin x + \cos x < 1$ for all x in the interval $\left(0, \frac{\pi}{2}\right)$. □

If we want to show that the proposition is false, then all we need to do is show that there is **some** choice of x that fails the proposition $P(x)$. Such an x is called a *counterexample.*

Example 1.17. (Proof of "$(\exists x)P(x)$" by contradiction) We outline this as well:

Proof. `Assume instead that` $(\forall x) \sim P(x)$`.`

⋮

`Deduce` Q `is true.  (Again, this doesn't necessarily depend on` x`)`

⋮

`Also find` $\sim Q$ `to be true.`
`Hence, we draw the contradiction` $Q \wedge (\sim Q)$`.`
`Therefore,` $(\forall x) \sim P(x)$ `is false.`
`Conclude that` $(\exists x)P(x)$ `is true.` □

Let's do a practical example. The Jersey turnpikes have a toll system in place where if you check in at a toll booth too soon, you automatically get a speeding ticket. This puts the following proposition to use:

> "Two toll booths are 15 miles apart. If you check into the second booth less than 15 minutes after you checked into the first, then there must be some point between the two booths where you were traveling faster than the posted speed limit of 60 miles per hour."

Observe that we can rewrite this as

> "There exists a time t between 0 and 15 minutes such that your speed was greater than 60 miles per hour."

Proof. Let $T < 15$ be the time it took you to travel from the first booth to the second. Now suppose instead that you were traveling at a speed x no greater than 60 miles per hour for the whole T-minute trip.

Observe now that the furthest distance you can travel is given by

$$(x \text{ miles/min})(T \text{ min}) \leq (1 \text{ mile/min})(T \text{ minutes} = T \text{ miles} < 15 \text{ miles}.$$

That is, the distance between the toll booths is both strictly less than and equal to 15 miles. This contradiction shows that you couldn't have been traveling no more than the speed limit for your entire trip. Therefore, you were speeding at some point during the trip. □

Notice here how I had to introduce some notation to make my proof concise. If I didn't label these values, my constant referencing would have been much messier and would have made my proof a lot harder to read.

1.4.3 Common two-step techniques

Now that we have covered the basic techniques, there are still a few common ones that need to be mentioned before we continue.

Example 1.18. (Proof of "$P \Leftrightarrow Q$") This is outlined as follows:

Proof. `First show that` $P \Rightarrow Q$ `is true however you'd like.`
`Then show that` $Q \Rightarrow P$ `is also true however you'd like.`
`We've shown this is equivalent to` $P \Leftrightarrow Q$`.`
`Conclude that` $P \Leftrightarrow Q$ `is true.` □

Let's try this out on

"Let n be an integer. It is the case that n is even if and only if n^2 is even."

Proof. Both directions we're proven as examples above. The *forward/only-if direction* was shown via direct proof. The *backward/if direction* was shown via contraposition. Therefore, the statement must be true. □

We mostly see this two-step proof when we are trying to show that two sets are the same. We'll come back to this idea soon.

Finally, we extend one of our previous methods.

Example 1.19. (Proof of "$(\exists! x)P(x)$") This is outlined as follows:

Proof. `First show that` $(\exists x)P(x)$ `is true however you'd like.`
`Now establish that` $(\forall y)(\forall z)[\{P(y) \wedge P(z)\} \Rightarrow \{y = z\}]$ `is true. Typically done as follows:`
`Assume` y `and` z `make` $P(y)$ `and` $P(z)$ `true.`
$\vdots$
`Deduce that` $y = z$`.`
`Conclude that` $(\exists! x)P(x)$ `is true.` □

Let's show an elementary fact rigorously:

"There exists a uniqe real number x such that $5 + x = 0$."

Proof. Observe that $x = -5$ satisfies $5 + x = 0$.
Now suppose that y and z both satisfy $5+y = 0$ and $5+z = 0$. Since both expressions are equal to zero, we must have $5 + y = 5 + z$ so that $y = z$. Therefore, $x = -5$ is the unique real number satisfying this equality. □

An Important Note To Students

From this point on in the class, the student should focus using an appropriate amount of rigor and proper writing style. As one of the goals of the class is to develop these skills, **the student should be frequently asking for feedback about his or her writing.** Becoming familiar with proof writing will not happen overnight, so I recommend that the student do as many assigned exercises as possible to gain such a familiarity and become comfortable with this type of exposition. It should almost feel more like an English class in debate than a Math class.

2 Foundational Set Theory

The next section of notes comes from the Bartle and Sherbert's well received text *Introduction to Real Analysis*. Again, the edition is irrelevant seeing how these ideas have become a staple of mathematical education over the past century and haven't changed much, if at all.

2.1 Sets and Functions

One of the most pervasive ideas in all areas of math is the notion of a set. The idea is to have all objects that satisfy a certain property all in one convenient package. Once these packages are made, we can start manipulating them in effective and meaningful ways.

Definition 2.1. A *set* a collection of objects. If x is an object found in a set A, then we say that x is an *element* of A and denote this by $x \in A$. If a set contains no objects, then it is called the *empty set* and is denoted by $\emptyset$.

Two sets A and B are said to be *equal* if they contain the same elements. This is denoted by $A = B$.

Supposing we have some established set of objects, perhaps we don't want to consider all of them. Maybe we only want to consider those that satisfy an extra proposition. This yields the notion of a subset.

Definition 2.2. Given a pair of sets A and B, if every element of B is also an element of A, we say that B is a *subset* of A and denote this by $B \subset A$. If we allow for the possibility that A and B are the same set, then we use the notation $B \subseteq A$. If this is not the case, then we say that B is a *proper subset* of A.

Remark 2.1. 1. A common way of constructing a set $S$is to state which objects x are being considered (the universe U in which they exist) and then to propose a proposition $P(x)$ that the elements of the set should satisfy. The way we denote this is by the equivalent notations

$$S = \{x \in U \mid P(x)\} = \{x \in U \colon P(x)\}.$$

Sometimes the universe is omitted from the construction, but it is obvious from the context what that universe should be.

2. When showing that $B \subseteq A$, we simply prove the proposition

"$(\forall x \in B)(x \in A)$."

To show that two sets A and B are equal, we need to show that $A \subseteq B$ and $B \subseteq A$. This amounts to the equivalence

$$(\forall x \in B)(x \in A) \wedge (\forall x \in A)(x \in B)$$

$$\Leftrightarrow$$

$$(\forall x \in A)(\forall y \in B)[(x \in B) \wedge (y \in A)].$$

Example 2.1. The following are all sets:

1. The *natural numbers* $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$
2. The *integers* $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
3. The *rational numbers* $\mathbb{Q} = \left\{\frac{p}{q} \mid p, q \in \mathbb{Z} \text{ and } q \neq 0\right\}$
4. The *real numbers* $\mathbb{R} = \ldots$ let's come back to this one.
5. $S_1 = \{6, 7, 8, 9, 10\}$
6. $S_2 = \{x \in \mathbb{N} \mid x \leq 5\} = \{1, 2, 3, 4, 5\}$
7. $S_3 = \{x \in \mathbb{Q} \mid x^2 < 2\}$
8. $S_4 = \{x \in \mathbb{R} \mid x^2 < 0\} = \emptyset$.

There is some standard notation for sets that satisfy certain propositions when we involved more than one set beyond the notion of subset described above.

Definition 2.3. Let A and B be sets.

The *union* of A and B is the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

The *intersection* of A and B is the set

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

If $A \cap B = \emptyset$, then we say that A and B are *disjoint*.

The *complement of B in A* is the set

$$A - B = A \backslash B = \{x \in A \mid x \notin B\}.$$

The *Cartesian product* of A and B is the set

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

A slightly different yet related definition is given as follows:

Definition 2.4. Let I be a set, called the *index set*, and let $\{A_\alpha\}_{\alpha \in I}$ be a collection of sets indexed by I.

The *union* of all the sets $\{A_\alpha\}_{\alpha \in I}$ is denoted

$$\bigcup_{\alpha \in I} A_\alpha = \{x \mid x \in A_\alpha \text{ for some } \alpha \in I\}.$$

The *intersection* of all the sets $\{A_\alpha\}_{\alpha \in I}$ is denoted

$$\bigcap_{\alpha \in I} A_\alpha = \{x \mid x \in A_\alpha \text{ for all } \alpha \in I\}.$$

If the index set is $I_k = \{1, 2, \ldots, k\}$ or $\mathbb{N}$, then the union is denoted by

$$\bigcup_{n=1}^{k} A_n \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n$$

and the intersection is denoted by

$$\bigcap_{n=1}^{k} A_n \quad \text{and} \quad \bigcap_{n=1}^{\infty} A_n$$

Example 2.2. Consider from our previous example that:

1. $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} (\subset \mathbb{R})$.
2. Given $A = \{1, 2, 3, 4, 5, 6, 7\}$ and $B = \{4, 5, 6, 7, 8, 9, 10\}$, we have
   $$A \cap B = \{4, 5, 6, 7\}, \quad A \cup B = S_1 \cup S_2, \quad A - B = \{1, 2, 3\}, \quad \text{and } B - A = \{8, 9, 10\}.$$
3. Given $A = \{1, 2, 3\}$ and $B = \{4, 5\}$, we have
   $$A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$
   and
   $$B \times A = \{(4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (5, 3)\}.$$
   Notice that $A \times B \neq B \times A$. In fact, $A \not\subset B$ and $B \not\subset A$.
4. If $A_n = \{1, 2, \ldots, n\}$ is defined for all $n \in \mathbb{N}$, then
   $$\bigcup_{n=1}^{\infty} A_n = \mathbb{N} \quad \text{and} \quad \bigcap_{n=1}^{\infty} A_n = \{1\}.$$

Example 2.3. To demonstrate how we prove propositions involving sets, we consider DeMorgan's Laws for sets:

$$A - (B \cup C) = (A - B) \cap (A - C)$$

and

$$A - (B \cap C) = (A - B) \cup (A - C).$$

We prove the first equivalence:

Proof. "$\subseteq$": If the set $A - (B \cup C) = \emptyset$, there is nothing to show (this is the *vacuous inclusion*). Suppose then that it is nonempty. [Usually this part of the proof is omitted.]

Let $x \in A - (B \cup C)$. Then $x \in A$ but $x \not B \cup C$. By DeMorgan's Law for propositions, we know that the latter proposition is equivalent to $x \notin B$ and $x \notin C$. Then $x \in A$ and $x \notin B$, and also $x \in A$ and $x \notin C$. Hence, $x \in (A - B)$ and $x \in (A - C)$. Therefore, $x \in (A - B) \cap (A - C)$.

"$\subseteq$": Now let $x \in (A - B) \cap (A - C)$. Then $x \in A$ and $x \notin B$, and also $x \in A$ and $x \notin C$. This is consolidated as $x \in A$, and also $x \notin B$ and $x \notin C$. We can further simplify this proposition as $x \in A$ and $x \notin B \cup C$. Therefore, $x \in A - (B \cup C)$.

These two steps show that the two sets in questions are equal. $\square$

We leave the proof of the other law to the reader.

We now make rigorous a few familiar ideas.

Definition 2.5. Let A and B be sets. A *function* f from A to B is a subset of $A \times B$ such that for each $a \in A$ there exists a unique $b \in B$ where $(a, b) \in f$. This is often denoted by $f \colon A \to B$. We also denoted the ordered pair (a, b) as $a \mapsto b$ and use the relation $f(a) = b$ in this case.

The set A is called the *domain*, and the set B is called the *codomain*. The *range* of the function is the set

$$Rng(f) = \{b \in B \mid (a, b) \in f \text{ for some } a \in A\}.$$

If we want to consider the outputs of many elements of the domain simultaneously, or if we want to know where all the elements of a part of the codomain were mapped from, we can extend this definition.

Definition 2.6. Let $f\colon A \to B$, $C \subset A$, and $D \subset B$.
The *image of C under f* is the set

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

The *preimage of D under f* is the set

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

Example 2.4.

1. The set $\{(1, x), (2, x), (3, y)\}$ is a function with domain $\{1, 2, 3\}$ and range $\{x, y\}$.

2. For any nonempty set X, a function $s\colon \mathbb{N} \to X$ is called a *sequence* of elements in X. This is usually denoted instead by $s(n) = s_n$.

3. The empty set $\emptyset$ is also considered a function. Its domain and range are both empty.

4. Define the squaring function $f\colon \mathbb{R} \to \mathbb{R}$ as $f(x) = x^2$ for each $x \in \mathbb{R}$. The range of this function is the nonnegative reals, a proper subset of the codomain. Observe that for each positive real b, there are two values $\pm\sqrt{b}$ in the domain that map to it.

5. Let $C = \{2, -1, 1, 4\}$ and $D = \{9, 25\}$. For the squaring function f above, we have $f(C) = \{1, 4, 16\}$ and $f^{-1}(D) = \{3, -3, 5, -5\}$.

Definition 2.7. Let $f\colon A \to B$ be a function.
The function f is *injective/one-to-one* if

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

We say f is an *injection*.
The function f is *surjective/onto* if $Rng(f) = B$. We say f is a *surjection*
The function f is *bijective* if it is both injective and surjective. We say f is a *bijection*.

Example 2.5. The squaring function $f\colon \mathbb{R} \to \mathbb{R}$ is not injective. However, $F\colon [0, \infty) \to \mathbb{R}$ given by $F(x) = x^2$ is.
The same function f isn't surjective, either. However, $G\colon \mathbb{R} \to [0, \infty)$ given by $G(x) = x^2$ is.
Combining these features into the functions $H_1\colon [0, \infty) \to [0, \infty)$ and $H_2\colon (-\infty, 0] \to [0, \infty)$ given by $H_1(x) = H_2(x) = x^2$ are both bijections.

There are a few other operations we can talk about regarding functions, but we will only need those discussed above within the context of this class. The will certainly come up in later courses in analysis, but the reader may be "reminded" of these operations of *composition*, *inversion*, and *restriction* as they arise.

2.2 The Natural Numbers and Mathematical Induction

We now begin our axiomatic approach to constructing familiar sets of numbers. Yes, the set $\mathbb{N}$ is merely the collection of positive integers - but notice that this "definition" is a bit self-referential... The set of integers is the collection of the number '0' along with the positive and negative versions of the natural numbers... Hence, we need to move away from this definition (or at least define the set of integers in a meaningful way). We take Peano's approach by devising a collection of properties, called *axioms*, that the elements of $\mathbb{N}$ must satisfy.

Definition 2.8. The *Peano Axioms/Postulates* for the set $\mathbb{N}$ of so-called *natural numbers* are given as follows:

N1) 1 belongs to $\mathbb{N}$.

N2) If n belongs to $\mathbb{N}$, then its *successor* $n+1$ also belongs to $\mathbb{N}$.

N3) 1 is not the successor of any element of $\mathbb{N}$.

N4) If $n, m \in \mathbb{N}$ have the same successor, then $n = m$.

N5) A subset of $\mathbb{N}$ which contains 1, and which contains $n+1$ whenever it contains n, must in fact be $\mathbb{N}$ itself.

It is this final axiom N5 that builds the notion of *mathematical induction*. The idea is to create a set $S \subseteq \mathbb{N}$ satisfying a proposition $P(n)$ for each n in S, and then we show that S satisfies N5. Our conclusion is that $P(n)$ holds for each $n \in \mathbb{N}$.

Theorem 2.1. (Proof by Mathematical Induction) Let $S \subseteq \mathbb{N}$ be the collection of natural numbers n satisfying a proposition $P(n)$. Suppose that it can be shown that

1. The number $1 \in S$, and

2. For every $k \in \mathbb{N}$, if $k \in S$, then $k+1 \in S$ as well,

then it is the case that $S = \mathbb{N}$ so that $P(n)$ is true for each $n \in \mathbb{N}$.

Think of induction as standing up a bunch of dominoes: We set up neighboring dominoes to knock each other over when the first is pushed (step (2) above) and then knock over the first one (step (1) above). Step (1) is called the *base case*, and step (2) is called the *induction step*. The proposition $P(n) = (n \in S \Rightarrow n+1 \in S)$ is called the *induction hypothesis*.

In order to prove this theorem, it is necessary to accept the following two statements as fact:

Theorem 2.2. (Well-ordering of $\mathbb{N}$) Every nonempty subset of $\mathbb{N}$ has a smallest element.

Theorem 2.3. (Predecessors in $\mathbb{N}$) Any element $n \in \mathbb{N}$ not equal to 1 is the successor of some element of $\mathbb{N}$.

Because their proofs lead away from the scope of this course, both of these facts shall be left without proof. It should be noted, though, that well-ordering can be shown to be equivalent to the principle of mathematical induction.

We can now prove Theorem 2.1:

Proof. (By contradiction) Assume the opposite, that there is some set S satisfying (1) and (2) and yet $S \neq \mathbb{N}$. Consider then the smallest element $n_0 \notin \mathbb{S}$. We know $n_0 \neq 1$ since $1 \in S$. Then Theorem 2.3 gives us a predecessor $n_0 - 1$ to n_0.

If $n_0 - 1 \notin S$, then this contradicts our choice of n_0. Assume then that $n_0 - 1 \in S$. This can't be true either since it would force $(n_0 - 1) + 1 = n_0 \in S$, again contradicting our choice of n_0. Hence, Theorem 2.1 must be true. □

Example 2.6. The classical example is to prove the fact that

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

for all $n \in \mathbb{N}$.

Proof. First consider that the proposition sums the first n natural numbers. Hence, the base case is $n = 1$ so that we need to show that 1 equals $\frac{1}{2} \cdot 1 \cdot 2$, which is indeed the case. Therefore, the base case is true.

Now suppose that, for some $n \in \mathbb{N}$, the equation is satisfied. Consider the sum of the first $n + 1$ natural numbers:

$$1 + 2 + \cdots + n + (n+1) = (1 + 2 + \cdots + n) + (n+1).$$

Our induction hypothesis shows that

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

$$= \frac{n^2 + 3n + 2}{2} = \frac{(n+1)(n+2)}{2}.$$

This is exactly the equation we're trying to prove when we take n to be $n + 1$ instead. That is, when we assume that the n^{th} case holds, the $(n+1)^{th}$ case also holds. Therefore, the equation is true for each $n \in \mathbb{N}$. □

We can also apply induction to a different choice of the base case. Of course, this will only prove the proposition for all $n \in \mathbb{N}$ no smaller than the base case.

Example 2.7. "Show that $2^n > 2n + 1$ for integers $n \geq 3$."

Proof. (Base case: $n = 3$) Observe that $2^3 = 8 > 7 = 2 \cdot 3 + 1$. Hence, the statement is true in the base case.

(Induction step) Suppose that $2^n > 2n + 1$ for some $n \geq 3$. Consider then that

$$2^{n+1} = 2 \cdot 2^n > 2(2n + 1) = 4n + 2.$$

Because we're assuming $n \geq 3$, we know that $2n + 2 > 3$. Hence,

$$2^{n+1} > 2n + (2n + 2) > 2n + 3 = 2(n+1) + 1.$$

Thus, the proposition holds for $n + 1$ as well.

Therefore, by mathematical induction, $2^n > 2n + 1$ for all $n \in \mathbb{N}$. □

There is one other form of induction that may be easier to use, especially when we consider recursive relations.

Theorem 2.4. (Strong Induction) Let $S \subseteq \mathbb{N}$ be such that

1. $1 \in S$, and

2. For every $n \in \mathbb{N}$, if $\{1, 2, \ldots, n\} \subset S$, then $n+1 \in S$ as well,

then $S = \mathbb{N}$.

Example 2.8. Recall that the Fibonacci sequence f_n is defined as follows:

$$f_1 = f_2 = 1 \quad \text{and} \quad f_{n+2} = f_{n+1} + f_n, \quad n \geq 3.$$

Prove that

$$f_{n+6} = 4f_{n+3} + f_n$$

for all $n \in \mathbb{N}$.

Proof. (Base case: $n = 1$) Observe first that $f_4 = 3$ and $f_7 = 13$. These indeed satisfy the relation $f_7 = 4f_4 + f_1$ so that the proposition is true for $n = 1$.

(Strong induction step) Assume that the proposition is true for all $k \in \mathbb{N}$ no greater than some $n \in \mathbb{N}$. Consider then that

$$\begin{aligned}
f_{(n+1)+6} &= f_{n+6} + f_{(n-1)+6} \\
&= (4f_{n+3} + f_n) + (4f_{(n-1)+3} + f_{n-1}) \\
&= 4(f_{n+3} + f_{n+2}) + (f_n + f_{n-1}) \\
&= 4f_{(n+1)+3} + f_{(n+1)}.
\end{aligned}$$

Hence, the proposition is also true for $n+1$. We conclude that the proposition is true for all $n \in \mathbb{N}$. $\square$

2.3 Finite and Infinite Sets

One of the things we are going to do frequently in this class is wrestle with the infinite. We are quite familiar with finite objects, but the thing that made Calculus so difficult to study at first was the fact that we hadn't found a way to rigorously deal with infinite objects. We begin by defining what we mean the "size" of a set.

Definition 2.9. Let S be a set. The number of elements in S, denoted by $|S|$ is called the *size/cardinality* of S.

1. The empty set $\emptyset$ has cardinality zero.

2. The set $N_k = \{1, 2, \ldots, k\}$ has cardinality $k \in \mathbb{N}$.

Definition 2.10. Two sets S_1 and S_2 have the same cardinality if there exists a bijection $f\colon S_1 \to S_2$. The function f is sometimes called a *counting function.*

The set S is said to be *finite* if it is empty or has cardinality k for some $k \in \mathbb{N}$.

Otherwise, S is said to be *infinite.*

The idea behind using a counting function is to begin with a set S_1 of known cardinality and then assign to each element of S_2 exactly one element of S_1. Therefore, each element of S_2 is "counted" by a unique element of S_1, and thus a *one-to-one correspondence* is established between the two sets.

Example 2.9. 1. The set $S = \{\exists, \forall, \emptyset, \exists!\}$ of mathematical symbols has cardinality 4 since the function

$$f\colon \mathbb{N}_4 \to S = \{(1,\emptyset),(2,\forall),(3,\exists!),(4,\exists)\}$$

is a bijection from a set of cardinality 4.

2. The set $E = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$ has cardinality 3 since

$$f\colon \mathbb{N}_3 \to E = \{(1,\emptyset),(2,\{\emptyset\}),(3,\{\{\emptyset\}\}\}$$

is a bijection from a set of cardinaliy 3.

3. The set $\mathbb{N}$ is infinite.

Proof. Suppose instead that $\mathbb{N} \neq \emptyset$ is finite. Then there must exist some $k \in \mathbb{N}$ and a bijection $f\colon \mathbb{N}_k \to \mathbb{N}$.

Consider the maximum value M in the range of f. Since $M \in \mathbb{N}$, $M+1 \in \mathbb{N}$ as well. However, by construction, $M+1$ can't be the image of any element of $\mathbb{N}_k$ under f. This shows that f is not a surjection and, hence, not a bijection. This contradiction shows that $\mathbb{N}$ is inifinite. □

An important theorem that we are going to make frequent use of is the following:

Theorem 2.5. Suppose that S and T are sets such that $T \subseteq S$. Then

1. If S is finite, then so is T, and

2. If T is inifinite, then so is S.

Proof. Observe that (2) is the contrapositive of (1). Hence, we merely prove the assertion (1) is true.

If $T =$, then of course it is finite. So we assume that $T \neq \emptyset$ and S is finite. We proceed by induction on the cardinality $|S|$ of S.

(Base case: $|S| = 1$) The only nonempty subset T of S is S itself. Since S is finite, $T = S$ must also be finite.

(Induction step) Suppose that if $|S| = n$, then every subset T of S is finite.

For a set S with $|S| = n+1$, there exists a bijection $f\colon \mathbb{N}_{n+1} \to S$. Assume that $T \subseteq S$. If $f(n+1) \notin T$, then $T \subseteq S_1 = S - \{f(n+1)\}$. Because $|S_1| = n$, the induction hypothesis shows that T is finite.

If instead $f(n+1) \in T$, then $T_1 = T - \{f(n+1)\}$ is a subset of S_1. Hence, the induction hypothesis once again shows that T_1 is finite. As T has one element more than the finite set T_1, T is also finite.

Therefore, the assertion that any subset T of a set S of cardinality n is true for all $n \in \mathbb{N}$, and hence is true for any finite set S. □

Notice that it is an immediate corollary to the theorem that both $\mathbb{Q}$ and $\mathbb{R}$ are infinite as well.

Now is the point where we take a turn into the obscure. We define those sets that we can "enumerate"; that is, those sets that admit a bijection to a subset of $\mathbb{N}$.

Definition 2.11. A set S is said to be *countable* if there exists a subset A of $\mathbb{N}$ to which we can construct a bijection between A and S. If we have $A = \mathbb{N}$, then we say S is *denumerable*.

If the set S is not countable, we call it *uncountable*.

Remark 2.2. It can also be shown that if $T \subseteq S$ and T is uncountable, then S is also uncountable. Hence, the a certain subcase of its contrapositive is true as well: If $T \subseteq S$ and S is denumerable, then T is also denumerable.

Example 2.10. The classical (surprising) examples are listed below.

