Analysis

Joseph R. Mileti

August 24, 2023

Contents

1 Introduction **5**
- 1.1 Historical Background . . . 5
- 1.2 Exploring the Real Numbers . . . 8
- 1.3 Ordered Fields and the Real Numbers . . . 12
- 1.4 Supremums and Infimums . . . 19
- 1.5 Countability and Uncountability . . . 27
- 1.6 The Absolute Value Function . . . 32

2 Sequences of Real Numbers **37**
- 2.1 Definitions and Basic Results . . . 38
- 2.2 Algebraic and Order-Theoretic Properties of Limits . . . 43
- 2.3 Infinite Limits . . . 52
- 2.4 Subsequences and Cauchy Sequences . . . 60

3 Series of Real Numbers **65**
- 3.1 Definitions and Basic Results . . . 65
- 3.2 Series of Nonnegative Terms . . . 74
- 3.3 Absolute Convergence . . . 79

4 Topology of the Real Line **91**
- 4.1 Interiors and Closures . . . 91
- 4.2 Open and Closed Sets . . . 95
- 4.3 Compact Sets . . . 99

5 Functions, Limits, and Continuity **105**
- 5.1 Limits . . . 105
- 5.2 Continuity . . . 110

6 Differentiation and Integration **119**
- 6.1 Differentiation . . . 119
- 6.2 The Riemann Integral . . . 128

7 Sequences and Series of Functions **143**
- 7.1 Pointwise Convergence . . . 143
- 7.2 Uniform Convergence . . . 147
- 7.3 Series of Functions . . . 156
- 7.4 Power Series . . . 159
- 7.5 Representing Functions as Power Series . . . 165

Chapter 1

Introduction

1.1 Historical Background

Mathematics is often considered the pinnacle of reasoning. Many people think that mathematical arguments are airtight. To hold this position, a mathematical proof should be different than a scientific argument. It should not be enough to have a technique that has worked in every specific instance you have encountered and then use this to inductively generalize to a universal truth, or to have a heuristic argument suggesting truth. Also, mathematics should not depend on our current scientific theories or philosophical inclinations.

The history of mathematics demonstrates that mathematicians have historically failed at this goal, at least up through much of the 19th century. In particular, the track record of calculus in this regard was dismal. When calculus was originally developed in the 17th century, mathematicians (including Newton and Leibniz) used some very shady arguments. Newton's conception of calculus depended fundamentally on what he called *fluxions*. In our modern understanding, these would correspond to the derivative of the variables with respect to time, but this was not carefully laid out and Newton took time and motion as primitive concepts. But then these fluxions were used to understand (and essentially define) motion, which caused some philosophical circularity.

Both Newton and Leibniz made use of infinitely small quantities which Leibniz called *differentials*. These objects where treated at one turn as if they were distinct from zero (so you could divide by them) and at another turn the same as zero (so that a differential times a number can be treated as if "equal" to zero). Hence, when finding the derivative of a function f at a point x, they would use a differential dx, compute

$$\frac{f(x+dx)-f(x)}{dx},$$

and then throw away any terms which still had a factor of dx. Some mathematicians tried to justify this by saying that a differential was a new kind of number that was strictly greater than zero, but an order of magnitude less than every positive real number. You might have utilized them in your understanding of calculus so far, perhaps by thinking of the area under a curve as being comprised of infinitely many infinitely thin rectangles of length dx. To complicate matters, they even used higher order differentials that were supposedly orders of magnitudes smaller than differentials.

This kind of double-think and reckless use of the infinite may start to sound a little mystical. After all, the above fuzzy idea of the integral suggests that we can "add up" infinitely many infinitely small numbers to produce at a standard real number. At the time, few mathematicians gave much thought to these philosophical concerns, since the system (usually) worked and provided a handy computational toolkit to solve many physical problems. The most memorable charge against these techniques was given by George Berkeley (at the time about to become an Anglican bishop) in 1734 in an essay addressed to Edmund Halley

(of Halley's comet) entitled *The Analyst, or A Discourse Addressed to an Infidel Mathematician*. Here are a couple of quotes:

> It must, indeed, be acknowledged that [Newton] used fluxions, like the scaffold of a building, as things to be laid aside or got rid of as soon as finite lines were found proportional to them. But then these finite exponents are found by the help of fluxions. Whatever therefore is got by such exponents and proportions is to be ascribed to fluxions: which must therefore be previously understood. And what are these fluxions? The velocities of evanescent increments? And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?
>
> He who can digest a second or third fluxion, a second or third difference, need not, methinks, be squeamish about any point in divinity.