1. Any finite set is countable.

2. The identity function $f\colon \mathbb{N} \to \mathbb{N}$ given by $n \mapsto n$ is clearly a bijection. Hence, $\mathbb{N}$ is denumerable.

3. Less obvious is the idea that the even naturals $2\mathbb{N} = \{2, 4, 6, \ldots\} \subset \mathbb{N}$, a proper subset with "half" as many elements of $\mathbb{N}$, has the same cardinality as $\mathbb{N}$! How? Consider that the function $f\colon 2\mathbb{N} \to \mathbb{N}$ given by $n \mapsto 2n$ is indeed a bijection between the two sets. The same can be said for the odd naturals.

4. In a similar (shocking) fashion as the above, the integers $\mathbb{Z}$ are also denumerable.

5. The set $\mathbb{N} \times \mathbb{N}$ is denumerable.

   Proof. Observe that each element $x \in \mathbb{N} \times \mathbb{N}$ can be written as (m, n) for some $m, n \in \mathbb{N}$. Consider that for any natural number $N \geq 2$, there are only $N - 1$ ordered pairs $(m, n) \in \mathbb{N} \times \mathbb{N}$ such that $m + n = N$. We informally construct a bijection $f\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ below.

   $N = 2$: First assign $1 \mapsto (1, 1)$.

   $N = 3$: Now assign $2 \mapsto (1, 2)$ and $3 \mapsto (2, 1)$.

   $N = 4$: Next assign $4 \mapsto (1, 3)$, $5 \mapsto (2, 2)$, and $6 \mapsto (3, 1)$.

   We observe a pattern now. At each value of $N \geq 2$, we are able to exhaustively count, in a one-to-one fashion, the number of ordered pairs $(m, n) \in \mathbb{N}$ with $m+n = N$. It isn't hard to convince oneself that for a given $(p, q) \in \mathbb{N} \times \mathbb{N}$, our construction of the bijection f will eventually come across that pair and assign some value k to it so that $f(k) = (p, q)$. Hence, this is also a surjective construction. Therefore, this constitutes a bijection so that $\mathbb{N} \times \mathbb{N}$ is denumerable. □

6. An even more surprising fact: The set of rational numbers $\mathbb{Q}$ is denumerable.

   Proof. First notice that because $\mathbb{N} \subseteq \mathbb{Q}$, $\mathbb{Q}$ can't be finite. We now rule out the possibility that $\mathbb{Q}$ is uncountable.

   Notice that each positive rational number x can be written as $\frac{p}{q}$ for some $(p, q) \in \mathbb{N} \times \mathbb{N}$. However, there are some ordered pairs that yield the same rational number. Hence, $\mathbb{Q}_{\geq}$ is denumerable.

   Finally, it can be shown that the union of two countable sets is countable. Hence, applying successive unions, we find $\mathbb{Q} = (\mathbb{Q}_{\leq} \cup \{0\}) \cup \mathbb{Q}_{\geq}$ must be countable and infinite. That is, $\mathbb{Q}$ is denumerable. □

7. The set of real numbers $\mathbb{R}$ is uncountable.

Proof. Cantor's diagonal proof is one of the classical proofs of this fact. It proceeds by contradiction and uses a brilliantly simple argument. Because we assume that real numbers have a unique decimal expansion, we must acknowledge that we aren't allowing decimal representations concluding with nonterminating nines (things like $0.\overline{9} = 1$ are not allowed).

We prove that $(0, 1)$ is uncountable. By comparison, this will allow us to conclude that $\mathbb{R}$ is also uncountable. Since $\left\{\frac{1}{n+1} \mid n \in \mathbb{N} - \{1\}\right\} \subset (0, 1)$, we know that $(0, 1)$ must be infinite by comparison - the question is just "how big of an inifity are they?" To establish that it's not countably infinite, suppose instead that it is. That is, assume there is a bijection $f : \mathbb{N} \to \mathbb{R}$ so that $(0, 1)$ is denumerable.

The existence of f allows us to "list" the elements of $(0, 1)$. For example, let's say the list begins as follows:

n	$f(n)$
1	$0.\boxed{1}792315648\ldots$
2	$0.7\boxed{0}53298532\ldots$
3	$0.22\boxed{2}2222222\ldots$
4	$0.851\boxed{1}032867\ldots$
5	$0.5000\boxed{0}00000\ldots$
6	$0.12345\boxed{2}6789\ldots$
$\vdots$	$\vdots$

The boxes around the digits indicate the locations where we would like to do something to that number. The idea here is to construct another number in $(0, 1)$. To do so, consider the function

$$g : \{0, 1, 2, \ldots, 9\} \to \{1, 2\}$$

$$g(x) = \begin{cases} 2, & \text{if } x = 1 \\ 1, & \text{if } x \neq 1. \end{cases}$$

We apply g to each of the boxed numbers in the list above - the n^{th} box is the n^{th} digit of $f(n)$ - and use the result to construct our new number. Based on this scheme, the number x that we may construct begins as

$$x = 0.211211\ldots \in (0, 1).$$

Notice that this number can't be equal to any of the numbers in our list since it was designed to be different from the n^{th} number in its n^{th} digit. Therefore, x can't be in the list so that f is not a surjection. Hence, it can't have been a bijection, and our theorem is proved. Both $(0, 1)$ and $\mathbb{R}$ must be uncountably infinite.

□

3 Properties of the Real Numbers

Now that we've gotten through the preliminaries, we can begin to develop some fundamental ideas of Real Analysis. Our goal for this section is to find a suitable characterization and potential definition of the familiar set of real numbers $\mathbb{R}$.

To do so, we need to state which properties we want these numbers to possess. These include their algebraic and ordering properties (which also carry over to the integers and rational numbers as they are both subsets of this larger set). The other key property that we need to establish is the idea of *completeness* - this is primarily what sets $\mathbb{R}$ and $\mathbb{Q}$ apart.

All notes from here on are based on those in the second edition of Ross's *Elementary Analysis: The Theory of Calculus*. This section is draw from the first chapter of this reference.

3.1 The World Outside of the Rational Numbers

The ancient Greeks named the set of positive integers the "natural numbers" since they could be viewed **naturally** as some physical quantity of things. They soon discovered, however, that this set of numbers was insufficient to study algebraic equations - Diophantine equation easily illustrate this inadequacy. In order to remedy this, they decided that not only natural values could be considered. In fact, an even more beautiful set of numbers emerged as ratios of the natural numbers: The so-called rational numbers. These, too, were physically realizable as unitary relations - things like "2 parts water to 5 parts flour" were representable as the ratio of $\frac{2}{5}$ water to flour.

Then they came to the geometry of simple shapes. They easily described the regular polygons, but something very serious troubled them. There were physical quantities that were realizable within these shapes that the rational numbers couldn't describe. The perfect beauty of the rational numbers was threatened to be sullied.

Consider the square with sides of length one. There is absolutely nothing wrong with its construction - it can be made simply with a compass and straightedge. However, we consider constructing a line between two opposite corners of the square. This divides the square into two congruent, isosceles, right triangles with leg lengths one. Label the length of the hypotenuse x. The celebrated theorem of their contemporary Pythagoras dictated that the lengths of these triangles must satisfy the (they assumed Diophantine) equation

$$1^2 + 1^2 = x^2 \quad \Leftrightarrow \quad 2 = x^2 \quad \Leftrightarrow \quad x^2 - 2 = 0.$$

Since they thought that the rational numbers were sufficient to describe nature, the ancient Greeks deduced that there was such a **rational** number (denoted $\sqrt{2}$) whose square was 2.

This equation is not the only one of its kind.

Definition 3.1. A number is said to be *algebraic* if it satisfies a polynomial equation

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 = 0,$$

where $c_0, c_1, \ldots, c_n \in \mathbb{Z}$, $c_n \neq 0$, and $n \in \mathbb{N}$.

We now present the surprising (and almost blasphemous) claim that the positive solution $x = \sqrt{2}$ to the algebraic equation $x^2 - 2 = 0$ is **not** a rational number.

Proposition 3.1. The number $\sqrt{2}$ is not rational.

Proof. Let's suppose instead (as the ancient Greeks did) that $\sqrt{2}$ was rational. Then we may write

$$\sqrt{2} = \frac{p}{q}$$

for some pair of natural numbers p and q. To make this choice unique, we impose the condition that $\gcd(p, q) = 1$. That is, p and q have no common factors apart from 1.

Now consider that we must have $2q^2 = p^2$ after squaring the relation above. Hence, p^2 must be even; and as we've already proven, this is equivalent to saying that p is also even. If this is indeed the case, then $p = 2k$ for some integer k.

Notice now that this implies that $2q^2 = (2k)^2 = 4k^2$. Then the equation $2q^2 - 4k^2 = 2(q^2 - 2k^2) = 0$ and the zero-product property of the integers together show that q^2 must also be even. Hence, q must be even. Therefore both p and q share a factor of 2. This contradicts our assumption that p and q had no factors in common other than 1. Therefore, our assumption that $\sqrt{2}$ was rational turns out to be false. □

This shattered the worldview of some mathematicians (and still is a bit unsettling to some today). **There are other numbers that are not rational**! We call such numbers *irrational.*

Unfortunately, the proof above is quite cumbersome to translate to other similar propositions. There is a saving grace, however, when we wish to show that an algebraic number is irrational.

Theorem 3.1. (Rational Roots) Suppose $c_0, c_1, \ldots, c_n \in \mathbb{Z}$ and $r \in \mathbb{Q}$ satisfying the polynomial equation

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 = 0,$$

where $n \geq 1$, $c_n \neq 0$, and $c_0 \neq 0$. Let $r = \frac{c}{d}$ where $c, d \in \mathbb{Z}$ with $\gcd(c, d) = 1$ and$d \neq 0$. Then c divides c_0 and d divides c_n.

Before we prove this fact, we make one quick remark about its use. This theorem is typically applied as part of a proof by contradiction, since it lists all possible solutions to the polynomial equation. Consider its use in the following examples:

Example 3.1. 1. We can apply the Rational Roots Theorem to our proposition that $\sqrt{2} \neq \mathbb{Q}$.

Proof. The equation we consider is

$$x^2 - 2 = 0,$$

whose coefficients are $c_2 = 1$, $c_1 = 0$, and $c_0 = -2$. Suppose that $\sqrt{2} = \frac{c}{d}$ solves the equation as per the conditions in the theorem.

The theorem states that d divides $c_2 = 1$. The only integers that do this are $d = \pm 1$. The theorem also states that c divides $c_0 = -2$. Hence, the only candidates for rational numbers that solve the equation $x^2 - 2 = 0$ are the numbers $r = \pm 1, \pm 2$. By substituting each value into the equation, we establish that none of them are its roots. Therefore, $\sqrt{2}$ cannot be a rational solution to the equation. Hence, $\sqrt{2} \notin \mathbb{Q}$. □

2. We claim that $a = \sqrt{2 + \sqrt[3]{5}}$ is not rational either.

   Proof. Observe (verify this for yourself) that a solves that algebraic equation

   $$x^6 - 6x^4 + 12x^2 - 13 = 0.$$

   Then we have $c_6 = 1$ and $c_0 = -13$. Hence, the rational roots theorem states that the only possible rational roots to this equation are $r = \pm 1, \pm 13$. It can be verified that indeed none of these values satisfies the equation (there is a slick way to do this for ± 13 - ask me about it in office hours). Therefore, $a \notin \mathbb{Q}$ since none of our rational options could possibly solve the equation. $\square$

Remark 3.1. Observe now that this method of proof to show that a number is irrational only applies to algebraic numbers. In order to do this in a more general setting, there are other methods (like Newtion's method and bifurcating) that can easily accomplish this; however, they are a bit more difficult to use as they involve proving limits of sequences of real numbers.

We now prove the Rational Roots Theorem.

Proof. Since $r = \frac{c}{d}$ solves the polynomial equation, we have the relation

$$c_n \left(\frac{c}{d}\right)^n + c_{n-1}\left(\frac{c}{d}\right)^{n-1} + \cdots + c_1\left(\frac{c}{d}\right) + c_0 = 0.$$

Multiplying both sides by d^n shows that

$$c_n c^n + c_{n-1}c^{n-1}d + c_{n-2}c^{n-2}d^2 + \cdots + c_1 c d^{n-1} + c_0 d^n = 0.$$

Solving for $c_0 d^n$, we find

$$c_0 d^n = -c(c_n c^{n-1} + c_{n-1}c^{n-2}d + c_{n-2}d^2 + \cdots + c_2 c d^{n-2} + c_1 d^{n-1}).$$

Hence, c divides $c_0 d^n$. Since $\gcd(c, d) = 1$, c must divide c_0.

Solving the equation instead for $c_n c^n$, we find

$$c_n c^n = -d(c_{n-1}c^{n-1} + c_{n-2}c^{n-1}d + \cdots + c_1 c d^{n-2} + c_0 d^{n-1}).$$

Hence, d divides $c_n c^n$. Again, because $\gcd(c, d) = 1$, d must divide c_n. $\square$

3.2 Properties of Ordered Fields

We've established that there are other numbers than the rationals that we would like to be able to use. However, just defining the specific elements of the set that we want is overly complicated by technicalities. Even for the rationals, this prospect is quite tedious. We adopt a different approach: Outline the properties that we require them to satisfy. Since the rationals are a poster child for success in algebra, we would like to include them in our construction of the so-called "real numbers." So at the very least, we need this set of real numbers to do the same things that the rationals can and more. We state without proof the "axioms" for the rational numbers as the following two theorems.

Theorem 3.2. (Algebraic Properties of $\mathbb{Q}$) The set $\mathbb{Q}$ of rational numbers is closed under and addition law $+$ and a multiplication law $\cdot$ and satisfies the following additive (A), multiplicative (M), and distributive (DL) properties:

A1) $a + (b + c) = (a + b) + c$ for all a, b, c,
A2) $a + b = b + a$ for all a, b,
A3) $a + 0 = a$ for all a,
A4) For each a, there exists a unique element $-a$ such that $a + (-a) = 0$,

M1) $a(bc) = (ab)c$ for all a, b, c,
M2) $ab = ba$ for all a, b,
M3) $a \cdot 1 = a$ for all a,
M4) For each $a \neq 0$, there exists a unique element a^{-1} such that $aa^{-1} = 1$,

DL) $a(b + c) = ab + ac$ for all a, b, c

Theorem 3.3. (Ordering Properties of $\mathbb{Q}$) The set $\mathbb{Q}$ admits an order structure $\leq$ that satisfies the following properties:

O1) Given a and b, either $a \leq b$ or $b \leq a$,
O2) If $a \leq b$ and $b \leq a$, then $a = b$,
O3) If $a \leq b$ and $b \leq c$, then $a \leq c$,
O4) If $a \leq b$, then $a + c \leq b + c$,
O5) If $a \leq b$ and $0 \leq c$, then $ac \leq bc$.

Remark 3.2.

1. The properties (A1) and (M1) are called the *associative laws*, and (A2) and (M2) are called the *commutative laws*.

   Properties (A3) and (M3) demonstrate the effect of addition and multiplication by the *additive identity* 0 and the *multiplicative identity* 1, respectively.

   Properties (A4) and (M4) establish the existence of *additive* and *multiplicative inverses*, respectively.

   Any set satisfying all nine of these algebraic properties is called a *field*. The rationals are the model for this kind of algebraic set, but there are plenty of other examples of fields.

2. Property (O3) is called the *transitive law*, and it is the critical comparison property that an ordering satisfies.

   A field that admits and ordering that satisfies all five of these ordering properties is called an *ordered field*.

When proving facts about an ordered field, one must be careful not to think his or her work as obvious or legal. The only properties that we are allowing this object to satisfy are those outlined above.

Observe that we still haven't defined the set of real numbers $\mathbb{R}$ yet. This is indeed intentional. The construction of $\mathbb{R}$ is traditionally done in one of two ways. Each method has is advantages and disadvantages.

When constructed as *Dedekind cuts*, the idea $\mathbb{R}$ "fills in the gaps" in $\mathbb{Q}$ is quite easily understood as this is built into the definition. However, the algebraic and ordering properties become horribly obscured so that nothing in the construction is familiar to the student.

When constructed axiomatically as we have described, the algebraic and ordering properties are quite clearly understood and intuitive to the student. The trade-off here is in understanding these "gaps" in $\mathbb{Q}$. Let's demonstrate the ease of this method by deducing some familiar properties of ordered fields; in particular, those of $\mathbb{Q}$ and what we are going to call $\mathbb{R}$.

Theorem 3.4. For all a, b, c in an ordered field:

(a) $a + c = b + c$ implies that $a = b$,

(b) $a \cdot 0 = 0$,

(c) $(-a)b = -(ab)$,

(d) $(-a)(-b) = ab$,

(e) $ac = bc$ and $c \neq 0$ implies $a = b$,

(f) $ab = 0$ implies that either $a = 0$ or $b = 0$.

Proof. (a) Suppose indeed that $a+c = b+c$. Then add $-c$ to both sides of this equation:

$$(a + c) + (-c) = (b + c) + (-c).$$

Properties (A1), (A4), and (A3) (in that order) allow us to continue the computation so that

$$a + (c + (-c)) = b + (c + (-c)) \quad \Leftrightarrow \quad a + 0 = b + 0 \quad \Leftrightarrow \quad a = b,$$

as was to be shown.

(b) Use (a) to establish this. Consider that (A3) and (DL) yield

$$a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0.$$

That is,

$$0 + a \cdot 0 = a \cdot 0 + a \cdot 0$$

so that (a) implies that $0 = a \cdot 0$

(c) (A4) and (DL) together give us

$$0 = 0 \cdot b = (a + (-a)) \cdot b = ab + (-a)b.$$

Hence, $(-a)b$ behaves like the unique additive inverse $-(ab)$ for the element ab. Therefore, $(-a)b = -(ab)$.

(d) In the same fashion as (c), notice that

$$0 = 0 \cdot (-b) = (a + (-a)) \cdot (-b) = a(-b) + (-a)(-b) = -(ab) + (-a)(-b),$$

where the last equality is an application of (c) above.

(e) Supposing that $ac = bc$, we see that

$$(ac)c^{-1} = (bc)c^{-1} \quad \Leftrightarrow \quad a(cc^{-1}) = b(cc^{-1}) \quad \Leftrightarrow \quad a \cdot 1 = b \cdot 1 \quad \Leftrightarrow \quad a = b,$$

as was to be shown.

(f) Suppose indeed that $ab = 0$. Without loss of generality (WLOG), suppose also that $b \neq 0$. Then b^{-1} exists. Hence, we can apply (e) to the equation $ab = 0 \cdot b$ so that $a = 0$. Therefore, the assertion is also true when for all cases of a and b.

□

Theorem 3.5. For all a, b, c in an ordered field:

(a) If $a \leq b$, then $-b \leq -a$,

(b) If $a \leq b$ and $c \leq 0$, then $bc \leq ac$,

(c) If $0 \leq a$ and $0 \leq b$, then $0 \leq ab$,

(d) $0 \leq a^2$,

(e) $0 < 1$ in a field with $0 \neq 1$,

(f) If $0 < a$, then $0 < a^{-1}$,

(g) If $0 < a < b$, then $0 < b^{-1} < a^{-1}$.

Define $a < b$ as meaning $a \leq b$ and $a \neq b$.

Proof. (a) Supposing that indeed $a \leq b$, we use (O4) and a combination of (A1)-(A4) to show that

$$a + [(-a) + (-b)] \leq b + [(-a) + (-b)] \quad \Leftrightarrow \quad (a + (-a)) + (-b) \leq (-a) + (b + (-b))$$

$$\Leftrightarrow \quad 0 + (-b) \leq 0 + (-a) \quad \Leftrightarrow -b \leq -a,$$

as was to be shown.

(b) By (a), $0 \leq -c$ so that we may apply (O5) to $a \leq b$: $a(-c) \leq b(-c)$. By the previous theorem, this is exactly $-(ac) \leq -(bc)$. Another application of (i) shows that $bc \leq ac$ as required.

(c) Suppose that $0 \leq a$ and $0 \leq b$. Another application of (i) where we take 'a'$= 0$ in "$a \leq b$." Hence, $0 \cdot b \leq a \cdot b$ so that $0 \leq ab$.

(d) If $0 \leq a$, we merely apply (iii) to get $0 \leq a^2$. If $a \leq 0$ instead, then $0 \leq -a$ so that (iii) implies $0 \leq (-a)^2$. The previous theorem indicates this is $0 \leq a^2$ as well.

(e) Suppose instead that $1 \leq 0$. Then $0 \leq -1$ so that (iii) shows that $0 \leq 1$. Then (O2) implies $0 = 1$, which is not allowed. Hence, the opposite statement that $0 < 1$ is true.

(f) Suppose that $0 < a$ but instead that $a^{-1} \leq 0$. Then $0 \leq -a^{-1}$ so that (iii) tells us $0 \leq a(-a^{-1}) = -(aa^{-1}) = -1$. That is, $1 \leq 0$. This contradiction to (e) shows that if $0 < a$, then $0 < a^{-1}$ as well.

(g) Suppose indeed that $0 < a < b$. That is, $0 < a$, $0 < b$ and $a < b$. From (f), we know that both $0 < a^{-1}$ and $0 < b^{-1}$. For the final relation, consider that (iii) shows that $0 < a^{-1}b^{-1}$. Then (O5) shows that

$$a(a^{-1}b^{-1}) < b(a^{-1}b^{-1}) \quad \Leftrightarrow \quad b^{-1} < a^{-1}.$$

Hence, we certainly have the relations $0 < b^{-1} < a^{-1}$.

□

We now take a quick detour to acknowledge the familiar idea of the absolute value, even though we haven't finished our construction of $\mathbb{R}$.

Definition 3.2. Define the function $|\cdot|\colon \mathbb{R} \to \mathbb{R}$ as

$$|x| = \begin{cases} x, & 0 \leq x \\ -x, & x < 0. \end{cases}$$

The quantity $|x|$ is called the *absolute value of* x.

The student should be familiar with the geometric intuition of this function: $|x|$ measures how far away x is from 0. That is, it returns the **distance** from 0 to x.

Definition 3.3. For $a, b \in \mathbb{R}$ we define $\mathrm{dist}(a, b) = |a - b|$ to be the *distance between a and b.*

We end section by proving some basic properties of this absolute value function.

Theorem 3.6. For all $a, b \in \mathbb{R}$:

(a) $|a| \geq 0$ with equality iff $a = 0$,

(b) $|ab| = |a| \cdot |b|$,

(c) (Triangle inequality) $|a + b| \leq |a| + |b|$.

Proof. Since the definition of $|\cdot|$ is piecewise, one might expect that the proofs of these facts be divided into cases... and one would be absolutely correct!

(a) If $0 \leq a$, then $0 \leq |a| = a$ by definition. If $a < 0$, then $0 \leq -a = |a|$ again by its definition.

To show the second assertion, consider the easy case first: $a = 0$ gives $|a| = 0$. Now suppose that $|a| = 0$. Observe that since $|a| = 0$ is equivalent to $0 \leq |a|$ and $|a| \leq$ together, we have $0 \leq |a|$ implies $0 \leq a$ and that $|a| \leq 0$ implies $a \leq 0$. Therefore, $a = 0$ as well.

(b) If $0 \leq a, b$, then $0 \leq ab = |a| \cdot |b|$.

If $a, b \leq 0$, then $0 \leq -a, -b$ so that $0 \leq (-a)(-b) = |a| \cdot |b|$.

WLOG, take $0 \leq a$ and $b \leq 0$ (the remaining case is similarly proven). Then $0 \leq -b = |b|$ and $0 \leq a(-b)$ so that $|a| \cdot |b| = a(-b) = -(ab) = |ab|$.

(c) Since either $|a| = a$ or $|a| = -a$, we have $-|a| \le a \le |a|$. The same is true for b. Then successive applications of (O4) show that

$$-|a| + (-|b|) \le a + b \le |a| + b \le |a| + |b|.$$

Therefore,

$$-(|a| + |b|) \le a + b \le |a| + |b|.$$

Then we have $|a + b| \le |a| + |b|$ as was to be shown.

$\square$

3.3 The Completeness Axiom

Now that we've establish all the algebraic and order properties that we want the real numbers to satisfy, we turn to the harder task of showing that $\mathbb{R}$ "fills in the gaps" in $\mathbb{Q}$. We do this by guaranteeing a certain property about nonempty subsets of $\mathbb{R}$ is satisfied. The following two definitions let up what we need for this goal.

Definition 3.4. Let S be a nonempty subset of $\mathbb{R}$.

If S contains a largest element M, then we call M the *maximum of S* and write $M = \max S = \max\limits_{s \in S} s$.

If S contains smallest element m, then we call m the *minimum of S* and write $m = \min S = \min\limits_{s \in S} s$.

Definition 3.5. Let S be a nonempty subset of $\mathbb{R}$.

If a real number M satisfies $s \le M$ for all $s \in S$, then M is called and *upper bound of S*. If such an M exists, we say that S is *bounded above*.

If a real number m satisfies $m \le s$ for all $s \in S$, then m is called and *lower bound of S*. If such an m exists, we say that S is *bounded below*.

The set S is said to be *bounded* if it is both bounded above and bounded below. That is, if there exist real numbers m and M such that $S \subseteq [m, M]$. If S is not bounded, we say it is *unbounded*.

Example 3.2.

1. Every finite, nonempty subset S of $\mathbb{R}$ has a maximum and minimum. For a specific example, consider $S = \{-2, 0, 500, \pi, \frac{1}{e}\}$. We see that the maximum of S is 500, and the minimum of S is -2.

   Any real number no smaller than the maximum of the S is an upper bound for S. Similarly, any real number no greater than the minimum of S is a lower bound of S. For the same specific S above, 2000 is an upper bound and -50 is a lower bound.

2. The familiar intervals $(a, b), (a, b], [a, b), [a, b] \subset \mathbb{R}$ with $a < b$ are all bounded. The sets $(a, b]$ and $[a, b]$ both have a maximum of b, and the sets $[a, b)$ and $[a, b]$ both have a minimum of a.

   Notice that the interval (a, b) has no maximum or minimum. However, all four intervals have the same upper and lower bounds.

3. Our fundamental sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are unbounded. However, $\mathbb{N}$ has a minimum of 1 but no maximum.

4. Consider the set $S = \{r \in \mathbb{Q} \mid 0 \leq r \leq \sqrt{2}\}$. This has a minimum of 0, but no maximum since $\sqrt{2} \notin S$. We shall see shortly that there are values that are rational numbers arbitrarily close to $\sqrt{2} \not\in \mathbb{Q}$.

Definition 3.6. Let S be a nonempty subset of $\mathbb{R}$.

If S is bounded above and S has a least upper bound, then we call it the *supremum of* S and denote it by $\sup S = \sup\limits_{s \in S} s$.

If S is bounded below and S has a greatest lower bound, then we call it the *infimum of* S and denote it by $\inf S = \inf\limits_{s \in S} s$.

Remark 3.3. 1. Though similar to the maximum and minimum values of a set, the supremum and infimum are defined even if they don't belong to the set. It is easy to see that if S is bounded above, the $M = \sup S$ iff (1) $s \leq M$ for all $s \in S$ and (2) $M_1 < M$ implies there exists $s_1 \in S$ such that $s_1 > M_1$.

2. When showing that a number $s_0 = \sup S$ for some set S, we typically do this in two steps. First we show that that s_0 is indeed an upper bound for S in the first place. The second step amounts to assume that we have a second upper bound t, and we then demonstrate that $s_0 \leq t$. A similar technique is used for proving the infimum of S is a specific value.

Example 3.3. 1. Sets with a maximum have a supremum equal to that maximum. Sets with a minimum has an infimum equal to that minimum.

2. The suprema and infima of the open, closed, and half-open intervals from the previous example are all the same.

3. For the set $S = \{r \in \mathbb{Q} \mid 0 \leq r \leq \sqrt{2}\}$, we have $\inf S = 0$ and $\sup S = \sqrt{2}$.

4. The set $A = \{r \in \mathbb{Q} \mid r^5 \leq 3\}$ bounded above, but not below. We have $\sup A = \sqrt[5]{3}$, but there is no infimum.

5. The set $B = \{x \in \mathbb{R} \mid x^2 < 7\} = (-\sqrt{7}, \sqrt{7})$ is bounded above and has $\sup B = \sqrt{7}$. Also, B is bounded below and has $\inf B = -\sqrt{7}$.

6. The set $C = \{-1, 2, \frac{-1}{3}, 4, \frac{-1}{5}, 6, \ldots\}$ is bounded below by -1 and has $\inf C = -1$; however, there is no upper bound and no supremum.

Observe that every nonempty subset of $\mathbb{R}$ that we've considered so far that is bounded above also has a supremum. This is no coincidence. This is the property that sets $\mathbb{R}$ apart from $\mathbb{Q}$: Completeness.

Theorem 3.7. (Completeness Axiom) Every nonempty subset S of $\mathbb{R}$ that is bounded above has a least upper bound. In other words, $\sup S$ exists and is a real number.

The fact that a supremum exists for any nonempty subset that is bounded above may not be too surprising, but the thing that makes this interesting is that this supremum may wind up existing outside of the rational numbers! For example, given $S = \{r \in \mathbb{Q} \mid 0 \leq r^2 \leq 2\} \subset \mathbb{Q}$, there is no supremum in $\mathbb{Q}$. However, if we instead assume that $S \subset \mathbb{R}$, then the supremum exists and is $\sup S = \sqrt{2}$. Then we finally have the formal definition of $\mathbb{R}$.

Definition 3.7. The set $\mathbb{R}$ of *real numbers* is the ordered field whose elements are all the possible upper bounds to non empty subsets of $\mathbb{Q}$ that are bounded above.

Corollary 3.1. Every nonempty subset S of $\mathbb{R}$ that is bounded below has a greatest lower bound $\inf S$ that is real.

Proof. Define $-S = \{-s \mid s \in S\}$. Then since S was bounded below, $-S$ must be bounded above. Because $S \neq \emptyset$, the Completeness Axiom shows that $\sup -S$ exists. We just need to show now that $\boxed{\inf S = -\sup(-S)}$.

Let $s_0 = \sup -S$. We claim that

$$-s_0 \leq s, \ \forall s \in S,$$

and that if $t \leq s$ for all $s \in S$, then $t \leq -s_0$.

Since $s_0 = \sup -S$, we know that for all $-s \in -S$, $-s \leq s_0$ so that $-s_0 \leq s$. Therefore, $-s_0$ is a lower bound for S.

Suppose that t is another lower bound for S, $t \leq s$ for all $s \in S$. Then $-s \leq -t$ for all $-s \in -S$ shows that $-t$ is an upper bound for $-S$. Since $s_0 = \sup -S$, we know that $-s_0 \leq -t$. Hence, $t \leq s_0$ so that s_0 is the greatest among all lower bounds for S. Therefore, $\inf S = \sup -S$. □

We now show some foundational consequences to the Completeness Axiom - ones that relate the real numbers to the naturals and rationals. In particular, we need to establish the Archimedean Property and (more importantly) the Density of the rationals in $\mathbb{R}$. The first suggests the real numbers don't "extend beyond the reach" of the natural numbers - the natural numbers can "sneak past and underneath" them if we allow them to take on large enough values. The second makes rigorous the idea that the real numbers "fill in the gaps" of the rational numbers. Together these two properties show that the real numbers truly extend the properties of the integers and rationals to be able to accomplish more.

Theorem 3.8. (Archimedean Property) If $a, b > 0$, then for some positive integer n we have $na > b$.

Proof. Suppose intstead that this is not possible. That is, asuume there is some pair of reals $a, b > 0$ where $an \leq b$ for all $n \in \mathbb{N}$. Then consider that the set

$$S = \{na \mid n \in \mathbb{N}\}$$

is nonempty and bounded above by b, hence we can apply the completeness axiom to it. Then $\sup S = s_0 \in \mathbb{R}$ exists.

Observe now that because $a > 0$, we must has $s_0 - a < s_0$. Because s_0 is the **least** upper bound, $s_0 - a$ cannot be an upper bound for S. As cuh, there must be some $N \in \mathbb{N}$ such that $s_0 - a < Na \in S$. Therefore, $s_0 < Na + a = (N+1)a \in \mathbb{S}$. That is, s_0 is also **not** a lower bound for S. This contradiction shows that our assumption that the Archimedean property fails must be false. □

Theorem 3.9. (Density of $\mathbb{Q}$ in $\mathbb{R}$) If $a, b \in \mathbb{R}$ and $a < b$, then there is an $r \in \mathbb{Q}$ such that $a < r < b$.

Proof. Given that the rational number r we are searching for has the form $r = \frac{m}{n}$ for some $m, n \in \mathbb{Z}$ with $n > 0$, we are really trying to show that

$$an < m < bn.$$

Because $a < b$, we have $b - a > 0$ so that the Archimedean property guarantess the existence of an $n \in \mathbb{N}$ such that $n(b - a) > 1$. Hence, we have found an $n \in \mathbb{N}$ such that

$$an + 1 < bn.$$

Intuition dictates that there must be an integer between an and bn since they are more than 1 apart. Intuition does not constitute a proof, though, so we need to prove that such an integer exists.

By the Archimedean property, there exists an integer $k > \max\{|an|, |bn|\}$ satisfying the chain of inequalities

$$-k < an < bn < k.$$

Let $K = \{j \in \mathbb{Z} \mid -k \leq j \leq k\}$ and $S = \{j \in K \mid an < j\}$. It is evident that K is finite, so $J \subseteq K$ must also be finite. Then define $m = \min S$ - we now show that this is the m we are looking for.

This m gives us the chain of inequalities $-k < an < m$. Because $m > -k$, $m - 1 \in K$ so that $an < m - 1$ is false by our choice of m. Then $m - 1 \leq an$ combines with our first observation that $an + 1 < bn$ to yield

$$m \leq an + 1 < bn.$$

Thus, $an < m$ and $m < bn$ as was to be shown. $\square$

It should be noted that the proof above hinges on the Archimdean property twice; hence, it depends on the completeness axiom. There are ordered fields that do not satisfy the completeness axiom so that we can't prove either of these last two theorems about them. This is what sets $\mathbb{R}$ apart as an ordered field. It is big enough to to everything that $\mathbb{Q}$ can, and yet it satisfies the Archimedean property and allows for rational numbers to fit in wherever possible. It should be noted that the same argument for $\mathbb{Q}$ to be dense in $\mathbb{R}$ can be extended to irrational numbers as well: For every pair of rational numbers $a < b$, there is an **irrational** number between them.

3.3.1 A Brief Interlude for the Infinities

Now is a great time to say a word about the idea of a positive and negative infinity and their relation to $\mathbb{R}$. I will say so loudly and clearly right now:

The symbols $+\infty$ and $-\infty$ are NOT real numbers!

Did you hear me? Good! $\smile$

That being said, the symbols $\pm\infty$ don't satisfy the algebraic axioms that $\mathbb{R}$ does. However, this doesn't mean $\pm\infty$ can't have any semblance of ordering properties. We can append them to $\mathbb{R}$ to create the set of *extended real numbers*

$$\widehat{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

and allow these two extra symbols to constitute the universal upper and lower bounds:

$$-\infty \leq a \leq +\infty \quad \forall a \in \widehat{\mathbb{R}}.$$

This means that **as an ordered set** $\widehat{\mathbb{R}}$ satisfies (O1), (O2), and (O3). This also allows us access to their familiar use in Calculus; though, again, $\widehat{R}$ should not be allowed any algebraic properties for mathematical consistency. Therefore, **theorems about** $\mathbb{R}$ **don't apply to** $\widehat{\mathbb{R}}$. Hence, we will have to redefine the notions supremum and infimum in $\mathbb{R}$ to extend to $\widehat{\mathbb{R}}$.

Definition 3.8. For a subset $S \subseteq \mathbb{R}$ that is not bounded above, we define the *supremum of S* to be $\sup S = +\infty$.

For a subset $S \subseteq \mathbb{R}$ that is not bounded below, we define the *infimum of S* to be $\inf S = -\infty$.

Unless we explicitly state a result applies for $\widehat{\mathbb{R}}$, the student should assume that all definitions, theorems, and proofs are valid only in $\mathbb{R}$.

4 Sequences of Real Numbers

One of the primary tools of Analysis is the *sequence*. Many, if not all, major definitions can be rephrased in terms of them and, in many cases, are much simpler to use in practice. In this chapter, we examine these objects in careful detail and make use of the completeness of $\mathbb{R}$ to establish different characterizations of what we mean for them to *converge*.

4.1 Convergence Preliminaries

Let's first make the definition of such an object rigorous.

Definition 4.1. Let X be a set of objects. A function $s\colon \mathbb{N} \to X$ is called a *sequence* and is denoted (s_n), where s_n is traditional shorthand for $s(n)$. When $X = \mathbb{R}$, we say that (s_n) is a *real-valued* (or just *real*) sequence.

That is, a sequence is a function whose domain is the natural numbers. It can even be stated so that its domain is a well-ordered (each nonempty subset has a smallest element), countably infinite set. Unless otherwise stated, all sequences mentioned hereafter will be real-valued.

Example 4.1. The following are numerous ways that we can define sets:

1. We can list elements with a clear pattern:
$$(s_n) = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\}.$$
In such cases the domain is implicitly taken to be $\mathbb{N}$.

2. We can create a set with elements enumerated by the natural numbers:
$$(s_n) = \{(-1)^n \mid n \in \mathbb{N}\}.$$

3. We can define a sequence recursively:
$$s_1 = 2 \quad \text{and} \quad s_{n+1} = \frac{1 + s_n}{2}, \quad \text{for } n \in \mathbb{N}.$$

4. To acknowledge the remark before the example, we can begin indexing the sequence at a number other than $n = 1$:
$$(s_n) = \left\{\frac{1}{\ln(\ln(\ln n))} \mid n \in \mathbb{N} - \{1, 2\}\right\}.$$

5. As a (perhaps not so) surprising example, the familiar series from Calculus can be thought of as a "limit" (whatever THAT is...) of a sequence of partial sums:
$$(s_n) = \left(\sum_{k=0}^{n} \frac{(-1)^k}{k^2}\right).$$

6. We can also use familiar functions to aid us in describing terms of a sequence:

$$(s_n) = (\sin n + \cos n).$$

Again, when it is not specified, the domain of the sequence is assumed to be $\mathbb{N}$.

We now come to our first major obstacle in Analysis: How do we describe the eventual/limiting behavior of a sequence? Our obvious goal is to find a good measure of what it means for a sequence to "settle around" or "converge to" a particular value. Intuitively this notion is obvious: The values of the sequence **eventually** start to **get close** to that value that we wish to call the *limit* of the sequence. Our intuitive thoughts don't constitute rigor, however, since they are too vague to be developed into propositions that we can satisfy. The motivation for being "close enough" to that limiting value is found in the following proposition:

Theorem 4.1. (No smallest, positive real) Suppose that $a \in \mathbb{R}$ satisfies the following condition:

$$0 \leq a < \epsilon \quad \text{for all real} \quad \epsilon > 0.$$

Then it must be the case that $a = 0$.

Proof. The proof is simple enough when done by contradiction. Suppose that the ϵ-condition holds but instead that $a > 0$. Then observe that $\epsilon = \frac{1}{2}a > 0$ is a value for which a does not satisfy the ϵ-condition. This contradiction shows that $a = 0$ as desired. □

All we need to do now is establish the idea of this condition happening "eventually." Perhaps we mean that for a given ϵ, **after some point in the sequence, all terms beyond will satisfy the condition**. This will serve as our definition of being "close enough" or "abritrarily close" (don't say "inifinitely close") to a value:

Definition 4.2. A sequence (s_n) is said to *converge* to $L \in \mathbb{R}$ if

for each $\epsilon > 0$ there exists a real N such that if $n \geq N$, then $|s_n - L| < \epsilon$.

This is more compactly written as $\lim_{n\to\infty} s_n = L$ or, even more simply, $s_n \to L$. The condition with the ϵ will be referred to in these notes as the *ϵ-condition.*

The number L is called the *limit* of the sequence (s_n).

A sequence that fails to satisfy the ϵ-condition for any real number is said to *diverge.* It is said for such a sequence that its limit *does not exist.*

Remark 4.1. 1. In plain English, this statement says

> "for however close ϵ I want to make the term s_n be to the limit L, there comes a point N in the sequence where each element after s_N is within ϵ of L."

That is to say that each ϵ gets its own N that satisfies this ϵ-condition - so N can be thought of as a function of ϵ. So the ϵ-condition is an infinite number of statements - one for each positive real number ϵ.

2. The "eventually" clause of our definition taken care of by the Archimedean property of $\mathbb{R}$. Whichever N we come up with in this definition, there will always be infinitely many $n \in \mathbb{N}$ that are guaranteed to be greater then N. Hence, we can simply take N to be a natural number.

3. We should also take a moment to mention that the ϵ-condition can be relaxed down to "$|s_n - L| \leq \epsilon$ from the strict inequality. This is due to the fact that if we can show the weak inequality for $\frac{\epsilon}{2}$, then we can extend to the strict inequality by merely noting that $\frac{\epsilon}{2} < \epsilon$.

Let's now see how this definition is applied to the first two examples we outlined above:

Example 4.2. 1. We would like to show that $\left(\frac{1}{n}\right) \to 0$.

Proof. If we are given $\epsilon > 0$, we can choose $N = \left\lceil \frac{1}{\epsilon} \right\rceil$ (the smallest integer greater than $\frac{1}{\epsilon}$) to satisfy the ϵ-condition. Supposing that $n \geq N$, we can see that

$$|s_n - L| = \left|\frac{1}{n} - 0\right| = \left|\frac{1}{n}\right| = \frac{1}{n} \leq \frac{1}{N} < \frac{1}{1/\epsilon} = \epsilon.$$

$\square$

2. We claim that $(s_n) = ((-1)^n)$ diverges.

Proof. Suppose instead that there exists some number $L \in \mathbb{R}$ such that $s_n \to L$. Then consider that for $\epsilon = \frac{1}{2}$ there must exist an $N \in \mathbb{N}$ such that $|s_n - L| < \frac{1}{2}$ for all $n \geq N$.

Now consider that $|s_{n+1} - s_n| = |\pm 2| = 2$ for any $n \in \mathbb{N}$. The triangle inequality allows us to split this absolute value through the limit L at the cost of an inequality:

$$|s_{n+1} - s_n| = |s_{n+1} - L + L - s_n| \leq |s_{n+1} - L| + |L - s_n|.$$

If we take $n = N$, we see that

$$2 = |s_{N+1} - s_N| \leq |s_{N+1} - L| + |s_N - L| < \frac{1}{2} + \frac{1}{2} = 1.$$

This contradiction shows us that (s_n) diverges. $\square$

Who's to say that a sequence can't have more than one limit? Maybe there's a sequence that has two... or three... or worse, an infinite number of limits! The following proposition prevents such an occurrence:

Proposition 4.1. (Limits are unique) Suppose a sequence (s_n) converges to two real numbers L_1 and L_2. Then we must have $L_1 = L_2$.

Proof. Suppose indeed that $s_n \to L_1$ and $s_n \to L_2$. Let $\epsilon > 0$. Then there exist $N_1, N_2 \in \mathbb{N}$ such that $n \geq N_1$ implies $|s_n - L_1| < \frac{\epsilon}{2}$ and $n \geq N_2$ implies $|s_n - L_2| < \frac{\epsilon}{2}$. Hence, for $N = \max\{N_1, N_2\}$, $n \geq N$ implies both that $|s_n - L_1| < \frac{\epsilon}{2}$ and $|s_n - L_2| < \frac{\epsilon}{2}$ simultaneously.

Now consider for $n \geq N$ we have

$$|L_1 - L_2| = |L_1 - s_n + s_n - L_2| \leq |L_1 - s_n| + |s_n - L_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

By the proposition shown earlier, this implies that $L_1 - L_2 = 0$. That is, $L_1 = L_2$. $\square$

4.2 How to Find the N

In order to show that a sequence (s_n) converges to a limit L, our goal is simple:

Given an $\epsilon > 0$, find some N where $n \geq N$ implies $|s_n - L| < \epsilon$.

We just need to get our hands on that N. This is usually much easier said than done, but it can be demystified by walking through some examples of how a formal proof of the ϵ-condition is written.

Example 4.3. Suppose we were asked to show that $\left(\frac{1}{n^3}\right) \to 0$. We need to give ourselves an $\epsilon > 0$ and try to come up with an N satisfying the ϵ-condition. So we need to do a little bit of scratch work in order to find a good candidate for N - **this scratch work should never go into a formal proof**!

If we had an $\epsilon > 0$, then we need our $n \geq N$ to satisfy the following chain of inequalities:

$$|s_n - L| = \left|\frac{1}{n^3} - 0\right| = \frac{1}{n^3} \leq \frac{1}{N^3} < \epsilon.$$

That is, we slowly work the n terms out of the inequalities until only N terms remain. Then we try to put an ϵ bound on N. To do so, formally solve

$$\frac{1}{N^3} < \epsilon \quad \Leftrightarrow \quad \frac{1}{\epsilon} < N^3 \quad \Leftrightarrow \quad \frac{1}{\sqrt[3]{\epsilon}} < N.$$

We can satisfy the $|s_n - L|$ inequalities by choosing **any** $N \in \mathbb{N}$ that satisfies the condition $N > \frac{1}{\sqrt[3]{\epsilon}}$. We choose $N = \left\lceil \frac{1}{\sqrt[3]{\epsilon}} \right\rceil$. We are now ready to construct a formal proof.

Proof. Let $\epsilon > 0$ (almost every formal proof of the ϵ-condition begins this way). Choose $N = \frac{1}{\sqrt[3]{\epsilon}}$. Then for any $n \geq N$, consider that

$$|s_n - L| = \left|\frac{1}{n^3} - 0\right| = \frac{1}{n^3} \leq \frac{1}{N^3} \leq \frac{1}{(1/\sqrt[3]{\epsilon})^3} = \epsilon.$$

Hence, $\frac{1}{n^3} \to 0$. $\square$

Observe how long the formal proof is compared to the scratch work before it. The formal proof really isn't all that long. Why? It's an existence proof - we just need to find the N and show that it's sufficient to prove what we want it to prove.

Example 4.4. Let's show that $\displaystyle\lim_{n\to\infty} \frac{2n^4 - 5n^2}{n^4 - 3} = 2$.

We examine the $|s_n - L|$ term to see if it simplifies in any "nice" way:

$$|s_n - L| = \left|\frac{2n^4 - 5n^2}{n^4 - 3} - 2\right| = \left|\frac{2n^4 - 5n^2}{n^4 - 3} - \frac{2(n^4 - 3)}{n^4 - 3}\right| = \left|\frac{6 - 5n^2}{n^4 - 3}\right| = \left|\frac{5n^2 - 6}{n^4 - 3}\right|.$$

We notice that this final expression vaguely resembles $\frac{n^2}{n^4} = \frac{1}{n^2}$. Our goal now becomes "find appropriate bounds on the numerator and denominator to be able to force an upper bound that looks like a constant multiple of $\frac{1}{n^2}$."

First notice that $5n^2 - 6 \leq 5n^2$ for all $n \in \mathbb{N}$. We're lucky at this point since we don't need to impose any restrictions on our choice of N. That's not always the case, as demonstrated below.

Now look to the denominator where we will have to find a **lower** bound to compare against the denominator in the expression above. Observe that $n^4 - 2 \geq \frac{1}{2}n^4$ for $n \geq 2$. This now changes what we can allow N to be since we need to guarantee this lower bound for $n^4 - 2$.

Using both of these observations, we find that if $n \geq N \geq 2$, then

$$|s_n - L| = \left|\frac{5n^2 - 6}{n^4 - 3}\right| \leq \left|\frac{5n^2}{\frac{1}{2}n^4}\right| = \left|\frac{10}{n^2}\right| \leq \left|\frac{10}{N^2}\right| < \epsilon$$

Hence, we need

$$\frac{10}{N^2} < \epsilon \quad \Leftrightarrow \quad \frac{10}{\epsilon} < N^2 \quad \Leftrightarrow \quad \sqrt{\frac{10}{\epsilon}} < N.$$

We can choose any $N \geq 2$ that satisfies this final inequality. Let's just choose $N = \max\left\{2, \left\lceil\sqrt{\frac{10}{\epsilon}}\right\rceil\right\}$. We can now construct a formal proof.

Proof. Let $\epsilon > 0$. Choose $N = \max\left\{2, \left\lceil\sqrt{\frac{10}{\epsilon}}\right\rceil\right\}$. Then for all $n \geq N$, we have

$$|s_n - L| = \left|\frac{2n^4 - 5n^2}{n^4 - 3}\right| = \left|\frac{5n^2 - 6}{n^4 - 3}\right| \leq \left|\frac{10}{n^2}\right| \leq \frac{10}{N^2} < \frac{10}{(\sqrt{10/\epsilon})^2} = \epsilon.$$

□

As much as I hate to use examples straight out of the textbook, Ross has used a rather classical example problem that I, too, would like to examine. However, he leaves some details to the reader. I think it more illustrative to show those details here so that the student can emulate the techniques and strategies later.

Example 4.5. Suppose (s_n) is a sequence of nonnegative real numbers that converges to $L \in \mathbb{R}$.

Claim: The limit is nonnegative as well. That is, $L \geq 0$.

Proof. Suppose to the contrary that $L < 0$. Then for $\epsilon = \frac{-L}{2} > 0$, we see that for n sufficiently large

$$|s_n - L| = s_n - L < \epsilon = \frac{-L}{2} \quad \Leftrightarrow \quad s_n < \frac{L}{2} < 0.$$

Hence, s_n wouldn't be nonnegative after a certain point. This contradiction indicates that $L \geq 0$. □

Now that we've established that all terms involved are nonnegative, we would like to show that we can apply a square root to all of them. That is:

Claim: It is the case that $\sqrt{s_n} \to \sqrt{L}$.

Remember that we need some scratch work first. We recall a useful "algebraic escape":

$$\sqrt{s_n} - \sqrt{L} = \frac{(\sqrt{s_n} - \sqrt{L})(\sqrt{s_n} + \sqrt{L})}{\sqrt{s_n} + \sqrt{L}} = \frac{s_n - L}{\sqrt{s_n} + \sqrt{L}}.$$

We notice here that we can bound the numerator by whatever $\epsilon > 0$ that we want because $s_n \to L$. When $L = 0$, however, the denominator gets arbitrarily small as well. Consider, though, that $|\sqrt{s_n} - L| = |\sqrt{s_n}|$ in this case. So we use the N for (s_n) that makes $|s_n - 0| < \epsilon^2$.

This motivates us to present our proof in two cases based on L. The case where $L > 0$ won't be an issue, though. Since $\sqrt{s_n} \geq 0$, we know that $\sqrt{L} < \sqrt{s_n} + \sqrt{L}$. Hence, we can bound

$$|\sqrt{s_n} - \sqrt{L}| \leq \frac{|s_n - L|}{\sqrt{s_n} + \sqrt{L}} \leq \frac{|s_n - L|}{\sqrt{L}}.$$

If this is to be less than ϵ, we would require an N such that $n \geq N$ implies

$$\frac{|s_n - L|}{\sqrt{L}} < \epsilon \quad \Leftrightarrow \quad |s_n - L| < \epsilon\sqrt{L}.$$

This is possible to find since $s_n \to L$. That is, given $\epsilon > 0$, we find N so large that $n \geq N$ implies $|s_n - L| < \frac{\epsilon}{\sqrt{L}}$. We can now write our formal proof.

Proof. Let $\epsilon > 0$.

Case $L > 0$: Take N so large that $|s_n - L| < \epsilon\sqrt{L}$ for all $n \geq N$. Then we see that when $n \geq N$,

$$|\sqrt{s_n} - \sqrt{L}| = \frac{|s_n - L|}{\sqrt{s_n} + \sqrt{L}} \leq \frac{|s_n - L|}{\sqrt{L}} < \frac{\epsilon\sqrt{L}}{\sqrt{L}} = \epsilon.$$

Hence, if $L > 0$, $\sqrt{s_n} \to \sqrt{L}$.

Case $L = 0$: Take N so large that $|s_n| < \epsilon^2$ for all $n \geq N$. Then we see that when $n \geq N$,

$$|\sqrt{s_n} - 0| = |\sqrt{s_n}| < \sqrt{\epsilon^2} = \epsilon.$$

Hence, $\sqrt{s_n} \to \sqrt{L}$ in this case as well. $\square$

In order to gain even more familiarity with this new way of proving propositions, I recommend that the student consult the textbook. I have left most of Ross's examples out of my notes so that the student has a plethora of examples to peruse and learn how these proofs are developed and written.

4.3 The Limit Theorems

We are now in a position where we can prove some elementary facts that were given (without proof) in the first-quarter Calculus class. These facts are the so-called *limit theorems* that allow us to evaluate the limits of many sequences (and, in turn, functions - we won't get to this version of the theorems in this class) by breaking them into smaller sequences and applying algebraic and ordering properties to evaluate the whole. In order to prove these facts, we need a fundamental result about convergent sequences:

Proposition 4.2. Convergent sequences are bounded.

Proof. Let (s_n) be a convergent sequence with $s_n \to L$. Since the ϵ-condition must hold for any $\epsilon > 0$, we observe that there must exist an $N \in \mathbb{N}$ such that

$$n \geq N \quad \Rightarrow \quad |s_n - L| < 1.$$

One version of the triangle inequality states that

$$||s_n| - |L|| \leq |s_n - L|.$$

Then we apply this to the above so that $|s_n| < |L| + 1$ for all $n \geq N$. Hence, each term after s_{N-1} is bounded by $|L| + 1$.

Define $M = \max\{|s_1|, |s_2|, \ldots, |s_{N-1}|, |L| + 1\}$. Then we must have $|s_n| \leq M$ for all $n \in \mathbb{N}$. Hence, (s_n) constitutes a bounded sequence. $\square$

We can now prove the limit theorems. That is, the $\lim\limits_{n\to\infty}$ operation on sequence is (1) linear (that is, homogeneous and additive) and (2) multiplicative.

Theorem 4.2. (Homogeneity) For a convergent sequence (s_n) and a constant $k \in \mathbb{R}$, the sequence (ks_n) converges. Moreover, $s_n \to L$ implies $ks_n \to kL$.

Proof. Let $\epsilon > 0$. If $k = 0$, then $(ks_n) = (0)$ and the result is trivial. Hence, we assume that $k \neq 0$.

Choose N so large that $|s_n - L| < \frac{\epsilon}{|k|}$ for all $n \geq N$. Then we verify that for $n \geq N$,

$$|ks_n - kL| = |k||s_n - L| < |k|\frac{\epsilon}{|k|} = \epsilon.$$

Hence, $ks_n \to kL$ as was to be shown. $\square$

Theorem 4.3. (Additivity) If the sequences (s_n) and (t_n) both converge, then so does $(s_n + t_n)$. Moreover, $s_n \to L_1$ and $t_n \to L_2$ implies $s_n + t_n \to L_1 + L_2$.

Proof. Let $\epsilon > 0$. Choose N_1 so large that $|s_n - L_1| < \frac{\epsilon}{2}$ for all $n \geq N_1$, and choose N_2 so large that $|t_n - L_2| < \frac{\epsilon}{2}$ for all $n \geq N_2$.

We define $N = \max N_1, N_2$ so that both s_n and t_n are within $\frac{\epsilon}{2}$ of their respective limit at the same time (when $n \geq N$). Then we verify that, when $n \geq N$,

$$|s_n + t_n - (L_1 + L_2)| = |(s_n - L_1) + (t_n - L_2)| \leq |s_n - L_1| + |t_n - L_2| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

so that $s_n + t_n \to L_1 + L_2$ as was to be shown. $\square$

Theorem 4.4. (Multiplicativity) If the sequences (s_n) and (t_n) both converge, then so does $(s_n t_n)$. Moreover, $s_n \to L_1$ and $t_n \to L_2$ implies $s_n t_n \to L_1 L_2$.

Proof. Let $\epsilon > 0$. First observe that both (s_n) and (t_n) are bounded as they are convergent. Say that (s_n) is bounded by $M > 0$.

Choose N_1 so large that $|t_n - L_2| < \frac{\epsilon}{2M}$ for all $n \geq N_1$, and (less intuitively) choose N_2 so large that $|s_n - L_1| < \frac{\epsilon}{2(|L_2|+1)}$ for all $n \geq N_2$. We then define $N = \max\{N_1, N_2\}$ so that $n \geq N$ guarantees these bounds simultaneously.

We now verify for $n \geq N$ that

$$|s_n t_n - L_1 L_2| = |s_n t_n - s_n L_2 + s_n L_2 - L_1 L_2|$$
$$\leq |s_n t_n - s_n L_2| + |s_n L_2 - L_1 L_2| = |s_n||t_n - L_2| + |L_2||s_n - L_1|$$
$$\leq M \cdot \frac{\epsilon}{2M} + |L_2| \cdot \frac{\epsilon}{2(|L_2| + 1)} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence, $s_n t_n \to L_1 L_2$ as was to be shown. $\square$

Now that we've established that the limit of a product is the product of the limits, we can deal with quotients $\frac{a}{b}$ by treating them as a product $a \cdot \frac{1}{b}$. Hence, we need to establish that the limit of a reciprocal is the reciprocal of the limit. This is a delicate matter, though, and we need to prove one fact about sequences with nonzero limits.

Proposition 4.3. A convergent sequence (s_n) satisfying $s_n \neq 0$ for all $n \in \mathbb{N}$ and $s_n \to L \neq 0$ forces $\inf\limits_{n\in\mathbb{N}}\{|s_n|\} > 0$.

Proof. [This proof looks very similar to the proof that shows convergent sequences are bounded.] Choose N so large that for all $n \geq N$ we have

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

When $n \geq N$, the triangle inequality implies $|s_n| \geq \frac{|L|}{2}$.

Define $m = \min\left\{|s_1|, |s_2|, \ldots, |s_{N-1}, \frac{|L|}{2}\right\}$. Then we must have $|s_n| \geq m$ for all $n \in \mathbb{N}$. Hence, $\inf\limits_{n\in\mathbb{N}}\{|s_n|\} > 0$. $\square$

Theorem 4.5. (Reciprocal Limit) If (s_n) converges to a limit $L \neq 0$ and satisfies $s_n \neq 0$ for all $n \in \mathbb{R}$, then $\left(\frac{1}{s_n}\right)$ converges to $\frac{1}{L}$.

Proof. Let $\epsilon > 0$. The proposition gives us a lower bound $m > 0$ to the set $\{|s_n| \mid n \in \mathbb{N}\}$.

Choose N so large that $|s_n - L| < \epsilon m|L|$ for all $n \geq N$. Then $n \geq N$ implies

$$\left|\frac{1}{s_n} - \frac{1}{L}\right| = \left|\frac{L - s_n}{s_n L}\right| \leq \frac{|s_n - L|}{m|L|} < \frac{\epsilon \cdot m|L|}{m|L|} = \epsilon$$

so that $\frac{1}{s_n} \to \frac{1}{L}$ as was to be shown. $\square$

Corollary 4.1. (Quotient Limit) Suppose (s_n) and (t_n) both converge and that $s_n \neq 0$ for all $n \in \mathbb{N}$ with $s_n \to L_1 \neq 0$ and $t_n \to L_2$. Then the sequence $\left(\frac{t_n}{s_n}\right)$ also converges, and $\frac{t_n}{s_n} \to \frac{L_2}{L_1}$.

Proof. We write $\frac{t_n}{s_n} = t_n \cdot \frac{1}{s_n}$ for all $n \in \mathbb{N}$. We can then apply the product theorem with the reciprocal theorem so that

$$\lim \frac{t_n}{s_n} = \lim t_n \cdot \frac{1}{s_n} = (\lim t_n)\left(\lim \frac{1}{s_n}\right) = L_2 \cdot \frac{1}{L_1} = \frac{L_2}{L_1},$$

as was to be shown. $\square$

This slough of theorems is enough to prove many standard limit calculations.

Proposition 4.4.

1. $\lim \frac{1}{n^p} = 0$ when $p > 0$.
2. $\lim a^n = 0$ for $|a| < 1$.
3. $\lim n^{1/n} = 1$.
4. $\lim a^{1/n} = 1$ for $a > 0$.

Proof.

1. Let $\epsilon > 0$. Choose $N > \left(\frac{1}{\epsilon}\right)^{1/p}$. Then for $n \geq N$,

$$\left|\frac{1}{n^p}\right| \leq \left|\frac{1}{N^p}\right| < \frac{1}{(1/\epsilon^{1/p})^p} = \epsilon$$

so that $\frac{1}{n^p} \to 0$ when $p > 0$ as was to be shown.

2. Let $\epsilon > 0$. If $a = 0$, then $(a^n) = (0)$ so that the theorem is trivial. Hence, we assume $a \neq 0$. In fact, we can find $b > 0$ so that $|a| = \frac{1}{b+1}$ (why?).

   Choose $N = \frac{1}{\epsilon b}$. Then the binomial theorem states $(1+b)^n \geq 1 + nb > nb$ for all $n \in \mathbb{N}$. This allows us to verify that for $n \geq N$

   $$|a^n - 0| = |a^n| = \frac{1}{(1+b)^n} < \frac{1}{nb} \leq \frac{1}{Nb} = \frac{1}{[1/(\epsilon b)]b} = \epsilon$$

   so that $a^n \to 0$ as was to be shown.

3. Let $\epsilon > 0$. We show the equivalent statement that $s_n = n^{1/n} - 1 \to 0$. Observe that each term in s_n is nonnegative. A quick bit of algebra shows $n = (1+s_n)^n$, and the binomial theorem yields

   $$n = (1+s_n)^n \geq 1 + ns_n + \frac{1}{2}n(n-1)s_n^2 > \frac{1}{2}n(n-1)s_n^2.$$

   Provided $n \geq 2$, we conclude further the nontrivial inequality $n > \frac{1}{2}n(n-1)s_n^2$. When $n \geq 2$, this gives

   $$s_n^2 < \frac{2}{n-1} \quad \Leftrightarrow \quad s_n < \sqrt{\frac{2}{n-1}}.$$

   By the Archimedean property, we may choose $N > \max\{\frac{3}{\epsilon}, 2, 3\}$. Because $\sqrt{x} \leq x$ when $x \geq 1$, we see that

   $$|s_n - 0| = s_n < \sqrt{\frac{2}{n-1}} \leq \frac{2}{n-1} \leq \frac{3}{n} \leq \frac{3}{N} \leq \frac{3}{(3/\epsilon)} = \epsilon$$

   so that $s_n \to 0$. Hence, $n^{1/n} \to 1$ as was to be shown.

4. Let $\epsilon > 0$.

   Case $a \geq 1$: By now you will have shown the "Squeeze Theorem" in homework. It is applied in this case as $1 \leq a^{1/n} \leq n^{1/n}$ when $n \geq a$. Since (c) showed that $n^{1/n} \to 1$, the Squeeze Theorem implies that $a^{1/n} \to 1$ as well.

   Case $a < 1$: Here $1 < \frac{1}{a}$. From (b) we conclude that $a^{1/n} \to 1$.

□

Now that we're done proving all of the necessary theorems, we can show how they are applied in practice.

Example 4.6. Recall from the previous section that we (quite laboriously) showed

$$\frac{2n^4 - 5n^2}{n^4 - 3} \to 2.$$

We can much more easily show this limit by applying the limit theorems. Unfortunately, neither of the sequences in the numerator or denominator are convergent. However, we can algebraically manipulate this expression so that these sequences do converge:

$$\frac{2n^4 - 5n^2}{n^4 - 3} = \frac{2n^4 - 5n^2}{n^4 - 3} \cdot \frac{1/n^4}{1/n^4} = \frac{2 - \frac{5}{n^2}}{1 - \frac{3}{n^4}}.$$

Therefore, the quotient theorem allows us to conclude that

$$\frac{2 - \frac{5}{n^2}}{1 - \frac{3}{n^4}} \to \frac{2 - 5 \cdot 0}{1 - 3 \cdot 0} = 2.$$

Example 4.7. We can generalize the above result as follows. Consider the sequence

$$\left(\frac{a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1} n^{q-1} + \cdots + b_1 n + b_0}\right)$$

where $p, q \geq 1$, $a_p \neq 0$, $b_q \neq 0$, and no value of n makes the denominator equal to zero.

Case $p < q$: Then the algebraic expression can be manipulated so that

$$\frac{a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1} n^{q-1} + \cdots + b_1 n + b_0} \cdot \frac{1/n^q}{1/n^q}$$

$$= \frac{a_p n^{p-q} + a_{p-1} n^{p-q-1} + \cdots + a_1 n^{1-q} + a_0 n^{-q}}{b_q + b_{q-1} n^{-1} + \cdots + b_1 n^{1-q} + b_0 n^{-q}} \to \frac{0}{b_q} = 0.$$

Case $p = q$: Then the algebraic expression can be manipulated so that

$$\frac{a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1} n^{q-1} + \cdots + b_1 n + b_0} \cdot \frac{1/n^q}{1/n^q}$$

$$= \frac{a_p + a_{p-1} n^{-1} + \cdots + a_1 n^{1-q} + a_0 n^{-q}}{b_q + b_{q-1} n^{-1} + \cdots + b_1 n^{1-q} + b_0 n^{-q}} \to \frac{a_p}{b_q}.$$

Case $p > q$: Then the algebraic expression can be manipulated so that

$$\frac{a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1} n^{q-1} + \cdots + b_1 n + b_0} \cdot \frac{1/n^q}{1/n^q}$$

$$= \frac{a_p n^{p-q} + a_{p-1} n^{p-q-1} + \cdots + a_1 n^{1-q} + a_0 n^{-q}}{b_q + b_{q-1} n^{-1} + \cdots + b_1 n^{1-q} + b_0 n^{-q}}.$$

Unfortunately, this final expression doesn't converge. The proof of this fact stems from the facts that every polynomial with positive leading coefficient is eventually surpassed (that is, bounded above) by a polynomial with a single term.

The last case of the above example exemplifies that we may need to extend our definition of a limit to the extended reals.

Definition 4.3. Given a sequence (s_n), we say

1. (s_n) *diverges to* $+\infty$: $\lim s_n = +\infty$ if for each $M > 0$ there is an N such that $n \geq N$ implies $s_n > M$. We denote this by $s_n \to +\infty$.

2. (s_n) *diverges to* $-\infty$: $\lim s_n = -\infty$ if for each $M < 0$ there is an N such that $n \geq N$ implies $s_n < M$. We denote this by $s_n \to -\infty$.

Example 4.8. In light of this definition, the final case of the previous example shows that

$$\frac{a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1} n^{q-1} + \cdots + b_1 n + b_0} \to \begin{cases} +\infty, & \text{if } \frac{a_p}{b_q} > 0 \\ -\infty, & \text{if } \frac{a_p}{b_q} < 0. \end{cases}$$

when $p > q$. We should emphasize here that **the limit theorems don't apply to** $\widehat{\mathbb{R}}$.

Example 4.9. To see how to apply these definitions to specific sequences, we prove the following claim formally.

Claim: $\lim \frac{2n^3-5}{7n-2} = +\infty$.

Proof. Let $M > 0$. Choose $N > \max\{2M, 2\}$. Then observe that if $n \geq N$, then we have

$$\frac{2n^3 - 5}{7n - 2} \geq \frac{2n^3}{8n} = \frac{1}{4}n^2 \geq \frac{1}{4}N^2 > \frac{1}{4}(2M)^2 = M.$$

Hence, the sequence diverges to $+\infty$. $\square$

Even though we can't apply the limit theorems directly, we can still come up with results that apply to sequences that diverge in a very particular way.

Theorem 4.6. Let (s_n) and (t_n) be sequences such that $s_n \to +\infty$ and $t_n \to L > 0$ (L can be $+\infty$). Then $s_n t_n \to +\infty$ as well.

Proof. Let $M > 0$. Since $t_n \to L > 0$, we can find an m such that $0 < m < L$. Choose N_1 so large that $t_n > m$ for all $n \geq N_1$.

Since $s_n \to +\infty$, there exists some N_2 such that $s_n > \frac{M}{m}$ for all $n \geq N_2$.

Define $N = \max N_1, N_2$. Then we must have for all $n \geq N$ that

$$s_n t_n > s_n \cdot m > \frac{M}{m} \cdot m = M.$$

Hence, $s_n t_n \to +\infty$ as was to be shown. $\square$

Theorem 4.7. For a sequence (s_n) with $s_n > 0$ for all $n \in \mathbb{N}$, $s_n \to +\infty$ if and only if $\frac{1}{s_n} \to 0$.

Proof. "$\Rightarrow$": Suppose $s_n \to +\infty$. Let $\epsilon > 0$. Take N so large that $s_n > \frac{1}{\epsilon}$ for all $n \geq N$. Then we see for $n \geq N$ that

$$\left| \frac{1}{s_n} - 0 \right| = \frac{1}{|s_n|} = \frac{1}{s_n} \leq \frac{1}{1/\epsilon} = \epsilon.$$

Hence, $\frac{1}{s_n} \to 0$ as was to be shown.

"$\Leftarrow$": Now suppose that $\frac{1}{s_n} \to 0$. Let $M > 0$. Choose N so large that $\left| \frac{1}{s_n} - 0 \right| = \frac{1}{s_n} < \frac{1}{M}$ for all $n \geq N$. Then $s_n > M$ for all $n \geq N$ so that $s_n \to +\infty$ as was to be shown. $\square$

4.4 No Limit? No Problem!

We have successfully defined what it means for a sequence of numbers to "eventually get close to so value"; that is, converge to a limit. However, there is a bit of a shortcoming to our definition. Consider the sequence

$$(s_n) = \left(\sum_{k=1}^{n} \frac{1}{k^2} \right).$$

The student should recognize the limit of this sequence as a series

$$s_n \to \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

From an unproven(!) theorem in Calculus, this series (a so-called *p-series* with $p > 1$) converges to some positive real number.

Observe that the theorem tells us that **a limit exists**, but it doesn't tell us what that limit is! The question becomes this:

> How do we establish that a sequence converges when **we don't know its limit**?

4.4.1 Monotone and Recursive Sequences

This is hard to do for a general sequence; however, there are certain types of sequences whose convergence is relatively easily established.

Definition 4.4. A sequence (s_n) is called *increasing/non-decreasing* if $s_{n+1} \geq s_n$ for all $n \in \mathbb{N}$. It is called *strictly increasing* if $s_{n+1} > s_n$ for all $n \in \mathbb{N}$.

A sequence (s_n) is called *decreasing/non-increasing* if $s_{n+1} \leq s_n$ for all $n \in \mathbb{N}$. It is called *strictly decreasing* if $s_{n+1} < s_n$ for all $n \in \mathbb{N}$.

A sequence that is either increasing or decreasing will be called a *monotone/monotonic* sequence. A sequence is called *strictly monotone* if it is either strictly increasing or strictly decreasing.

Example 4.10. The sequences $\left((1 - \frac{1}{n})^n\right)$, $\left(\sum_{k=1}^{n} \frac{1}{k^2}\right)$, and (n) are all increasing sequences (though not all so obviously as the last).

The sequences $\left(\frac{1}{n}\right)$, $\left(1 + \frac{1}{n^2}\right)$, and $((1+n)^{1/n})$ are all decreasing sequence (though the last one not quite so obviously as the first two).

None of the sequences $((-1)^n)$, $(\sin(n))$, or $((-1)^n n)$ are monotone.

We prove the following well-known theorem about monotone sequences.

Theorem 4.8. (Monotone Convergence) All bounded monotone sequences converge. More specifically:

1. All increasing sequences that are bounded above converge, and

2. All decreasing sequences that are bounded below converge.

Proof. Let $\epsilon > 0$, and denote the monotone sequence by (s_n).

(1): Let $S = \{s_n \mid n \in \mathbb{N}\}$ be the set of all terms of the sequence. Because this is nonempty and bounded above, $u = \sup S \in \mathbb{R}$. We show that $s_n \to u$ in this case.

Because $u - \epsilon$ is not an upper bound for S, there must exist some $N \in \mathbb{N}$ such that $s_N > u - \epsilon$. The fact that (s_n) is increasing shows that $s_n > u - \epsilon$ for all $n \geq N$. Since $u \geq s_n$ for all $n \in \mathbb{N}$, and for $n \geq N$ in particular, we have

$$u - \epsilon < s_n \leq u \quad \Rightarrow \quad |s_n - u| < \epsilon$$

for all $n \geq N$. Hence, $s_n \to u$ so that (s_n) converges as required.

(2): The proof of this case follows immediately by instead considering the set $-S = \{-s_n \mid n \in \mathbb{N}\}$ and showing that $s_n \to \inf S = \sup -S$. $\square$

Notice that we appealed the completeness of $\mathbb{R}$ in order to prove this fact. Indeed this theorem won't hold for incomplete ordered fields - just think of the sequence $\{1, 1.4, 1.41, 1.414, 1.4142, \ldots\} \subset \mathbb{Q}$ that approximates $\sqrt{2}$.

We say that a sequence (s_n) is *recursively defined* (or just *recursive*) if $s_{n+1} = f(s_n)$ for all $n \in \mathbb{N}$. The equation $s_{n+1} = f(s_n)$ is called a *recursion*. Such recursive sequences provide myriad examples where the Monotone Convergence Theorem may be applied.

Example 4.11. Consider the sequence defined by

$$s_1 = 4 \quad \text{and} \quad s_{n+1} = \frac{1}{3}(s_n + 2).$$

It is easy to see that because we begin the sequence with a positive number, all subsequent terms must also be positive.

Claim: (s_n) is bounded below.

Proof. We merely observe that $s_1 > 0$ and argue that no negative terms are ever used in the recursion. Hence, we conclude that $s_n \geq 0$ for $n \in \mathbb{N}$ so that the sequence is bounded below by zero. $\square$

Claim: (s_n) is decreasing.

Proof. (By induction) Notice that $s_1 = 4$ and $s_2 = 2$ so that $s_2 \leq s_1$. Now suppose that $s_{n+1} \leq s_n$ for some $n \in \mathbb{N}$. Then consider

$$s_{n+2} = \frac{1}{3}(s_{n+1} + 2) \leq \frac{1}{3}(s_n + 2) = s_{n+1}$$

so that $s_{n+2} \leq s_{n+1}$ as well so that the claim is proven. $\square$

By the Monotone Convergence Theorem, we see that (s_n) converges. Moreover, if we say that $s_n \to L$, we can compute

$$\lim s_{n+1} = \lim \frac{1}{3}(s_n + 2) \quad \Leftrightarrow \quad L = \frac{L+2}{3}$$

so that $L = 1$ is the limit of the sequence.

Example 4.12. (Ancient square root calculator) Consider the sequence defined as

$$s_1 = a > 0 \quad \text{and} \quad s_{n+1} = \frac{1}{2}\left(s_n + \frac{a}{s_n}\right).$$

It is clear that this sequence is nonnegative since it begins with a positive term. We want to show that this sequence converges. We do this by showing that $s_n^2 \geq a$ (so that (s_n) is bounded below when s_n is positive) and that (s_n) is a decreasing sequence.

Claim: $s_n^2 \geq a$ for all $n \in \mathbb{N}$.

Proof. Notice that the recursion can be arranged so that

$$s_n^2 - 2s_{n+1}s_n + a = 0.$$

Because each s_n satisfies this equation, it has real roots. This implies that the discriminant

$$4s_{n+1}^2 - 4a \geq 0.$$

This proves our claim. $\square$

Claim: (s_n) is decreasing.

Proof. Consider now that

$$s_n - s_{n+1} = s_n - \frac{1}{2}\left(s_n + \frac{a}{s_n}\right) = \frac{s_n^2 - a}{s_n} \geq 0,$$

which proves our claim. $\square$

Then the Monotone Convergence Theorem shows us that the sequence (s_n) converges to a limit $L \in \mathbb{R}$. Moreover, now that we have that limit, we can find it explicitly by letting $n \to \infty$ in the recursion:

$$\lim s_{n+1} = \lim \frac{1}{2}\left(s_n + \frac{a}{s_n}\right) \quad \Leftrightarrow \quad L = \frac{1}{2}\left(L + \frac{a}{L}\right).$$

Solving this for L yields $L = \pm\sqrt{a}$. We take the positive root since our nonnegative sequence can't have a negative limit. That is to say that $s_n \to +\sqrt{a}$.

This is a method for computing the square root of a positive number that date back to Mesopotamia before 1500 BCE(!!). As an illustrating example, we can compute $\sqrt{2}$ by checking that

$$s_1 = 2, \quad s_2 = \frac{3}{2} = 1.5, \quad s_3 = \frac{17}{12} = 1.41\overline{6}, \quad s_4 = \frac{577}{408} \cong 1.4142156862.$$

For comparison's sake, $\sqrt{2} = 1.4142135623\ldots$. This method finds $\sqrt{2}$ to **within five decimals in three iterations**!

We can also use the extended reals $\widehat{\mathbb{R}}$ to further our understanding of these sequences.

Proposition 4.5. Unbounded, monotone sequences diverge to either $+\infty$ or $-\infty$. In particular,

1. if (s_n) is unbounded above and increasing, then $\lim s_n = +\infty$, and
2. if (s_n) is unbounded below and decreasing, then $\lim s_n = -\infty$.

Proof. **(1)**: Let $M > 0$. Suppose indeed that (s_n) is both unbounded above and increasing. Then there exists an $N \in \mathbb{N}$ such that $s_N > M$. Because (s_n) is increasing, we must have $s_n \geq s_N > M$ for all $n \geq N$. Hence, $s_n \to +\infty$ as was to be shown.

(2): Let $M < 0$. Suppose indeed that (s_n) is both unbounded below and decreasing. Then there exists an $N \in \mathbb{N}$ such that $s_N < M$. Because (s_n) is decreasing, we must have $s_n \leq s_N < M$ for all $n \geq N$. Hence, $s_n \to -\infty$ as was to be shown. $\square$

This theorem helps us complete the thought that monotone sequences (s_n) are "well-behaved" as $n \to \infty$. That is to say that, if we use $\widehat{\mathbb{R}}$, **limits of monotone sequences always exist**!

4.4.2 The Limit Superior and Limit Inferior

Perhaps showing monotonicity or boundedness is too hard to show for some sequence (s_n). Because of our definition of the limit of a sequence, the "interesting" behavior occurs only for the "tail" of the sequence. That is, for the set of terms

$$T_N = \{s_n \mid n \geq N\}.$$

Notice that (T_N) is now a sequence not of real numbers but rather of sets. This sequence of *tails* of the sequence (s_n) can be studied with tools that we already have if we extract a real number from each set to recover a sequence of real numbers.

If we can't get a handle on the limit of the sequence, then perhaps we can understand the limiting behavior of the sequence better by studying these tails. To accomplish this, we analyze the sequences $(\sup T_N)$ and $(\inf T_N)$. [Notice that $(\sup T_N)$ is decreasing while $(\inf T_N)$ is increasing. (Why?)] By using $\widehat{\mathbb{R}}$, we can always speak of the limits of these sequences. These limits are of particular import to us.

Definition 4.5. Let (s_n) be a sequence of real numbers. For

$$T_N = \{s_n \mid n \geq N\}, \quad N \in \mathbb{N},$$

we define the *limit superior* as

$$\limsup s_n = \lim_{N\to\infty} \sup T_N \in \widehat{\mathbb{R}}$$

and the *limit inferior* as

$$\liminf s_n = \lim_{N\to\infty} \inf T_N \in \widehat{\mathbb{R}}$$

.

Observe that it is always the case that $\liminf s_n \leq \limsup s_n$. By defining these quantities as being able to take on infinite values, we've guaranteed that **the limit superior and limit inferior always exist**!

Example 4.13. 1. Let $(s_n) = ((-1)^n)$. Then $\limsup s_n = \lim(\sup\{-1, 1\}) = 1$ and $\liminf s_n = \lim(\inf\{-1, 1\}) = -1$.

2. Let $(s_n) = (n^{(-1)^n})$. Then $\limsup s_n = +\infty$ and $\liminf s_n = 0$.

3. Let $(s_n) = \left(\frac{n+1}{n}\right)$. Then $\limsup s_n = \liminf s_n = 1$.

Intuitively speaking, $\limsup s_n$ is the largest value that **infinitely many terms** of s_n can get close to. Similarly, $\liminf s_n$ is the smallest such value. This goes to show that $\limsup s_n \leq \sup\{s_n \mid n \in \mathbb{N}\}$ and $\liminf s_n \geq \inf\{s_n \mid n \in \mathbb{N}\}$ (why?). We will study both of these in greater detail shortly within the context of subsequences.

With these observations in mind, if these two values are the same, then we can say that the sequence is zeroing in on a certain value in the limit; that is, the sequence converges.

Theorem 4.9. Let (s_n) be a sequence.

1. If $s_n \to L \in \widehat{\mathbb{R}}$, then $\liminf s_n = L = \limsup s_n$.

2. If $\liminf s_n = \limsup s_n = L \in \widehat{\mathbb{R}}$, then $\lim s_n = L$.

Proof. Let's denote $u_N = \inf\{s_n \mid n \geq N\}$ and $v_N = \sup\{s_n \mid n \geq N\}$. Hence, take $u = \lim u_N = \liminf s_n$ and $v = \lim v_N = \limsup s_n$.

(1): As we are using $\widehat{\mathbb{R}}$, we must proceed in cases.

- **Case** $L = +\infty$: Let $M > 0$. Then there exists some $N \in \mathbb{N}$ such that $s_n > M$ for all $n \geq N$. Then it is clear that $u_n \geq M$ for all $n \geq N$ as well (that is, the infima can't dip below M). This shows that $u_n \to u = +\infty$.

  A similar argument shows that $v_n \to v = +\infty$ as well.

- **Case** $L = -\infty$: Let $M < 0$. Then there exists some $N \in \mathbb{N}$ such that $s_n < M$ for all $n \geq N$. Then it is clear that $v_n \leq M$ for all $n \geq N$ as well (that is, the suprema don't sneak above M). This shows that $v_n \to v = -\infty$.

  A similar argument shows that $u_n \to u = -\infty$ as well.

- **Case** $L \in \mathbb{R}$: Let $\epsilon > 0$. Then there exists some $N \in \mathbb{N}$ such that $|s_n - L| < \epsilon$ for all $n \geq N$. Then $s_n < L + \epsilon$ for all $n \geq N$ so that $L + \epsilon$ is an upper bound for all s_n with $n \geq N$. Hence, $v_n \leq L + \epsilon$ for all $n \geq N$ as well. By our remarks before the theorem, this yields $v \leq L + \epsilon$. In a similar fashion, we can show that $L - \epsilon < u$. Because $u \leq v$, we have

  $$L - \epsilon < u \leq v < L + \epsilon \quad \Leftrightarrow \quad 0 \leq |u - L| < \epsilon \quad \text{and} \quad 0 \leq |v - L| < \epsilon$$

  for arbitrary $\epsilon > 0$. Therefore, $u = v = L$ as was to be shown.

(2): Again, we proceed by cases.

- **Case** $u = v = +\infty$: Let $M > 0$. Then there exists an $N \in \mathbb{N}$ such that $u_n > M$ for all $n \geq N$. Therefore, $s_n \geq u_n > M$ for all $n \geq N$ as well. Hence, $s_n \to +\infty$ in this case.

- **Case** $u = v = -\infty$: Let $M < 0$. Then there exists an $N \in \mathbb{N}$ such that $v_n < M$ for all $n \geq N$. Therefore, $s_n \leq v_n < M$ for all $n \geq N$ as well. Hence, $s_n \to -\infty$ in this case.

- **Case** $u = v \in \mathbb{R}$: Let $\epsilon > 0$. Then there exists an $N_1 \in \mathbb{N}$ such that $|v - v_{N_1}| < \epsilon$. Then $v_n < v + \epsilon$ for all $n \geq N_1$. In a similar fashion, there exists an $N_2 \in \mathbb{N}$ such that $u - \epsilon < u_n$ for all $n \geq N_2$. To satisfy these both simultaneously, we choose $N = \max\{N_1, N_2\}$. Setting $L = u = v$, we have

  $$L - \epsilon < s_n < L + \epsilon \quad \Leftrightarrow \quad |s_n - L| < \epsilon$$

  for all $n \geq N$ so that $s_n \to L$.

□

4.4.3 Cauchy Sequences

In light of the previous section, we can glean a lot of information from the limiting behavior of a sequence merely by comparing elements of the sequence to one another. That is, **eventually the terms of the sequence will get arbitrarily close to one another**. This leads to perhaps the most important condition that a sequence of real numbers can satisfy.

Definition 4.6. A sequence (s_n) is called *Cauchy* if

for every $\epsilon > 0$ there exists an N such that $|s_n - s_m| < \epsilon$ for all $n, m \geq N$.

This version of the ϵ-condition is called the *Cauchy criterion*.

Remark 4.2. Notice that this definition doesn't allow us to pick and choose which pairs m, n we want to use. To show a sequence is Cauchy, we need to assume m, n are arbitrary indices beyond the N in the definition.

The idea of "Cauchyness" is that a sequence will converge if its **terms eventually get arbitrarily close to one another**. This is not the same as saying that consecutive terms eventually get arbitarily close. For example, consider the sequence $(\ln n)$. Then the difference

$$|\ln(n+1) - \ln n| = \ln\left(\frac{n+1}{n}\right) = \ln\left(1 + \frac{1}{n}\right)$$

can be shown to converge to zero, but the sequence itself diverges to $+\infty$.

Intuition serves us well here. The ϵ-condition and the Cauchy criterion are indeed equivalent. However, the proof of this fact must invoke the completeness of $\mathbb{R}$; though, the first implication is rather straightforward.

Proposition 4.6. Every convergent sequence is a Cauchy sequence.

Proof. Let $\epsilon > 0$. Denote the sequence in question by (s_n) and let $s_n \to L$. Because $s_n \to L$, there exists an $N \in \mathbb{N}$ such that $|s_n - L| < \frac{\epsilon}{2}$ for all $n \geq N$.

Observe here that

$$|s_n - s_m| = |(s_n - L) + (L - s_m)| \leq |s_n - L| + |s_m - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $n, m \geq N$. Hence, (s_n) is a Cauchy sequence. □

To establish the other implication that Cauchy sequences converge, are going to have to work a little harder.

Proposition 4.7. Cauchy sequences are bounded.

Proof. Let (s_n) be a Cauchy sequence. Then for $\epsilon = 1$, there exists an $N \in \mathbb{N}$ such that $|s_n - s_m| < 1$ for $n, m \geq N$. When $m = N + 1$, we see that $|s_n - s_N| < 1$ for $n \geq N$ so that the triangle inequality $|s_n| < |s_N| + 1$ for all $n \geq N$. As we showed in the convergent sequence case, we simply bound (s_n) by the greatest value among the first $N - 1$ terms of the sequence and the bound that we found when $\epsilon = 1$. That is to say that

$$|s_n| \leq M = \max\{|s_1|, |s_2|, \ldots, |s_{N-1}|, |s_N| + 1\}.$$

Hence, (s_n) is a bounded sequence. □

We are now in a position to prove the following (major!) theorem.

Theorem 4.10. (Completeness) Cauchy sequences are convergent sequences.

Proof. Let (s_n) be a Cauchy sequence. The previous proposition shows that (s_n) is bounded so that $\liminf s_n = u$ and $\limsup s_n = v$ are both real numbers. By a previous theorem, we merely need to show that $u = v$. We use the same notation as before that $u_N = \inf\{s_n \mid n \geq N\}$ and $v_N = \sup\{s_n \mid n \geq N\}$.

Let $\epsilon > 0$. Then there exists an $N \in \mathbb{N}$ such that $|s_n - s_m| < \epsilon$ for all $n, m \geq N$. In particular, we have $|s_n| < |s_m| + \epsilon$ for all $n \geq N$. Then $v_N \leq |s_m| + \epsilon$ for $m \geq N$. Hence, $v_N - \epsilon \leq u_N$. Therefore,

$$\limsup s_n \leq v_N \leq u_N + \epsilon \leq \liminf s_n + \epsilon$$

for all $\epsilon > 0$. Hence, $v \leq u$. The opposite inequality may be shown in a similar fashion so that $u = v$ as desired. □

Hidden in here are suprema and infima, so we really do need the completeness of $\mathbb{R}$ to make this work. We can now conclude the following corollary.

Corollary 4.2. (Cauchy $\Leftrightarrow$ convergent) The real sequence (s_n) is Cauchy if and only if it is convergent.

4.5 Subsequences

This topic most certainly deserves its own section. There are usually two different approaches one can use to deal with sequences: The lim sup approach and the *subsequence* approach. Both take a fair bit of winding up in order to use effectively, and both have their merits. We have already seen the utility of the lim sup approach to sequences when we proved that Cauchy sequences converge. We would now like to show that the same sorts of goals can be accomplished with the more general subsequence approach. [Technically speaking, these two approaches are identical; however, the generality of the subsequence approach makes it look so much different from the lim sup approach in practice that it can effictively been seen as a different approach.]

Definition 4.7. Let (s_n) be a sequence. A *subsequence* of (s_n) is a restriction of (s_n) to a countably infinite subset of $\mathbb{N}$.

Alternatively, $(s_{n_k})_{k\in\mathbb{N}}$ is a subsequence of (s_n) if $\{n_k \mid k \in \mathbb{N}\}$ satisfies

$$1 \leq n_1 < n_2 < \cdots < n_k < n_{k+1} < \cdots$$

Example 4.14.
1. The constant sequences (-1) and (1) are (convergent!) subsequences of the alternating sequence $((-1)^n)$.
2. The sequence $(2n) = \{2, 4, 6, \ldots\}$ is a subsequence of the sequence $((-1)^n n) = \{-1, 2, -3, 4, -5, 6 \ldots\}$.
3. The *N-tail sequence* $(s_n)_{n\geq N}$ is a subsequence of any sequence (s_n). These are the sequences that we consider for the definitions of lim sup and lim inf. Yes, the 1-tail subsequence of (s_n) is just (s_n) itself. Hence, a sequence is always a subsequence of itself.

The next theorem will be proven in two ways in order to illustrate some common techniques in Analysis. The first is more cavalier in the sense that we don't care about any specifics about the end object. The second takes a fair bit more finesse in order to guarantee a better-looking end object. Both methods are entirely viable techniques.

Proposition 4.8. Let (s_n) be a sequence.

1. If $t \in \mathbb{R}$, then there is a subsequence of (s_n) converging to L if and only if the set $\{n \in \mathbb{N} \mid |s_n - L| < \epsilon\}$ is infinite for all $\epsilon > 0$.
2. If the sequence (s_n) is unbounded above, then it has a subsequence diverging to $+\infty$.
3. If the sequence (s_n) is unbounded below, then it has a subsequence diverging to $-\infty$.

Proof.
1. "$\Rightarrow$": If there is a subsequence (s_{n_k}) such that $s_{n_k} \to L$, then the definition of convergence gives us infinitely many n's for any $\epsilon > 0$ we choose. Therefore, $S_\epsilon = \{n \in \mathbb{N} \mid |s_n - L| < \epsilon\}$ is infinite for all $\epsilon > 0$.

   "$\Leftarrow$": Suppose now that for each $\epsilon > 0$ the set S_ϵ is infinite. Observe that $S_\epsilon \subseteq S_\delta$ whenever $\epsilon < \delta$ (since $|s_n - L| < \epsilon < \delta$ for $n \in S_\epsilon$). We shall construct the subsequence (s_{n_k}) mentioned in the statement.

Choose the first index of the subsequence n_1 from S_1. Then choose the second index n_2 from $S_{1/2}$ so that $n_2 > n_1$. To continue the pattern, choose the k^{th} index n_k from $S_{1/k}$ such that $n_k > n_{k-1}$. This is possible at each step because S_ϵ has infinitely many indices to choose from. The claim now is that this subsequence (s_{n_k}) converges to L.

Let $\epsilon > 0$. Choose $K > \frac{1}{\epsilon}$. Then for all $k \geq K$ we have

$$|s_{n_k} - L| < \frac{1}{k} \leq \frac{1}{K} \leq \frac{1}{1/\epsilon} = \epsilon$$

so that $s_{n_k} \to L$ as we claimed.

2. We construct the subsequence (s_{n_k}) in question. Define $A_M = \{n \in \mathbb{N} \mid s_n > M\}$. Since (s_n) is unbounded above, A_M is infinite for each $M > 0$. Choose the first index n_1 from A_1. Choose the second index n_2 from A_2 such that $n_2 > n_2$. Following the pattern, choose the k^{th} index n_k from S_k such that $n_k > n_{k-1}$. We claim that the subsequence (s_{n_k}) diverges to $+\infty$.

   Let $M > 0$. Choose $K > M$. Then for all $k \geq K$ we have $s_{n_k} > k \geq K > M$. Hence, $(s_{n_k}) \to +\infty$ as desired.

3. This case is proven in a fashion similar to (2).

□

Corollary 4.3. In each case above, the subsequence may be chosen so that it is monotone.

Proof. We construct the subsequence more cautiously here.

1. We show the fact that if S_ϵ is infinite (using the same notation as above), then there is a **monotone** subsequence (s_{n_k}) of (s_n) that converges to L.

   For starters, suppose $\mathcal{L} = \{n \in \mathbb{N} \mid s_n = L\}$ is infinite (notice that $\mathcal{L} \subseteq S_\epsilon$ for all $\epsilon > 0$). Then of course we can construct a subsequence whose limit is L. Just take an increasing sequence (n_k) from $\mathcal{L}$ such that $(s_{n_k}) = (L)$, which is a monotone (increasing **and** decreasing) sequence that converges to L.

   When $\mathcal{L}$ is finite, we consider S_ϵ as the union of

$$S_\epsilon^- = \{n \in \mathbb{N} \mid L - \epsilon < s_n < L\},$$
$$S_\epsilon^+ = \{n \in \mathbb{N} \mid L < s_n < L + \epsilon\},$$

   and $\mathcal{L}$. That is, $S = S_\epsilon^- \cup \mathcal{L} \cup S_\epsilon^+$. Because S_ϵ is infinite and $\mathcal{L}$ is finite, either S_ϵ^- is infinite for all $\epsilon > 0$ or S_ϵ^+ is infinite for all $\epsilon > 0$ (otherwise, for small enough ϵ, S_ϵ would be finite).

   Let's assume that S_ϵ^- is infinite. We construct (s_{n_k}) *inductively* as follows.

   (a) Select n_1 so that $L - 1 < s_{n_1} < L$.

   (b) Suppose that we have already selected $n_1 < n_2 < \cdots < n_{k-1}$ such that

$$\max\left\{s_{n_{j-1}}, L - \frac{1}{j}\right\} \leq s_{n_j} < L \quad \text{for} \quad j = 2, \ldots, k-1. \tag{1}$$

   Choose $\epsilon = \max\left\{s_{n_{k-1}}, L - \frac{1}{k}\right\}$. Select $n_k > n_{k-1}$ so that condition (1) is satisfied for $j = k$.

This two-step procedure constructs a monotone subsequence by design. In a similar fashion as the proof of the theorem, the sequence must also converge to L.

2. Again we define $A_M = \{n \in \mathbb{N} \mid s_n > M\}$, each of which is infinite. Choose $n_1 \in A_1$ as before. We then choose $n_2 \in A_2$ so that $n_2 > n_1$ and $s_{n_2} > s_{n_1}$. This can be done simply because (s_n) is unbounded above.

   Given n_{k-1}, we choose n_k so large that $n_k > n_{k-1}$ and $s_{n_k} > \max\{s_{n_{k-1}}, k\}$ - again possible by the unboundedness of (s_n).

3. Again, this case is proven in a similar fashion to (2).

□

There are a number of (sometimes very surprising!) results that immediately follow from this proposition.

Example 4.15. Because $\mathbb{Q}$ is denumerable, we can find a bijection $f : \mathbb{N} \to \mathbb{Q}$. That is, (f_n) is a sequence that "counts" the rationals (this is sometimes called an *enumeration* of $\mathbb{Q}$).

Choose any rational number $r \in \mathbb{R}$. Create the sets $S_\epsilon = \{n \in \mathbb{N} \mid |f_n - r| < \epsilon\}$. This proposition in the set can be rewritten as $r - \epsilon < f_n < r + \epsilon$ - that is, $f_n \in (r - \epsilon, r + \epsilon)$.

Claim: There exists a subsequence (f_{n_k}) that converges to r.

Proof. All we need to do by the proposition is show that each set S_ϵ is infinite. Observe that the density of $\mathbb{Q}$ shows that there is **at least one** rational $R_1 \in (r - \epsilon, r + \epsilon)$. Then density again allows us to find yet another rational $R_2 \in (R_1, r+\epsilon) \subset (r-\epsilon, r+\epsilon)$. Cutting the interval smaller and smaller in this fashion, we have shown a way to identify infinitely many rational numbers (R_n) inside the interval $(r-\epsilon, r+\epsilon)$. Hence, the proposition tells us that there exists a subsequence of (f_n) that converges (monotonically) to any $r \in \mathbb{R}$. □

To summarize what we've just shown: *For any enumeration (f_n) of $\mathbb{Q}$, each $r \in \mathbb{R}$ can be realized as the limit of some subsequence of (f_n).*

Example 4.16. Let (s_n) be a sequence of positive real numbers such that $\inf\{s_n \mid n \in \mathbb{N}\} = 0$. We simply take S_ϵ to have $L = 0$ and show that each S_ϵ is infinite.

If there was an $\epsilon_0 > 0$ where $S_\epsilon \neq \emptyset$ was finite ($S_{\epsilon_0} = \emptyset$ means all $s_n > \epsilon_0$), then $\inf\{s_n \mid n \in \mathbb{N}\} = \min S_{\epsilon_0} > 0$ (why?). This contradiction shows that S_ϵ is infinite for each $\epsilon > 0$ so that the conditions of (1) in the theorem are satisfied.

The theorem then guarantees us a "path" of positive numbers that decreases (they definitely don't increase) monotonically to zero.

Our goal is to eventually characterize the limit superior and limit inferior in terms of subsequences. In order to do this, we need to set up a little more foundational material.

Proposition 4.9. If (s_n) converges, then every subsequence converges to the same limit.

Proof. Suppose indeed that $s_n \to L$. Let $\epsilon > 0$ and let (s_{n_k}) be any subsequence of (s_n). It is clear that $n_k \geq k$ for any $k \in \mathbb{N}$ (why?). Then because $s_n \to L$, there exists an $N \in \mathbb{N}$ such that $|s_n - L| < \epsilon$ for all $n \geq N$. Oberve that the same N shows that $|s_{n_k} - L| < \epsilon$ for all $k \geq N$ (since $n_k \geq k \geq N$). Hence, $s_{n_k} \to L$ as was to be shown. □

We alluded to this theorem in one of our examples above.

Proposition 4.10. Every sequence (s_n) has a monotonic subsequence.

Proof. Call the n^{th} term of (s_n) *dominant* if it is greater than every subsequent term. That is,

$$s_m < s_n \quad \text{for all} \quad m > n.$$

Case 1: Suppose there are infinitely many dominant terms. The subsequence we are looking for consists of these terms. It can be constructed so that $s_{n_{k+1}} < s_{n_k}$ for all $k \in \mathbb{N}$. That is, (s_{n_k}) is decreasing subsequence of (s_n).

Case 2: If there are only finitely many dominant terms, choose n_1 so be the first index beyond the last dominant term. Because s_{n_1} isn't dominant, there exists an index $m > n_1$ such that $s_m \geq s_{n_1}$. Set $n_2 = m$. Because s_{n_2} isn't a peak either, we can find $n_3 > n_2$ where $s_{n_3} \geq s_{n_2}$. Continuing in this fashion, we can construct an increasing subsequence (s_{n_k}) of (s_n). □

We finally come to the famous theorem of Bolzano and Weierstrass. It is now an immediate consequence of our prior work.

Theorem 4.11. (Bolzano-Weierstrass) Every bounded sequence of real numbers has a convergent subsequence.

Proof. By the previous proposition, the bounded sequence must have a monotone subsequence. By the Monotone Convergence Theorem, this subsequence must converge. □

Similar theorems hold for other kinds of sequences as well. For example, there is a Bolzano-Weierstrass Theorem for sequences in the the plane $\mathbb{R}^2$ (and more generally $\mathbb{R}^n$). The idea of having a convergent subsequence carries over to other settings, but monotonicity doesn't in general. So we acknowledge here that our proof doesn't necessarily extend to general spaces. Hence, having a convergent subsequence is in some sense more "universal" than monotonicity.

Definition 4.8. Let (s_n) be a real sequence. A *subsequential limit* $L \in \widehat{\mathbb{R}}$ is the limit of some subsequence of (s_n).

It should be noted that a convergent sequence has exactly one subsequential limit. If the sequence diverges, more interesting things can occur.

Example 4.17.

1. The sequence $\left(\sin\left(\frac{\pi n}{3}\right)\right)$ has three(!) subsequential limits: $\pm\frac{\sqrt{3}}{2}, 0$.
2. The sequence $((-1)^n n)$ has two subsequential limits: $\pm\infty$.
3. Any enumeration of $\mathbb{Q}$ has (uncountably) infinitely many subsequential limits.
4. It's also possible to have finite **and** infinite subsequential limits: $(n^{(-1)^n})$ has subsequential limits $+\infty$ and 0.

There's an uncanny resemblance between subsequential limits and the limit superior and limit inferior - a resemblance made clear by the following proposition.

Proposition 4.11. Let (s_n) be a sequence. There exist a monotone subsequence whose limit is $\limsup s_n$ and a monotone subsequence whose limit is $\liminf s_n$.

Proof. We use the same notation from before that $u_N = \inf\{s_n \mid n \geq N\}$, $v_N = \sup\{s_n \mid n \geq N\}$, $u_N \to u$, and $v_N \to v$.

- **Case (s_n) unbounded above or below**: If (s_n) is unbounded above, then we know that $\limsup s_n = +\infty$. We've already shown that such sequences have a monotone subsequence diverging to $+\infty = \limsup s_n$.

  Similarly, when (s_n) is unbounded below, then there exists a monotone subsequence of (s_n) diverging to $-\infty$.

- **Case (s_n) bounded above or below**: Let $\epsilon > 0$. If (s_n) is bounded above, then $\limsup s_n = v \in \mathbb{R}$. There must exist an $N_1 \in \mathbb{N}$ such that $v_n < u+\epsilon$ for all $n \geq N_1$. Hence, $s_n < v + \epsilon$ for all $n \geq N_1$.

  In light of a previous proposition, we consider the set

  $$S = \{n \in \mathbb{N} \mid v - \epsilon < s_n < v + \epsilon\}.$$

  If S is finite, there must exist an $N_2 \in \mathbb{N}$ with $N_2 > N_1$ such that $s_n \leq v - \epsilon$ whenever $n \geq N_2$. With such an upper bound for the s_n terms where $n \geq N_2$, we have $v_n \leq u - \epsilon$ for each $n \geq N_2$. This shows that $v < v - \epsilon$ for any $\epsilon > 0$, which is absurd. Hence, S is infinite for all $\epsilon > 0$ so that there must exist a monotone subsequence of (s_n) converging to u.

  The case where (s_n) is bounded below is proved similarly.

□

Proposition 4.12. Let (s_n) be a sequence and S denote the set of its subsequential limits.

1. $S \neq \emptyset$,
2. $\sup S = \limsup s_n$ and $\inf S = \liminf s_n$, and
3. $\lim s_n$ exists if and only if S has exactly one element, namely $\lim s_n$.

Proof.

1. Because $\limsup s_n$ and $\liminf s_n$ both exist (in $\widehat{\mathbb{R}}$!) as limits of subsequences of (s_n). Hence $S \neq \emptyset$ as desired.

2. Let $L \in S$ be the limit of a subsequence (s_{n_k}) of (s_n). Because (s_{n_k}) is convergent, $\limsup s_{n_k} = \liminf s_{n_k} = L$. We acknowledge the obvious fact that $\{s_{n_k} \mid k \geq N\} \subseteq \{s_n \mid n \geq N\}$ for all $N \in \mathbb{N}$. Then we must have

   $$\liminf s_n \leq \liminf s_{n_k} = L = \limsup s_{n_k} \leq \limsup s_n$$

   Because this chain of inequalities was independent of L, we can replace the L in the middle by the inquality $\inf S \leq \sup S$ so that

   $$\liminf s_n \leq \inf S \leq \sup S \leq \limsup s_n.$$

   Since $\limsup s_n$ and $\liminf S$ are both in S, we conclude that

   $$\limsup s_n = \sup S \quad \text{and} \quad \liminf s_n = \inf S.$$

3. The fact that S contains only one element, call it L, we conclude that $\liminf s_n = L = \limsup s_n$ so that (s_n). This is equivalent to $s_n \to L$ by our work in the previous section.

□

With these propositions in tow, we can more intuitively understand the limit superior and limit inferior: **They are the greatest and smallest subsequential limits, respectively, of a sequence**. This agrees with our earlier thoughts about these being the extreme values where infinitely many terms of the sequence accumulate. Hence, another (sometimes easier) way to determine $\liminf s_n$ and $\limsup s_n$ is to examine all subsequential limits of (s_n) and find the infimum and supremum of them, respectively.

The next proposition shows that the limit of a sequence of subsequential limits is once again a subsequential limit - that is, the set of real(!) subsequential limits is a *sequentially closed set.*

Proposition 4.13. Let S be the set of subseqential limits of a sequence (s_n). If (t_n) is a sequence in $S \cap \mathbb{R}$ with $t_n \to L$, then $L \in S$.

Proof. Because the limit is possibly infinite, we divide our proof into the cases where $L \in \mathbb{R}$ and when $L = \pm\infty$.

- **Case** $L \in \mathbb{R}$: For any $\epsilon > 0$, the interval $(L-\epsilon, L+\epsilon)$ must contain infinitely many terms of the sequence. Choose one of them, say t_n. Define $\delta = \min\{\epsilon - (t_n - L), \epsilon + (t_n - L)\}$. Hence, $(t_n - \delta, t_n + \delta) \subseteq (L - \epsilon, L + \epsilon)$.

  Because t_n is a subsequential limit, the interval $(t_n - \delta, t_n + \delta)$ is infinite so that $(L - \epsilon, L + \epsilon)$ is also infinite. Therefore, L must also be a subsequential limit so that $L \in S$

- If $L = \pm\infty$, then (s_n) is unbounded either above or below. Either way, there must exist a (monotone) subsequence converging to L.

$\square$

4.6 The Special Case of Series

This section could be a chapter in its own rite, but we incorporate it into this chapter for a very simple reason: series are merely the limits of sequences of partial summations. What makes their treatment look so much different that of sequences is that we don't necessarily use the terms of the sequence to study them - rather we use the terms of the summation since they are more easily handled in practice. Our goal is then to see what follows about sequences of partial sums when we only examine the terms of the summation.

4.6.1 From Sequences to Series

In case we don't remember this from Calculus, we recall a few preliminary definitions.

Definition 4.9. Let (a_n) be a sequence of real numbers indexed starting at $n = m \in \mathbb{Z}$. Define the sequence

$$s_n = a_m + a_{m+1} + a_{m+2} + \cdots + a_{m+n} = \sum_{k=m}^{n} a_k,$$

also starting at $n = m$. The formal limit $S = \lim s_n$ is called a *formal series* and is denoted

$$S = \lim s_n = \lim_{n\to\infty} \left(\sum_{k=m}^{n} a_k \right) = \sum_{k=m}^{\infty} a_k = \sum a_n$$

The term s_n is called the n^{th} *partial sum of* S.

If the limit of a formal series exists and is real, we say the series *converges*. Otherwise, we say that it *diverges*.

Definition 4.10. Let $S = \sum a_n$ be a series. If $\sum |a_n|$ converges, we say that S is *absolutely convergent*. If S converges but $\sum |a_n|$ diverges, we say that S is *conditionally convergent*.

Example 4.18. Recall the classic example of the *geometric series*. The partial sums are given as

$$\sum_{k=0}^{n} ar^k = a + ar + ar^2 + \cdots + ar^n$$

for constants $a, r \in \mathbb{R}$. When $r \neq 1$, we see that

$$\sum_{k=0}^{n} ar^k = a\frac{1-r^{n+1}}{1-r};$$

if $r = 1$, then this sum is merely an.

When $|r| < 1$, the series (absolutely) converges to

$$S = \sum ar^n = \frac{a}{1-r}$$

since $|r^n| \to 0$ in this case. If $a \neq 0$ and $|r| \geq 1$, then this series diverges.

Example 4.19. Another standard example of series are the so-called *p-series*. These are defined by the partial sums

$$\sum_{k=1}^{n} \frac{1}{k^p},$$

where $p \in \mathbb{R}$. The series (absolutely) converges if and only if $p > 1$.

For example, it can be shown that

$$\sum \frac{1}{n^2} = \frac{\pi^2}{6}$$

and

$$\sum \frac{1}{n^4} = \frac{\pi^4}{90}.$$

However, when $p = 1$ the series is called the *harmonic series*, and it diverges to $+\infty$.

4.6.2 Testing Absolute Convergence

It is important to realize that it is often easier to prove that a series converges than it is to find the limit. To cater to this issue, we use the same notions convergence as before.

Definition 4.11. A formal series $\sum a_n$ satisfies the *Cauchy criterion* if its sequence of partial sums (s_n) is Cauchy:

for all $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $m, n \geq N$ implies $|s_n - s_m| < \epsilon$.

There is nothing wrong with assuming that $n > m$ since the case where $n = m$ trivially satisfies the criterion for any series. Hence, we can rewrite the Cauchy criterion as

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

Because series are just a special case of sequences of real numbers, we can apply all the same work from before to them.

Proposition 4.14. A series converges if and only if it satisfies the Cauchy criterion.

Theorem 4.12. (Divergence Test) If the terms of the sequence don't converge to zero, then the series diverges.

Proof. (We prove the contrapositive of this statement.) Let $\epsilon > 0$, and suppose the series $\sum a_n$ converges. Then it satisfies the Cauchy criterion so that there exists an $N \in \mathbb{N}$.

$$\left| \sum_{k=m}^{n} a_k \right| = |a_m + a_{m+1} + \cdots + a_n| < \epsilon$$

for all $m, n \geq N$. Specifically, this holds for $n = m + 1$ so that $|a_m| = |a_m - 0| < \epsilon$ for all $m \geq N$. Hence, $a_n \to 0$ as was to be shown. $\square$

Although it is quite tempting to use the converse of this statement (that if the terms tend to zero, then the series converges), but the harmonic series is a counterexample to this fact. Since the Divergence Test isn't enough to show that a series converges, we need to devise other means of telling when a series converges. A simple test for convergence of a series with to (1) bound it above by a series with nonnegative terms that is known to converge or (2) bound it below by a series with nonnegative terms that is know to diverge.

Theorem 4.13. (Comparison Test) Let $\sum a_n$ be a formal series where all $a_n \geq 0$.

1. If $\sum a_n$ converges and $|b_n| \leq a_n$ for all n, then $\sum b_n$ converges as well.
2. If $\sum a_n = +\infty$ and $b_n \geq a_n$ for all n, then $\sum b_n = +\infty$ as well.

Proof.

1. Let $\epsilon > 0$. Because $\sum a_n$ converges, the Cauchy criterion gives us an N for this particular ϵ. Then for all $n, m \geq N$, we have

$$\left| \sum_{k=m}^{n} b_k \right| \leq \sum_{k=m}^{n} |b_k| \leq \sum_{k=m}^{n} a_k < \epsilon$$

by repeated application of the triangle inequality and the use of our assumption. Then $\sum b_n$ also satisfies the Cauchy criterion for the same N. Hence, $\sum b_n$ also converges.

2. Let (s_n) and (t_n) be the sequences of partial sums for $\sum a_n$ and $\sum b_n$, respectively. Then $b_n \geq a_n$, it is evident that $t_n \geq s_n$. Because $s_n \to +\infty$, we must also have $t_n \to +\infty$ so that $\sum b_n = +\infty$ as well.

$\square$

Corollary 4.4. Absolutely convergent series converge.

Proof. Let $\sum b_n$ be an absolutely convergent series. Then take $a_n = |b_n|$. Then $|b_n| \leq a_n$ by contruction so that $\sum b_n$ must converge. $\square$

In order to establish our "repository of series of known convergence" that we can compare against, we need to show a property about ratios of consecutive terms in a sequence.

Proposition 4.15. Let (s_n) be any sequence of nonzero real numbers. Then we have

$$\liminf \left|\frac{s_{n+1}}{s_n}\right| \leq \liminf |s_n|^{1/n} \leq \limsup |s_n|^{1/n} \leq \limsup \left|\frac{s_{n+1}}{s_n}\right|.$$

Proof. It is immediate that $\liminf |s_n|^{1/n} \leq \limsup |s_n|^{1/n}$. We now prove the final inequality.

Let $\alpha = \limsup |s_n|^{1/n}$ and $L = \limsup \left|\frac{s_{n+1}}{s_n}\right|$. If $L = +\infty$, then we are done; so we assume that $L < +\infty$.

Let $M > L$. By its definition

$$L = \limsup \left|\frac{s_{n+1}}{s_n}\right| = {}_{N\to\infty} \sup \left\{\left|\frac{s_{n+1}}{s_n}\right| : n \geq N\right\} < M.$$

Hence, there must be an $N \in \mathbb{N}$ such that

$$\sup \left\{\left|\frac{s_{n+1}}{s_n}\right|\right\} < M.$$

That is to say that $\left|\frac{s_{n+1}}{s_n}\right| < M$ for all $n \geq N$. When $n \geq N$, we can write

$$|s_n| = \left|\frac{s_n}{s_{n-1}}\right| \cdot \left|\frac{s_{n-1}}{s_{n-2}}\right| \cdot \dots \cdot \left|\frac{s_{N+1}}{s_N}\right| \cdot |s_N|.$$

There are $n - N$ terms in this expression, so

$$|s_n| < M^{n-N}|s_N|$$

whenever $n \geq N$; that is,

$$|s_n|^{1/n} < M^{-N}|s_N|^{1/n}.$$

Because $M^{-N}|s_N| > 0$, the right-hand side tends to 1 in the limit. Hence, $\alpha \leq M$ so that $\alpha \leq L$.

The first inequality is shown similarly. $\square$

Corollary 4.5. If $\lim \left|\frac{s_{n+1}}{s_n}\right|$ exists, then $\lim |s_n|^{1/n}$ as well. Moreover, the two limits are equal.

Proof. Let $\lim \left|\frac{s_{n+1}}{s_n}\right| = L$. Then the four values in the proposition are the same, hence equal to L. Since the middle inequality is actually an equality, $\lim |s_n|^{1/n}$ exists and is equal to L. $\square$

We can now state perhaps the most important of the series tests.

Theorem 4.14. (Ratio Test) A series $\sum a_n$ of nonzero terms

1. converges absolutely if $\limsup \left|\frac{a_{n+1}}{a_n}\right| < 1$, and
2. diverges if $\liminf \left|\frac{a_{n+1}}{a_n}\right| > 1$.
3. Otherwise $\liminf \left|\frac{a_{n+1}}{a_n}\right| \le 1 \le \limsup \left|\frac{a_{n+1}}{a_n}\right|$ and the test is inconclusive.

Theorem 4.15. (Root Test) Let $\sum a_n$ be a series and let $\alpha = \limsup |a_n|^{1/n}$. The series $\sum a_n$

1. converges absolutely if $\alpha < 1$, and
2. diverges if $\alpha > 1$.
3. Otherwise $\alpha = 1$ and the test is inconclusive.

Both of these theorems are proven in a similar fashion. We prove the Root Test first and then extend that proof to prove the Ratio Test.

Proof. (Proof of the Root Test)

1. Take $\alpha < 1$ and choose $\epsilon > 0$ so that $\alpha + \epsilon < 1$. Then there is an $N \in \mathbb{N}$ satisfying
$$\alpha - \epsilon < \sup\{|a_n|^{1/n} \mid n \ge N\} < \alpha + \epsilon.$$
Specifically, when $n \ge N$, $|a_n|^{1/n} < \alpha + \epsilon$ so that
$$|a_n| < (\alpha + \epsilon)^n.$$
Because $0 < \alpha + \epsilon < 1$, the geometric series $\sum_{n=N+1}^{\infty} (\alpha + \epsilon)^n$ converges and bounds the series $\sum_{n=N+1}^{\infty} a_n$ so that it, too, converges. Hence, by adding finitely many terms, the series $\sum a_n$ converges.
2. When $\alpha > 1$, α is a subsequential limit of $(|a_n|^{1/n})$. Hence, $|a_n| > 1$ for infinitely many a_n so that $a_n \not\to 0$. Then the Divergence Test shows that $\sum a_n$ diverges.
3. As a counterexample, consider the series $\sum \frac{1}{n}$ and $\sum \frac{1}{n^2}$. It can be shown that $\alpha = 1$ for both of these series. However, the former diverges while the latter converges as it is a p-series for $p > 1$. Hence, the borderline case of $\alpha = 1$ doesn't dictate the convergence of the series.

□

Proof. (Proof of the Ratio Test) Again let $\alpha = \limsup |a_n|^{1/n}$. The proposition before the Root Test shows that
$$\liminf \left|\frac{a_{n+1}}{a_n}\right| \le \alpha \le \limsup \left|\frac{a_{n+1}}{a_n}\right|.$$
When $\limsup \left|\frac{a_{n+1}}{a_n}\right| < 1$, then $\alpha < 1$ and the series converges by the Root Test. When $\liminf \left|\frac{a_{n+1}}{a_n}\right| > 1$, then $\alpha > 1$ and the Root Test again tells us that the series diverges. The same conclusion from (3) is reached by the proof above. □

4.6.3 Alternating Series and Integral Tests

This section is the "grab bag" of leftover ideas that don't really fit in with the other sections. The tests we outline are designed to pick up the slack where the previous tests have proven ineffective or inconclusive. Their contents fit in particularly well within the study of Taylor series - not the series of function themselves, but rather the individual real sequences they bind together.

Theorem 4.16. (Integral Test) Let $f\colon (0,\infty) \to [0,\infty)$ be a nonincreadsing, continuous function, and define $a_n = f(n)$ for $n \in \mathbb{N}$. Then the series $\sum a_n$ converges if and only if $\int_1^\infty f(x)\,dx$ converges. Additionally, the series $\sum a_n$ diverges to $+\infty$ if and only if $\int_1^\infty f(x)\,dx$ diverges to $+\infty$.

Proof. This is simply an application of the Comparison Test.

"$\Rightarrow$": Suppose $\sum a_n$ converges to some limit $L \in \mathbb{R}$. Consider that $\sum a_n$ is merely the left-endpoint (upper) Riemann sum for the integral

$$0 \leq \int_1^\infty f(x)\,dx \leq \sum_{n=1}^\infty f(n) \cdot \Delta x = L,$$

where $\Delta x = 1$. Then because f is non-increasing, the integral is a monotone increasing. Since it is bounded above by the series, the integral must indeed converge to some value (that the test does not determine!) between 0 and L.

"$\Leftarrow$": Now suppose that $\int_1^\infty f(x)\,dx$ converges to the value $L \in \mathbb{R}$. This time, the series $\sum a_n$ may be seen as a right-endpoint (lower) Riemann sum starting at $x = 0$ instead. That is,

$$0 \leq \sum_{n=1}^\infty f(n) \cdot \Delta x \leq \int_1^\infty f(x)\,dx = L$$

with $\Delta x = 1$ again. This time the series is monotone non-decreasing and bounded above. Monotone Convergence tells us that the series must converge as well.

The proof for the diverging series follows similarly as a comparison with appropriate (divergent) lower bounds. $\square$

The other test is for alternating series.

Definition 4.12. A series $\sum a_n$ is called *alternating* if the terms can be written as $a_n = (-1)^{n+1} b_n$ for some nonnegative sequence (b_n).

Theorem 4.17. (Alternating Series Test) If the non-increasing sequence (a_n) of terms in the alternating series $\sum (-1)^{n+1} a_n$ converge to zero, then the series converges (to $s \in \mathbb{R}$, say) as well. Moreover, the partial sums $s_n = \sum_{k=1}^n a_n$ satisfy $|s - s_n| \leq a_n$ for all $n \in \mathbb{N}$.

Proof. To show the convergence of the series, consider that the subsequence (s_{2n}) of partial sums is increasing (since $s_{2n+2} - s_{2n} = a_{2n+2} + a_{2n+1} \geq 0$. In a similar fashion, (s_{2n-1}) is a decreasing sequence. The claim then is that the odd-index partial sums always bound the even-index partial sums. That is, $s_{2m} \leq s_{2n+1}$ for all $m, n \in \mathbb{N}$.

To see this, observe that $s_{2n} \leq s_{2n+1}$ for all $n \in \mathbb{N}$ since $s_{2n+1} - s_{2n} = a_{2n+1} \geq 0$. When $m \leq n$, we see $s_{2m} \leq s_{2n} \leq s_{2n+1}$ since (s_{2n}) is increasing. Otherwise, $s_{2n+1} \geq s_{2m+1} \leq s_{2m}$ because (s_{2n-1}) is decreasing. In particular, we have the following: (s_{2n}) is increasing and bounded above by s_3, and (s_{2n+1}) is decreasing and bounded below by

s_2 - hence, Monotone Convergence shows the subsequential limits exist. Given the two subsequential limits $s = \lim s_{2n}$ and $t = \lim s_{2n+1}$, we see that

$$t - s = \lim s_{2n+1} - \lim s_n = \lim(s_{2n+1} - s_{2n}) = \lim a_{2n+1} = 0.$$

Hence, $t - s = 0$ so that $s = t$.

Because $s_{2n} \leq s \leq s_{2n+1}$, we see that $s_{2n+1} - s$ and $s - s_{2n}$ are both bounded by $s_{2n+1} - s_{2n} = a_{2n+1} \leq a_{2n}$. Then $|s - s_n| \leq a_n \to 0$ so that $s_n \to s$. □

4.6.4 Examples

Now that we've established several ways that we can study series, we should set out a few examples that may or may not look familiar from Calculus.

Example 4.20. Because we used the Geometric Series to prove the Ratio Test (and, subsequently, the Root Test as well), we cannot actually use either of these tests to determine that a Geometric Series converges or diverges. Thankfully we have already shown in a previous example that such series converge if and only if the growth constant r falls within the interval $(-1, 1)$.

For the series with $r < 0$, the Alternating Series Test is also valid. When $r \leq -1$, the function $f(x) = r^x$ is certainly non-decreasing; but when $r \in (-1, 0)$, f is decreasing so that the alternating series $\sum r^n$ converges.

Geometric series are also one of the few series which may compute. Recall that if the series converges, then we have

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}.$$

Example 4.21. Recall the p-series $\sum \frac{1}{n^p}$. As promised earlier, we determine when such a series converges now. This can be done easily with the Integral Test. Letting $f(x) = x^{-p}$ for $p \neq 1$, we compute

$$\int_1^{\infty} x^{-p}\, dx = \lim_{x \to \infty} \frac{1}{p-1} \left[1 - x^{1-p}\right].$$

Hence, the p-series converges (to $\frac{1}{p-1}$) if and only if $p > 1$.

Moreover, if $p = 1$, then the integral is that of the natural logarithm. Hence, the series diverges in this case as well.

We now have our established base of known convergent and divergent series that we can call upon for comparison. Let's try them out, shall we?

Example 4.22. Consider the series $\sum a_n = \sum \frac{2n}{n^3 - 9}$. Just as we proved the limit for a sequence determined by a rational function $\frac{p(n)}{q(n)}$, we can make similar comparisons to bound it and show that the series converges.

Consider that a_n can be bounded above by $\frac{2n}{n^3/2} = \frac{4}{n^2}$ when $n \geq 3$. Hence, we can compare $\sum a_n$ against the convergent p-series $\sum \frac{1}{n^2}$ to conclude that indeed $\sum a_n$ converges (though to what limit we do not know).

Example 4.23. Consider the series $\sum a_n = \sum \frac{n^3}{n^4+1}$. We appeal to the uber-powerful Ratio Test here. We consider the ratio

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{(n+1)^3}{(n+1)^4 + 1} \cdot \frac{n^4+1}{n^3}\right| = \left(\frac{n+1}{n}\right)^3 \cdot \frac{n^4+1}{(n+1)^4+1}$$

. However, the limit of this sequence of absolute ratios tends to 1 - hence, the Ratio Test tells us nothing (NOOOOOO!!!!!). However, we observe that this series resembles the (divergent) p-series $\sum \frac{1}{n}$ - the harmonic series - so we try to compare against this series.

To do this, consider that $n^4 + 1 \leq n^4 + n^4 = 2n^4$ for $n \geq 1$. Then we see that $a_n = \frac{n^3}{n^4+1} \geq \frac{n^3}{2n^4} = \frac{1}{2n}$. Hence, because $\frac{1}{2}\sum \frac{1}{n}$ diverges and is a lower bound for $\sum a_n$, we see that $\sum a_n$ has no choice but to diverge as well.

Example 4.24. In light of the previous example, rational functions may not be the best time to apply the Ratio (or Root) Test. Here is a time where it turns out to be much more useful.

Given the series $\sum a_n = \sum \frac{n}{3^n}$, we consider

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{n+1}{3^{n+1}} \cdot \frac{3^n}{n}\right| = \frac{n+1}{3n} \to \frac{1}{3}$$

so that $\limsup \left|\frac{a_{n+1}}{a_n}\right| = \lim \left|\frac{a_{n+1}}{a_n}\right| = \frac{1}{3} < 1$. Therefore, the series must converge (absolutely).

It turns out that this series may be compared against certain Geometric Series by bounding the numerator by a suitable exponential function.

Example 4.25. Now we come across a sequence that would be hard to deal with using only the knowledge from introductory Calculus - find a series that works for the Root Test works when the Ratio Test fails. Let's take a look at the series $\sum a_n = \sum 2^{(-1)^n - n}$. Yes, this series can easily be compared to the (convergent) Geometric Series $\sum 2^{1-n} = \sum 2\left(\frac{1}{2}\right)^n$; however, that was not the point of the example.

Observe that

$$\left|\frac{a_{n+1}}{a_n}\right| = \begin{cases} \frac{1}{8}, & n \text{ even} \\ 2, & n \text{ odd.} \end{cases}$$

This allows us to conclude that

$$\frac{1}{8} = \liminf \left|\frac{a_{n+1}}{a_n}\right| < 1 < \limsup \left|\frac{a_{n+1}}{a_n}\right| = 2,$$

wherein the Ratio Test is useless.

However, we see that

$$a_n^{1/n} = \begin{cases} 2^{\frac{1}{n}-1}, & n \text{ even} \\ 2^{-\frac{1}{n}-1}, & n \text{ odd.} \end{cases}$$

Hence, $a_n^{1/n} \to \frac{1}{2}$ so that $\limsup a_n^{1/n} = \frac{1}{2} < 1$ wherein the Ratio Test concludes that the series converges (absolutely)!

4.6.5 Rearrangements of Series

There is a rather striking peculiarity that is seen in series that usually isn't covered in fundamental Calculus as students lack the rigor and tools to be able to explain it. It is motivated by the following:

> It can be shown (using a Taylor series) that $\sum \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln 2$. However, it can also be shown that by rearranging the terms of this

series, the alternating harmonic series can also sum to $e \cong 2.718281828\ldots$. Yet another rearrangement can sum this series to $1\ldots$ and another can yield -20. In fact, the alternating harmonic series can be made to converge to **any** real number at all - and this includes $\pm\infty$, which aren't even real numbers!

We only need to recall the definition of a *permutation* in order to state the theorem we want to prove.

Definition 4.13. Let S be a set. A *permutation of S* is merely a bijection $\sigma\colon A \to A$; that is, a one-to-one and onto function of A to itself.

The following theorem is due to Riemann.

Theorem 4.18. (Riemann Rearrangement) Let $\sum a_n$ be a conditionally convergent series, and choose any number $M \in \hat{\mathbb{R}}$. There exists a permutation $\sigma\mathbb{N} \to \mathbb{N}$ such that $\sum a_{\sigma(n)} = M$.

Proof. Because $\sum a_n$ is conditionally convergent, there must be infinitely many terms that are positive and infinitely many terms that are negative. Otherwise, there would be a tail whose absolute series converges, thus contradicting the fact that $\sum |a_n|$ must diverge.

Divide the terms of the series into those two categories thusly:

$$a_n^+ = \frac{a_n + |a_n|}{2}, \quad a_n^- = \frac{a_n - |a_n|}{2}$$

so that $a_n^+ = 0$ if $a_n \le 0$ and $a_n^- = 0$ if $a_n \ge 0$. We claim that both series $\sum a_n^+$ and $\sum a_n^-$ diverge to their respective infinities.

To see this, consider if one of these series converges: The series $\sum a_n$ would then diverge because it would have a divergent tail. If both converged, the series would be absolutely convergent. Hence, both positive- and negative-term series must diverge. It is at this point that we can state that if $M = \pm\infty$, then we can just take a permutation that sums the positive or negative terms, respectively, before the others. [There is a subtlety in being able to do this, but it is a minor point and will not be addressed further here.]

Then we choose $M \in \mathbb{R}$. We construct the permuatation σ inductively. First, we add positive terms until our partial sum exceeds M. We keep track of which indices of a_n are used here and "fill in" σ with the appropriate indices. We then add negative terms until the partial sum falls below M, keeping track of which indices are added along the way. The idea is to continue this process ad infinitum. Our job is to show that this process yields the result we want. That is, we need to show that

1. It is possible to continue this process ad infinitum, and

2. The process yield a series whose limit is M.

The first of these points is not difficult to establish. Because both $\sum a_n^+$ and $\sum a_n^-$ diverge, there are certainly enough terms in the respective series to push our rearrangement back and forth past M as we'd like.

The second point is equally easy to establish. Because $\sum a_n$ converges, the terms $a_n \to 0$. Hence, the amount that each term contributes to the partial sum can be controlled by an $\epsilon > 0$. That is, because σ is a permutation, the a_n terms are all accounted for in the rearrangement and don't repeat along the way. That is, for any given $N \in \mathbb{N}$, **eventually** N is surpassed as an index of the terms of the rearrangment. Therefore, for any $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that all contributions of to the partial sum **at some step of the process** are less than ϵ. Hence, the Cauchy criterion for series convergence allows us to make the contributions as small as we'd like so that the series must converge. □

That wasn't the most rigorous proof in the world, but it conveys the idea of how the rigorous proof should look if the reader wants more detail.

Students will show in homework that **absolutely convergent series tend to the same limit regardless of rearrangments of the terms of the series**.

5 Topology of Metric Spaces

We finally arrive at the last unit of this course: Topology. In order to do this section justice, I need to explain one of the overarching goals of math.

The first setting in which this theme becomes apparent is in an introductory course on Linear Algebra. That is, when one studies a vector space, one usually prescribes what is called a *basis* for this space. This is what allows us to do **computations** within the space - they describe the coordinate axes. Moreover, there are **infinitely many bases** that we may choose for our vector space.

As such, our view of the space may be skewed depending on the choice of coordinates we make. This is to be expected with the amount of freedom we had in our choice of coordinates. This is perhaps most evident when one learns how to diagonalize a matrix. Eigenvectors prescribe a "preferable" basis of coordinates wherein a particular linear transformation is most easily viewed.

However, it should be noted that the vector space and all it contains exists with or without these coordinates. Come to think of it, maybe we're limiting our scope of vector spaces by describing them with such cumbersome coordinates - a rather unsettling thought indeed! Our goal then is to try to "do linear algebra" **without coordinates**!

The same principle can be applied in numerous other contexts. Within the context of Analysis, we are trying to "do Calculus" without **distances**. The first step in getting to this goal is to figure out which properties of real numbers can be explained without distances - namely, the properties that are critical to the study of Analysis but don't require a notion of distance to describe them. This is what the field of Topology is all about. While metric spaces won't accomplish our end goal, they hash out which properties of real numbers are indispensable in our pursuit of that goal.

5.1 $\mathbb{R}^n$ As A Metric Space

The real numbers are an excellent example of a set that has very desirable properties and can be used effectively in applications. This set is not the only one that is useful in practice. There are many sets that satisfy the same sort of completeness that we've seen in $\mathbb{R}$ - sets like higher-dimensional Euclidean spaces $\mathbb{R}^n$, the set of real functions $\mathbb{R}^{\mathbb{R}}$ and its subsets (among which are the familiar continuous and differentiable functions), and some rather abstract sets that one usually does not consider in practice but are still useful as sources of counterexamples and counter-intuition in advanced Analysis. We will focus on the first of these initially and then move on to the more general setting shortly.

Although $\mathbb{R}$ and $\mathbb{R}^n$ $(n > 1)$ are quite similar, $\mathbb{R}^n$ does not enjoy the same ordering properties that $\mathbb{R}$ does. As these higher-dimensional spaces appear quite frequently in practice, we would still like to determine the extent to which they may be used. We started with a distance function on $\mathbb{R}$, and we are going to need a similar function for $\mathbb{R}^n$. The following definition is a collection of properties that we require such a distance function to satisfy.

Definition 5.1. Let S be a set and suppose we have a fuction $d\colon S \times S \to \mathbb{R}$ that satisfies

1. (Positive definiteness) $d(x,x) = 0$ for all $x \in S$ and $d(x,y) > 0$ for distinct $x, y \in S$,
2. (Symmetry) $d(x,y) = d(y,x)$ for all $x, y \in S$, and
3. (Triangle Inequality) $d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in S$.

The function d is called a *metric* for S, and the pair (S, d) is called a *metric space.*

Because of its abstraction, we should show which sets and functions pair together to give metric spaces.

Example 5.1. 1. Obviously, $(\mathbb{R}, |\cdot|)$ is the metric space that we've been working with this entire time - the $|\cdot|$ is exactly given by $d(x, y) = |x-y|$. It is the literal geometric meaning of distance that we are familiar with. We proved that this metric indeed satisfies all three of the above properties way back at the beginning of Chapter 2 of these notes.

2. We now show that $\mathbb{R}^n$ equipped with the proper distance function is also a metric space. It isn't too surprising that we try to mimic the success of $(\mathbb{R}, |\cdot|)$ by using exactly the same function... But we need to adapt it to this new set. If we have

$$\vec{x} = (x_1, \dots, x_n) \quad \text{and} \quad \vec{y} = (y_1, \dots, y_n),$$

then we can define the distance function d using the absolute value as before:

$$d(\vec{x}, \vec{y}) = \left[\sum_{j=1}^{n} (x_j - y_j)^2\right]^{1/2},$$

which is the *standard* or *Euclidean metric* on $\mathbb{R}^n$. It follows directly from the use of the Pythagorean Theorem for higher dimensions.

To establish that this is a metric, we observe that (1) a sum of square is zero if and only if each term was itself zero to begin with and (2) the squaring function is symmetric so that $(x_j - y_j)^2 = (y_j - x_j)^2$. Hence, the first two properties are satisfied. The triangle inequality is a little trickier to prove, but we show that as well.

Proof. First consider that

$$0 \leq \sum_{k=1}^{n} (x_k - \lambda y_k)^2 = \lambda^2 \sum_{k=1}^{n} y_k^2 + \lambda \sum_{k=1}^{n} (-2) x_k y_k + \sum_{k=1}^{n} x_k^2,$$

where $\lambda \in \mathbb{R}$. The student can easily verify (and really should... do it now) that, for $a, b, c \in \mathbb{R}$, $a\lambda^2 + b\lambda + c \geq 0$ for all λ implies that $b^2 - 4ac \geq 0$. Applying this to the first expression yields the observation that

$$4\left(\sum_{k=1}^{n} x_k y_k\right)^2 - 4\left(\sum_{k=1}^{n} x_k^2\right)\left(\sum_{k=1}^{n} y_k^2\right) \leq 0$$

so that we get the famous *Cauchy-Schwarz Inequality*

$$\sum_{k=1}^{n} x_k y_k \leq \left(\sum_{k=1}^{n} x_k^2\right)^{1/2} \left(\sum_{k=1}^{n} y_k^2\right)^{1/2},$$

which the student may recognize better from the formula $\vec{x} \cdot \vec{y} \leq |\vec{x}||\vec{y}|$.

The final step is to use this inequality as follows:

$$d(\vec{x},\vec{z})^2 = \sum_{k=1}^{n}(x_k - z_k)^2 = \sum_{k=1}^{n}(x_k - y_k + y_k - z_k)^2$$

$$= \sum_{k=1}^{n}[(x_k - y_k)^2 + 2(x_k - y_k)(y_k - z_k) + (y_k - z_k)^2]$$

$$\leq \sum_{k=1}^{n}(x_k - y_k)^2 + 2\left(\sum_{k=1}^{n}(x_k - y_k)^2\right)^{1/2}\left(\sum_{k=1}^{n}(y_k - z_k)^2\right)^{1/2} + \sum_{k=1}^{n}(y_k - z_k)^2$$

$$= \left[\left(\sum_{k=1}(x_k - y_k)^2\right)^{1/2} + \left(\sum_{k=1}^{n}(y_k - z_k)^2\right)^{1/2}\right]^2 = [d(\vec{x},\vec{y})^2 + d(\vec{y},\vec{z})^2]^2,$$

wherein our result follows. □

But let's not lose sight of the fact that this is merely the final step in showing that this standard metric is indeed a metric for $\mathbb{R}^n$.

3. Let's try something a little more abstract. Let S be **any** nonempty set. Define for S the function

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

This is called the *discrete metric* for S. It is simply a function that tells us whether or not two elements of S are the same. It is left to the reader to fill in the details to see that this is indeed a metric for S.

Now that we've defined this object, we would like to be able to mimic what we've done in $\mathbb{R}$. That is, we'd like to be able to consider sequences of elements of metric spaces. As such, we need a notion of convergence for these spaces. Thankfully, it looks very similar to how we define this for $\mathbb{R}$.

Definition 5.2. A sequence (s_n) in a metric space (S,d) *converges* to the *limit* $s \in S$ if $\lim d(s_n, s) = 0$. That is to say that for all $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$n \geq N \quad \text{implies} \quad d(s_n, s) < \epsilon.$$

This sequence is *Cauchy* if for each $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$m, n \geq N \quad \text{implies} \quad d(s_m, s_n) < \epsilon.$$

The metric space (S,d) is said to be *complete* if every Cauchy sequence converges to some element of S.

This notion of completeness is very closely tied to the Completeness Axiom of the real numbers. In fact, we could have used this idea of convergent Cauchy sequences instead of suprema of bounded sets of real numbers and wound up with exactly the same theory and results. Perhaps not surprisingly, these ideas can be shown to be equivalent.

We now show that $\mathbb{R}$ and $\mathbb{R}^n$ share the completeness property. In particular, we determine the convergence of a vector sequence by the convergence of the real sequences of its components.

Proposition 5.1. A sequence $(\vec{x}^{(n)})$ in $\mathbb{R}^k$ converges if and only if for each $j = 1, 2, \ldots, k$, the sequence $(x_j^{(n)})$ converges in $\mathbb{R}$. This sequence is Cauchy if and only if each component sequence is Cauchy in $\mathbb{R}$.

Proof. (Convergence) Let $\epsilon > 0$ and suppose that $(\vec{x}^{(n)})$ converges to some $\vec{X} \in \mathbb{R}^k$ where $\vec{X} = (X_1, \ldots, X_k)$. Then there is some $N \in \mathbb{N}$ such that $n \geq N$ implies that

$$d(\vec{x}^{(n)}, \vec{X})^2 = \sum_{j=1}^{k} (x_j^{(n)} - X_j)^2 < \epsilon.$$

. Observe then that $|x_j^{(n)} - X_j|^2 \leq d(\vec{x}^{(n)} - \vec{X})^2$ for each $j = 1, \ldots, k$ so that each $(x_j^{(n)})$ converges to X_j.

Now let $\epsilon > 0$ and suppose each component sequence $(x_j^{(n)})$ converges to some X_j. For $\frac{\epsilon}{\sqrt{k}}$, there exist N_j for each component sequence $(x_j^{(n)})$ such that $d(x_j^{(n)}, X_j) < \frac{\epsilon}{\sqrt{k}}$ for each $n \geq N_j$. We simply choose $N = \max\{N_1, \ldots, N_k\}$ to guarantee that inequality for each component sequence simultaneously. Then when $n \geq N$, we find

$$d(\vec{x}^{(n)}, \vec{X})^2 = \sum_{j=1}^{k} (x_j^{(n)} - X_j)^2 < k \cdot \frac{\epsilon^2}{k} = \epsilon^2$$

so that $\vec{x}^{(n)} \to \vec{X}$.

(Cauchy) This looks fairly similar to the above. Let $\epsilon > 0$ and suppose that $(\vec{x}^{(n)})$ is Cauchy in $\mathbb{R}^k$. Then there exists an $N \in \mathbb{N}$ such that $m, n \geq N$ implies that $d(\vec{x}^{(m)}, \vec{x}^{(n)}) < \epsilon$. That is to say that

$$(x_j^{(m)} - x_j^{(n)})^2 \leq d(\vec{x}^{(m)}, \vec{x}^{(n)})^2 < \epsilon^2$$

for each j. Hence, each component sequence is also Cauchy for the same pairings of ϵ and N.

Finally suppose that each component sequence is Cauchy in $\mathbb{R}$ and choose $\epsilon > 0$. Consider then that there exist $N_1, \ldots, N_j \in \mathbb{N}$ such that $|x_j^{(m)} - x_j^{(n)}| < \frac{\epsilon}{\sqrt{k}}$ whenever $m, n \geq N_j$. Hence, choosing $N = \max\{N_1, \ldots, N_k\}$, we have as before that

$$d(\vec{x}^{(m)}, \vec{x}^{(n)})^2 < k \cdot \frac{\epsilon^2}{k} = \epsilon^2$$

so that $(\vec{x}^{(n)})$ is Cauchy in $\mathbb{R}^k$. $\square$

With this proposition proven, we've actually done all the heavy lifting for proving facts about $\mathbb{R}^n$. The important ones being about completeness and convergent subsequences.

Theorem 5.1. The metric space $\mathbb{R}^k$ (with the Euclidean metric) is complete.

Proof. Let $(\vec{x}^{(n)})$ be a Cauchy sequence in $\mathbb{R}^k$. Then each of its component sequences is Cauchy in $\mathbb{R}$ by the proposition. Hence, the completeness of $\mathbb{R}$ shows that each of these component sequences converges in $\mathbb{R}$. The proposition then shows that $(\vec{x}^{(n)})$ converges in $\mathbb{R}^k$ as well. $\square$

To state the $\mathbb{R}^n$ Bolzano-Weierstrass Theorem, we need to know what it means for a sequence to be bounded in $\mathbb{R}^n$. Our definition is intuitive but doesn't actually depend on the standard metric.

Definition 5.3. A subset $S \subseteq \mathbb{R}^k$ is *bounded* if there exists an $M > 0$ such that

$$\max\{|x_j| \mid j = 1, \ldots, k\} \leq M$$

for all $\vec{x} = (x_1, \ldots, x_k) \in S$.

That is, there exists a k-dimensional cube (centered at the origin) with side length $2M$ that contains S.

Theorem 5.2. Every bounded sequence in $\mathbb{R}^k$ has a convergent subsequence.

Proof. Let $(\vec{x}^{(n)})$ be a bounded sequence in $\mathbb{R}^k$. Then each component sequence is also bounded in $\mathbb{R}$. The Bolzano-Weierstrass Theorem for $\mathbb{R}$ shows the first of these component sequences has a convergent subsequence $(x_1^{(n)})$. Replace $(\vec{x}^{(n)})$ by the subsequence where $(x_1^{(n)})$ converges. Continue paring down $(\vec{x}^{(n)})$ via the existence of its component sequences' convergent subsequences. Once we do this for $(x_k^{(n)})$, we will have a subsequence of $(\vec{x}^{(n)})$ that guarantees each of its component sequences converges. By the proposition, this subsequence of $(\vec{x}^{(n)})$ must also converge. $\square$

5.2 Open and Closed Sets

Now that we've established that $\mathbb{R}^n$ enjoys just about all the important properties that $\mathbb{R}$ does, we would like to know if other spaces have these properties; and if they don't, we want to know what makes them so different. Let's recall some examples.

Example 5.2.

1. We now know that $\mathbb{R}^n$ is a metric space when we give it the standard Euclidean metric.

2. We saw the *discrete metric* could be applied to any nonempty set S. That is,

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

3. Although we haven't talked about continuity, the student should be quite familiar with the concept so that we can talk about the space of continuous functions $C([0, 1])$ on the closed interval $[0, 1]$. This is a metric space if we say that

$$d(f, g) = \left(\int_0^1 |f(x) - g(x)|^p \, dx \right)^{1/p}$$

for any $1 \leq p \leq +\infty$. In the case of $p = +\infty$, we use the notation

$$d_\infty(f, g) = \max_{x \in [0,1]} |f(x) - g(x)|.$$

That is, there are sets that admit infinitely many metrics!

How can we even begin to study the properties about these other spaces. Are they bounded? Are they complete? Do the accomplish something that $\mathbb{R}^n$ cannot? In order to answer these questions, we need to extend our standard framework in $\mathbb{R}^n$ to a more general setting. One way we can do this is by talking about open and closed subsets of a metric space.

Definition 5.4. Let (S, d) be a metric space and $E \subseteq S$. An element $s_0 \in E$ is *interior to E* if there exists some $r > 0$ such that

$$B_r(s_0) = \{s \in S \mid d(s, s_0) < r\} \subseteq E.$$

The set $B_r(s_0)$ is called the *open ball of radius r centered at s_0*.

The set of all points interior to E is called the *interior* of E and is denoted E^o. We say that E is *open in S* if every point of E is interior to E. That is, $E \subseteq S$ is open if $E^o = E$.

Naturally, if there are open sets, then there must be some notion of "closed" sets. While tempting to say that a closed set is not open, this is not the must useful definition for us to work with. The one presented is much better suited to our needs - we will see why shortly.

Definition 5.5. Let (S, d) be a metric space with $E \subseteq S$. We say E is *closed* if its complement $S \backslash E$ is open in S.

The *closure of E in S*, denoted E^-, is the intersection of all closed sets containing E. The *boundary* of E, denoted ∂E is the set $E^- \backslash E^o$. The elements of ∂E are called *boundary points of E*.

To see why we chose this as our definition, we highlight the properties that result from this definition.

Proposition 5.2. For a metric space (S, d),

(a) The empty set $\emptyset$ and S are both open.

(b) The union of an arbitrary collection of open sets $\{U_\alpha\}$ is also open.

(c) The intersection of a **finite** collection of open sets $\{U_k\}_{k=1}^n$ is also open.

Proof. (a) For the empty set $\emptyset$, the condition for all points of $\emptyset$ to be interior are vacuously satisfied so that $\emptyset$ is open in S. Every point $x \in S$ is interior to S by using the ball $B_1(x)$ so that S is open as well.

(b) Let $x \in \bigcup_\alpha U_\alpha$. Then $x \in U_i$ for some $i = \alpha$ in the collection. Because U_i is open, x is interior to U_i so that there exists a ball $B_r(x) \subseteq U_i$. Therefore, $B_r(x) \subseteq U_i \subseteq \bigcup_\alpha U_\alpha$, so that x is interior to the union as well. Therefore, the union is open.

(c) Let $x \in \bigcap_{k=1}^n U_k$. Because each U_k is open, there exists a ball $B_{r_k}(x) \subseteq U_k$. If we let $r = \min\{r_1, \ldots, r_n\}$, then $B_r(x) \subseteq U_k$ for each $k = 1, \ldots, n$ so that x is interior to the intersection. Therefore, the intersection is open.

$\square$

Corollary 5.1. For a metric space (S, d),

(a) The empty set $\emptyset$ and S itself are both closed.

(b) The union of a **finite** collection of closed sets is also closed.

(c) The intersection of an arbitrary collection of closed sets is also closed.

Proof. (a) follows from noticing that $\emptyset = S \backslash S$ and $S = S \backslash \emptyset$, both of which were shown to be open above. (b) and (c) both follow from an application of DeMorgan's Laws. $\square$

Example 5.3. 1. Following the definition of $B_r(x)$, the open balls in $\mathbb{R}$ (using the standard metric) are merely the open intervals that we are familiar with from Calculus. The open balls in $\mathbb{R}^n$ are the "solid balls" without their shell boundaries - hence the name "ball".

We know from the above that $\mathbb{R}$ is open in itself. That is, any $x \in \mathbb{R}$ has a unit ball centered around it, and that ball is still completely contained inside $\mathbb{R}$. The same argument still applies to $\mathbb{R}^n$ with the Euclidean metric.

2. We can use a different metric for $\mathbb{R}^2$. Define

$$d_\infty((x_1, y_1), (x_2, y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}.$$

This is called the ℓ^∞ metric on $\mathbb{R}^2$. The open balls in this metric are open squares.

We can define yet another metric for $\mathbb{R}^2$. Define

$$d_1((x_1, y_1), (x_2, y_2)) = |x_2 - x_1| + |y_2 - y_1|.$$

This is called the ℓ^1 metric on $\mathbb{R}^2$. The open balls in this metric can be shown to be open diamonds.

3. It is important to note that when we say S is open, we need to specify **in which set** it is open. Case in point: Consider the interval $(0, 1]$. This set is not open in $\mathbb{R}$. (Why?) However, this set is open in the set $[0, 1]$ when we give it the *relative metric*:

> Given a metric space (S, d) and a subset $A \subseteq S$, we define the subspace metric $d' = d|_A$ to be the restriction of d to A. The pair (A, d_A) is a metric space in this way.

It can be elaborated upon that the open balls of $(A, d|_A)$ are exactly the intersections of the open balls of (S, d) intersected with A.

In light of this definition, consider then that for $((0, 1], d_{(0,1]})$ the subsets $(a, 1]$ - for any $a < 1$ - are all open subsets of $(0, 1]$ since they are all the intersection of $(a, 2)$ with $(0, 1]$.

4. The following exemplifies why we restrict the intersection of open sets to the finite. Consider the countably infinite intersection

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

Although each set of the intersection is open (nested as well), the intersection of all of them produces the **closed** set $[0, 1]$.

A similar counterexample may be found for the infinite union of closed sets - the union

$$\bigcup_{k=1}^{\infty} \left[\frac{1}{k}, 1 - \frac{1}{k} \right] = (0, 1),$$

for example.

Remark 5.1. Notice that our definitions of open and closed are different from their colloquial use. That is, **sets are not like doors**. Just because a set is not open, that does not mean that the set is closed and vice versa. Similarly, it is entirely possible for sets to be both open and closed (sometimes called *clopen* - I wish I was kidding...); worse yet, it's possible that a set is neither open nor closed!

We chose these definitions to retain closure under unions. That is, no matter how many times you append an open set to another, the result is still open. In a similar fashion, no matter how many times you intersect a closed set by closed sets, the result is still closed - a fact which we make use of shortly.

The book wants the leave the proofs of the following to the reader as exercises. However, I feel that the student should see examples of how these proofs look so that they may mimic the technique for their own devices later. This proposition shows how we may "humanize" the definition of closure - that is, give it an alternate characterization.

Proposition 5.3. Let (S, d) be a metric space with $E \subseteq S$.

(a) E is closed if and only if $E = E^-$.

(b) E is closed if and only if it contains the limit of every convergent sequence of points in E (not surprisingly, these are called *limit points*).

(c) An element is in E^- if and only if it is the limit of some sequence in E.

(d) A point is in ∂E if and only if it belongs to both E^- and $(S \backslash E)^-$.

Proof. (a) Suppose that E is closed. The definition of E^- allows us to conclude that $E \subseteq E^-$ so that we need only show that $E^- \subseteq E$. Because E is closed, consider that in the definition

$$E^- = \bigcap_{E \subseteq K,\ K \text{ closed}} K$$

E itself is one of these K. Then we see that $E^- \subseteq E$ so that $E = E^-$ as desired.

Now suppose $E = E^-$. The fact that E is closed follows from the corollary above.

(b) Suppose that E is closed, and let $x \in S$ be the limit of some sequence $(x_n) \subseteq E$. Suppose to the contrary that $x \notin E$. Then because $S \backslash E$ is open, there exists some ball $B_r(x) \subseteq S \backslash E$. However, since $x_n \to x$, we must be able to find an $N \in \mathbb{N}$ such that $n \geq N$ implies that $d(x_n, x) < \frac{r}{2}$ - meaning that infinitely many x_n leave E. This contradiction shows that $x \in E$ as well.

Conversely, suppose E contains all of its limit points. Choose any $x \in S \backslash E$. Because x is not a limit point of E, there must exist some ball $B_r(x) \subseteq S \backslash E$. Then $S \backslash E$ is open so that E is closed.

(c) First suppose that $x \in E^-$. If x is not a limit point of E, then there must exist a ball $B_r(x) \subseteq S \backslash E$. Then $S \backslash B_r(x)$ is a closed set that contains E, wherein x must lie. This contradiction shows that x must be a limit point of E.

Now suppose that x is a limit point of E that does not lie in E^-. Because E^- is closed, $S \backslash E^-$ is open so that there exists a ball $B_r(x) \subseteq S \backslash E^-$. That shows that $B_r(x) \cap E = \emptyset$, which contradicts x being a limit point of E.

(d) Suppose that $x \in \partial E$. By definition, $\partial E = E^{-}\backslash E^{o}$ so that $x \in E^{-}$. Because $(S\backslash E)^{-}$ is closed, if $x \notin (S\backslash E)^{-}$, then there exists a ball $B_r(x) \subset S\backslash(S\backslash E)^{-} \subseteq S\backslash(S\backslash E) = E$ so that $x \in E^{o}$. This contradicition shows that $x \in (S\backslash E)^{-}$ as well.

Now suppose that $x \in E^{-} \cap (S\backslash E)^{-}$. Then again we find $x \in E^{-}$ so that we need to show that $x \notin E^{o}$. If $x \in E^{o}$, then there is a ball $B_r(x) \subseteq E$. Hence, $S\backslash E \subseteq S\backslash B_r(x)$. In order for $x \in (S\backslash E)^{-}$, we would require $x \in S\backslash B_r(x)$ since this set is closed and contains $S\backslash E$. This contradiction shows that $x \notin E^{o}$. Hence, $x \in E^{-}\backslash E^{o}$.

$\square$

Example 5.4. We close out the section with a rather remarkable set, one with which the student is perhaps already quite familar. The *Cantor set* is defined as follows:

> Begin with the closed unit interval $F_0 = [0, 1]$. From F_0 remove the open middle-third interval $\left(\frac{1}{3}, \frac{2}{3}\right)$ so that we are left with $F_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{1}{3}, 1\right]$. To get F_2, remove from each of these intervals the open middle-thirds. That is,
>
> $$F_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$
>
> Continue this process, removing the open middle-thirds from each component of F_n, to produce F_{n+1}. The limit of this sequence of intervals F_∞ is called the *Cantor set*.

Consider first that the total lengths of the intervals in F_n is $\left(\frac{2}{3}\right)^{n-1}$. That is to say that these lengths tend to zero as $n \to \infty$. Why this is such a peculiar observation has to do with the fact that this F_∞ in uncountable - it can't be written as a sequence! Additionally, F_∞ has empty interior so that $F_\infty = \partial F_\infty$; and because $F_\infty = \bigcap_{k=0}^{\infty} F_k$ is the intersection of closed, bounded, and nonempty subsets of $\mathbb{R}$, we can guarantee that it, too, is closed, bounded, and nonempty. This claim is proven in a more general setting below.

Theorem 5.3. Let (F_n) be a sequence of closed, bounded, nonempty subsets of $\mathbb{R}^k$ such that $F_1 \supseteq F_2 \supseteq \cdots$. Then the intersection $F = \bigcap_{n=1}^{\infty} F_n$ is also closed, bounded, and nonempty.

Proof. We know that F is bounded since F_1 is bounded and the intersection only has the potential to make it smaller. We also know that F is closed since arbitrary intersections of closed sets are still closed. The only tricky part here is to show that the intersection is nonempty.

For each n, choose an element $\vec{x}_n \in F_n$. Then $(\vec{x}_n)$ constitutes a bounded sequence in $\mathbb{R}^k$. The Bolzano-Weierstrass Theorem proven in the previous section shows there exists a subsequence $(\vec{x}_{n_m})$ that converges to some element $\vec{x}_0 \in \mathbb{R}^k$. We claim that this $\vec{x}_0$ lies in F.

To see this, consider that if we fix any $N \in \mathbb{N}$, then $m \geq N$ implies that $n_m \geq N$ so that $\vec{x}_{n_m} \in F_{n_m} \subseteq F_N$. Then our subsequence can be seen as a subsequence of F_N for each N. By the theorem above, $\vec{x}_0$ is a limit point for each F_N so that $\vec{x}_0 \in F_N$ for each $N \in \mathbb{N}$. That is, $\vec{x}_0$ belongs to F so that $F \neq \emptyset$. $\square$

5.3 Compact Sets

As we move on to more and more abstract spaces, we still want to be carrying along with us the things that have worked so well - things like completeness and the Bolzano-Weierstrass Theorem. So far, we've shown that $\mathbb{R}^n$ has such properties, but we would like to know if other spaces share these kinds of properties or at least some semblance of them. To so this, we require the notion of compactness.

Definition 5.6. Let (S,d) be a metric space. A collection $\mathcal{U}$ of open sets is said to be an *open cover* for $E \subseteq S$ if $E \subseteq \bigcup\{U \mid U \in \mathcal{U}\}$. A subset $\mathcal{V} \subseteq \mathcal{U}$ is said to be an *open subcover* of E if $E \subseteq \bigcup\{V \mid V \in \mathcal{V}\}$.

The set E is called *compact* is every open cover of E has a finite subcover.

At first, one looks at this definition and wonders how this is helpful... Like... At all! This is again the power of re-characterization. This is the property that extends to more general settings, but we need to do some work to show that it's the right thing even in the settings that we already know and love. Our saving grace in this endeavor is in the Heine-Borel Theorem that we prove shortly. In order to prove this, we need to know that our usual closed hypercubes in $\mathbb{R}^k$ are indeed compact (this way we have a point of comparison).

Proposition 5.4. Every k-cell $F \subseteq \mathbb{R}^k$ is compact.

Proof. Recall that a k-cell F takes the form

$$F = \{\vec{x} \in \mathbb{R}^k \mid x_j \in [a_j, b_j] \text{ for } j = 1, \ldots, k\}$$

for some intervals $[a_j, b_j]$ with $j = 1, \ldots, k$. Using the Pythagorean theorem, we say that

$$\delta = \left[\sum_{j=1}^{k}(b_j - a_j)^2\right]^{1/2}$$

is the *diameter* of F, denoted $diam(F)$.

Supposing to the contrary that F is not compact, there must exist an open cover $\mathcal{U}$ of F such that no finite subset of $\mathcal{U}$ covers F. We can realize F as the union of 2^k smaller k-cells with diameter $\frac{\delta}{2}$.

At least of of these smaller k-cells must be covered by infinitely many sets from $\mathcal{U}$. Call this cell F_1. We can now realize F_1 as a union of 2^k k-cells of diameter $\frac{\delta}{4}$, one of which must be covered by infinitely many sets from $\mathcal{U}$. Call this one F_2.

We can continue this process to obtain a nested sequence (F_n) of k-cells. That is, $F_n \subseteq F_{n+1}$ for all $n \in \mathbb{N}$. Moreover, the diameter of these cells $diam(F_n) = \frac{\delta}{2^n} \to 0$, and none of the F_n may be covered by finitely many sets from $\mathcal{U}$.

By the last theorem of the previous section, the intersection $\mathcal{F} = \bigcap_{k=1}^{\infty} F_n \neq \emptyset$ so that there is some $\vec{x}_0 \in \mathcal{F}$. Because $\mathcal{U}$ is a cover for F, this $\vec{x}_0$ resides in some $U_0 \in \mathcal{U}$. Moreover, $\vec{x}_0$ is interior to U_0 so that we have a ball $B_r(\vec{x}_0) \subseteq U_0$. For all $n \geq N$ such that $\frac{\delta}{2^N} < r$, F_n **is then covered by a single element of** $\mathcal{U}$. This contradiction of the construction of (F_n) shows that F could not have been compact to begin with, and our proposition is proven. $\square$

We are now equipped to prove the famous Heine-Borel Theoerm.

Theorem 5.4. (Heine-Borel) A subset $E \subseteq \mathbb{R}^k$ is compact if and only if it is closed and bounded.

Proof. First suppose E is compact. Consider the countable collection $\mathcal{U} = \{U_m\}_{m=1}^{\infty}$ of open k-cells

$$U_m = \{\vec{x} \in \mathbb{R}^k \mid \max_j |x_j| < m\}.$$

Because $\mathcal{U}$ covers $\mathbb{R}^k$, it also constitutes and open cover for E. The compactness of E allows us to find a finite subcover so that there is a maximal U_{m_0}. That is, there exists an m_0 such that $E \subseteq U_{m_0}$. Therefore, E is bounded by m_0.

To show E is closed, choose any point $\vec{x}_0 \in \mathbb{R}^k \backslash E$. Define the collection $\mathcal{V} = \{V_m\}_{m=1}^{\infty}$ where

$$V_m = \left\{\vec{x} \in \mathbb{R}^l \mid d(\vec{x}, \vec{x}_0) > \frac{1}{m}\right\}.$$

Hence, $\mathcal{V}$ is an open cover for E seeing how $\bigcup_{m=1}^{\infty} V_m = \mathbb{R}^k \backslash \{\vec{x}_0\} \supseteq E$. We appeal to the compactness of E again to find some m_0 such that $E \subseteq V_{m_0}$. This shows that $\vec{x}_0$ is interior to the set

$$\left\{\vec{x}_0 \in \mathbb{R}^k \mid d(\vec{x}, \vec{x}_0) < \frac{1}{m_0}\right\}.$$

This shows that each point of $\mathbb{R}^k \backslash E$ is an interior point so that $\mathbb{R}^k \backslash E$ is closed. Hence, E must be closed.

Now suppose that E is closed and bounded. Observe that the boundedness of E shows that there exists some k-cell F such that $E \subseteq F$. Given some open cover $\mathcal{U}$ of E, choose an open set F' that entirely contains F (a larger open k-cell would suffice) and to $\mathcal{U}$ the set $F' \backslash E$ to obtain $\mathcal{U}' = \mathcal{U} \cup \{F' \backslash E\}$. Then $\mathcal{U}'$ constitutes an open cover for both E and F. Because we showed above that F is compact, there must exist a finite subcover $\mathcal{V} \subset \mathcal{U}'$ of F that necessarily contains $F' \backslash E$. By removing the set $F' \backslash E$ from $\mathcal{V}$, we obtain a finite subcover of E so that E must indeed be compact. □

Example 5.5.

1. This theorem has completely classified the open subsets of $\mathbb{R}^k$. In particular, we know one special subclass of compact subsets of $\mathbb{R}^k$. Because k-cells are compact, they are both closed and bounded. Hence, any finite union of them must also be closed and bounded so that this union is also compact. In particular, for $\mathbb{R}$, **finite unions of closed intervals are compact**.

2. These unions of k-cells are not the only examples of compact subsets of $\mathbb{R}^k$. We found a rather strange example in the previous section: The Cantor set! Because this is a subset of $[0, 1]$ as we've defined it, the Cantor set is bounded. Additionally, we showed that it must be a closed set. Therefore, **the Cantor set is compact**.

5.4 Separation and Connected Sets

This is a short section, but it is a relevant notion that will be used in later courses. To motivate it, we need to dispel what our intuition naturally thinks must be separated.

Example 5.6. Consider the two subsets

$$A = (-1, 0) \cup (0, 1) \subset \mathbb{R}$$

and

$$B = \{(0,1)\} \cup \{(x,0) \mid x \in (0,1]\} \cup \left(\bigcup_{k=1}^{\infty} \left\{ \left(\frac{1}{k}, y \right) \mid y \in [0,1] \right\} \right)$$

of $\mathbb{R}$ (B is called a *comb space*). We would like to know if these two sets are what we would like to call *separated.* The intuitive notion of this idea is that there is some "space" between some pair of subsets that allows us to partition the set into two pieces. These two pieces need to not "touch" one another.

Naturally, the first place one thinks to look is the intersections of pairs of subsets of A and B. Are there any that are disjoint? For A this answer is obvious using the construction of the set. For B this is also obvious: Choose the "comb" B_1 and the singleton set $B_2 = \{(0,1)\}$. Now is where our intuition begins to break down.

If we're basing our notion of separation on whether or not there exists a two-set partition of the set, then **every set** (of more than one element) would be such a separation... But that surely can't be the whole story.

Consider that for A, the subsets $(-1,0)$ and $(0,1)$ not only partition the set - there is some "space" inbetween them. To describe this more fully, consider the closure $(0,1)^- = [-1,0]$. Not only is $(-1,0)$ disjoint from $(0,1)$, its closure $[-1,0]$ is disjoint from $(0,1)$ as well. This is the idea of separation that we'll use.

The candidate separation $B = B_1 \cup B_2$ proposed above does not satisfy this condition on the closures. Observe that while $B_1 \cap B_2^- = \emptyset$, the intersection $B_1^- \cap B_2 = B_2 \neq \emptyset$. Hence, there is no extra "space" between the two sets.

It turns out that there is no way to separate B. When we can't separate a set, we are going to say that such a set is *connected.* The following definition will rigorously define everything we've just outlined.

Definition 5.7. Let (S,d) be a metric space with $A, B \subseteq S$. We say that A and B are *separated* if both $A^- \cap B = A \cap B^- = \emptyset$.

A set $E \subset S$ is *disconnected* if it can be written as the union of two nonempty, separated sets A and B. The partition $\{A, B\}$ is called a *separation of* $E = A \cup B$.

When the set $E \subseteq S$ is not disconnected, we say that E is *connected.*

Remark 5.2. It seems strange that we are defining "connection" via its opposite. We are so accustomed to connection in our physical experience that we think it the easier thing to define. However, it turns out that we can describe what it means to be disconnected much more easily that we can describe what it means to be connected, hence why our definition is given as it is.

In light of this remark, we wonder if there is a way to define connectedness without using a negation. It turns out there is an alternative characterization of connectedness that does exactly this.

Theorem 5.5. Let (S,d) be a metric space with $E \subseteq S$. Then E is connected if and only if, for all $A, B \neq \emptyset$ where $E = A \cup B$, there always exists a convergent sequence $(x_n) \subset A$ or $(y_n) \subseteq B$ such that $x_n \to x \in B$ or $y_n \to y \in A$.

Proof. Suppose that E is disconnected. Then there exist nonempty sets $A, B \subseteq E$ with $E = A \cup B$ such that $A \cap B^- = A^- \cap B =$. Without loss of generality, suppose that $(x_n) \subset A$ converges to $x \in S$. We showed before that $x \in A^-$, but this means that $x \notin B$ as this is disjoint from A^-. Hence, no sequence in A can converge to any element of B. This also shows no sequence of B can converge to any element of A.

Now suppose that there exists a pair of nonempty sets A and B such that $E = A \cup B$ but instead that no convergent sequence of A converges to a limit in B and no convergent subsequence of B converges to a limit in A. That is to say exactly that A^- and B share no elements in common and neither do A and B^-. Therefore, $\{A, B\}$ is a separation of E, thus making E disconnected. $\square$

5.5 Topological Spaces

We now come to the point of everything we've done so far. We've moved from $\mathbb{R}$ to general metric spaces, and now we are now going to increase the level of abstract in order to increase the generality of the concepts we can discuss. We are moving away from the notion of distance. In order to retain any semblance of what we've studied so far, we are going to have to replace this by something just as strong if not stronger. But what?!

Consider the concepts that we have from metric spaces: Convergence, separation, and (perhaps most importantly) compactness. All of these concepts have one thing in common: Open sets. If we ditch distance as something that we can impart to a set, we simply cannot remove the notion of openness. So for a set X, what if we just dictate which sets are "open"? We can't just arbitrarily choose these sets from $\mathcal{P}(X)$, we need to guarantee that they satisfy those closure properties under arbitrary unions and finite intersections. We also need to outright guarantee that both $\emptyset$ and X itself are open sets. Let's wrap this up into one neat package - a definition!

Definition 5.8. Let X be a set. A subset $\mathcal{T} \subset \mathcal{P}(X)$ is called a *topology* for X if $\mathcal{T}$ satisfied the following three properties:

1. $\emptyset, X \in \mathcal{T}$,
2. Any union of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.
3. Any finite intersection of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

The pair $(X, \mathcal{T})$ is called a *topological space*. The elements of $\mathcal{T}$ are called *open sets*. The complement of an element of $\mathcal{T}$ is called a *closed set* of $(X, \mathcal{T})$.

Example 5.7. We've already seen a few examples of topologies for $\mathbb{R}^n$. They are all part of the class of *metric topologies* for $\mathbb{R}^n$, among which are the following:

(a) Using the standard Euclidean metric on $\mathbb{R}^n$

$$d_2((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \left(\sum_{j=1}^{n} (y_j - x_j)^2 \right)^{1/2},$$

we define the open sets to be the open balls

$$B_r(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^n \mid d_2(\vec{x}, \vec{x}_0) < r\}.$$

These are exactly our usual notions of solid spheres.

(b) Recall that the ℓ^∞ *metric* on $\mathbb{R}^n$ is defined by

$$d_\infty((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \max_j |y_j - x_j|.$$

The open sets are yet again given by the open balls

$$B_r(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^n \mid d_\infty(\vec{x}, \vec{x}_0) < r.$$

The closures of these balls are exactly the n-cells that we spoke of two sections ago.

(c) Recall that the ℓ^1 *metric* on $\mathbb{R}^n$ is defined by

$$d_1((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sum_{j=1}^{n} |y_j - x_j|.$$

The open set are again the open balls

$$B_r(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^n \mid d_1(\vec{x}, \vec{x}_0) < r\}.$$

These are the open diamonds discussed when we talked about this metric before.

Sets whose topology arises from a metric topology are called *metrizable spaces.* To expand on this idea, there are topologies of $\mathbb{R}$ that are not the metric topology for any choice of metric on $\mathbb{R}$. These are the *nonmetrizable* spaces. The question of metrizability of a space is rather difficult in general, so we state (without proof) one example of a nonmetrizable space:

> The *Zariski topology* $\mathcal{Z}$ on $\mathbb{R}$ contains the subsets of $\mathbb{R}$ that are the $\mathbb{R}$-complement of the set of zeros of some real polynomial. This is a nonmetrizable topological space.

It is, however, possible for the reader to verify that this is indeed a topology

Example 5.8. As with any abstract idea, there are some "stupid" examples that may or may not be terribly enlightening. In the context of a topology, there are some obvious small and large examples that satisfy the definition.

For any nonempty set X, we can use the bare minimum number of sets from $\mathcal{P}(X)$ that is allowed - two! We can use the set $\{\emptyset, X\}$ to create the smallest possible topology for X - the *indiscrete topology.* That is, **the only open subsets of X are itself and $\emptyset$.** This isn't terribly useful, but it is still an important example to consider in general.

For the same nonempty set X, we can use the maximum number of sets from $\mathcal{P}(X)$ that is allowed by using $\mathcal{P}(X)$ itself as the topology. This topology of X is called the *discrete topology* for X. That is, **every subset of X is open in X.**

Because of the notion of an interior point, we had to try to find a way to fit open balls inside certain subsets of metric spaces. We would like to abstract this concept for topological spaces.

Definition 5.9. For a topological space X, a subset $S \subset X$ is called a *neighborhood* of a point $x \in X$ if there exists an open set U such that $x \in U \subseteq S$. A point $x \in X$ is said to be *interior* to S if S is a neighborhood of x. The set of all points interior to S is called the *interior* of S and is denoted $int(S)$.

It is apparent that $int(S) \subseteq S$ as the open set U containing the point in question sits entirely within S. We would then like to know if this topological notion of "interiorness" encapsulates what it meant for metric spaces. Indeed, it does!

Proposition 5.5. Let S be a subset of a topological space X.

(a) S is open if and only if $S = int(S)$.

(b) $int(S)$ is open.

Proof. (a) Suppose S is open, then we can just take $U = S$ in the definition above for each point $x \in S$. Hence, each $x \in S$ belongs to $int(S)$ as well.

Suppose now that $S = int(S)$. For each $x \in S$, there exists an open set $U_x \subseteq S$ that contains x. Consider the union of all such sets over all points of S. Since each $x \in S$ belongs to one of these sets, we see that $\cup_{x \in S} U_x = S$. Therefore, seeing as any union of open sets is still open, S must be open.

(b) In light of part (a), we show that $int(S) \subseteq int(int(S))$. Let $x \in int(S)$. Then there is an open set $U_x \subseteq S$ containing x. For each $x' \in U_x$, the set U_x may be used as the set U in the definition of an interior point so that each $x' \in U_x$ is interior to $int(S)$. Hence, $x \in int(int(S))$ so that $int(S)$ is open.

□

Now that we have the notion of an open set, it is only natural to ask if the "closed" sets do what we expect them to do. And, again, we are not disappointed - but first we need a quick definition.

Definition 5.10. A point $x \in X$ is *adherent* to $S \subseteq X$ if any neighborhood U of x meets S so that $S \cap U \neq \emptyset$. The *closure* S^- of S is the set of points adherent to S.

Again, it should be clear that $S \subseteq S^-$ since any point of S will intersect any neighborhood of x in S.

Proposition 5.6. Let S be a subset of a topological space X.

(a) S is closed if and only if $S = S^-$.

(b) S^- is closed.

Proof. (a) Suppose that S is closed. Then $X \backslash S$ is open and cannot intersect S. Hence, no point of $X \backslash S$ is adherent to S. Therefore, $S^- \subseteq S$ so that $S = S^-$.

Now suppose that $S = S^-$. Let $x \in X \backslash S$. Then $x \notin S^-$ either so that, for any such x, there exists some neighborhood U_x of x such that $U_x \cap S = \emptyset$. Then the union of all such U_x must coincide with $X \backslash S$. This union of open sets is open so that $X \backslash S$ is open. Hence, S is closed.

(b) Let $x \in X \backslash S^-$. Then there is an open neighborhood U_x of x that cannot meet S since $X \backslash S^- \subseteq X \backslash S$. Since U_x is a neighborhood of each of its points, no point of U_x can be adherent to S, so that $U_x \subseteq X \backslash S^-$. Finding such U_x for each $x \in X \backslash S^-$ allows us to cover $X \backslash S^-$ with such neighborhoods. Then $X \backslash S^-$ is open so that S^- is closed.

□

Now that we know that this new definition of closure agrees with the one from metric spaces, we can start to do some analysis!

Definition 5.11. A sequence (x_n) in a topological space X *converges to* $x \in X$ if, for every open neighborhood U of x, there is an $N \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq N$.

We now see the function of the open set: **Open sets act are our ϵ bounds for convergence**.

Proposition 5.7. If S is a subset of a topological space X and if $(x_n) \subseteq S$ converges to $x \in X$, then $x \in S^-$.

Proof. Because S^- is closed, $X \backslash S^-$ is open. Since (x_n) does not enter $X \backslash S \subseteq X \backslash S^-$, the sequence cannot converge in $X \backslash S^-$. $\square$

This demonstrates just one way that abstracting our ideas from metric spaces to a more general setting improves our understanding: **The theory is easier to explain**!