It was not until the middle of the 19th century that mathematicians began to provide a rigorous account of the real numbers and calculus that can be recognized today. Their eventual success in this endeavor capped two centuries of investigations, and stands as one of the great achievements of mathematics. The underlying transformation of the fundamental definitions and results altered the subject to such a degree that it was rebranded as *analysis*. Nonetheless, you may ask yourself whether the effort of carefully working out and formalizing the details behind Calculus is worth the effort. After all, we know calculus works, as its use in the sciences (especially physics) demonstrates. Sure, it might be intellectually and historically interesting, but does it actually matter to anybody other than those ivory-tower pure mathematicians? Isn't it all just an excuse for mathematicians to keep their jobs?

Perhaps surprisingly, the historical push for a more careful and rigorous treatment of the concepts of Calculus came from a very applied problem in physics. Fourier tried to model heat flow across a long thin rectangular plate. If we have a certain temperature distribution $f(x)$ across one short edge, and we assume that the heat is constant across the two long edges, how does the heat propagate across the plate? When solving the underlying partial differential equations, Fourier found solutions involving a constant together with linear combinations of sines and cosines of different frequencies (that is, of $\sin x, \sin 2x, \sin 3x, \dots$ and $\cos x, \cos 2x, \cos 3x, \dots$) along the one short edge. As a result, if $f(x)$ can be expressed in terms of a linear combination of such sines and cosines, then Fourier had solved the problem. So which functions $f(x)$ can be expressed as such a trigonometric series? If we allow only *finite* linear combinations, then the class of such functions is limited. But what if we think about *infinite* linear combinations, i.e. expressions of the form

$$c + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx.$$

Although such an expression might look complicated, mathematicians had up to this point made extensive use of infinite series of the form

$$\sum_{n=0}^{\infty} a_n x^n,$$

and had developed a robust theory of which functions can be represented as such an "infinite polynomial". In particular, they understood (at least through intuition and heuristic arguments) that essentially all "reasonable" functions can be represented in this way. In fact, many results in the early days of Calculus made extensive use of such expressions. But these trigonometric series seemed more exotic. Although they had arisen several times in 18th century work on vibrating strings and other waves, many mathematicians believed that few natural functions could be expressed in such a way. But Fourier, in his study of heat flow, came to believe the opposite, i.e. that every "reasonable" function can be so expressed. However, it took decades for mathematicians to both carefully define and to understand what it means to say that the above trigonometric series "equals" some given function. And by working out these details, mathematicians

were forced to grapple with concepts of *limits* and *convergence* that could not be handled with the same nonchalance that characterized earlier work.

The study of these series laid the groundwork for a field that is now known as *Fourier analysis*, and eventually the wider field of *harmonic analysis*. Although the origins of these investigations came from problems involving waves and heat flow, they have become essential components of many applied problems. If you visualize the above sine and cosine functions as waves of different frequencies, then you can imagine trying to break up a general function into these underlying constituent waves, just like we break up a vector as a linear combination of a given basis (in fact, the constant function 1 and the above sine and cosine function form a linearly independent sequence in the vector space of functions). For example, if we have a complicated sound wave produced by some instruments, can we determine the underlying simple waves that make it up? The discrete version of this problem is an essential part of creating MPEG and JPEG files. The idea is to take a complicated source (whether an audio file or a digital photo), break it up into a combination of sines and cosines, and then omit the "less important" high frequency terms from the sum in order to save space without losing noticeable quality. The essential algorithm behind these computations is the *fast Fourier transform*, which is one of the most important and widely used algorithms of our time.

Although Fourier's question about representing general functions as infinite trigonometric series provided the impetus for many mathematicians to study the foundations of Calculus, the fruits of their labor ended up going far beyond such series. For example, in order to apply the tools of calculus to modern science, it has become necessary to understand how to do calculus on spaces more complicated than $\mathbb{R}$. Of course, there's $\mathbb{R}^n$, but there's also the complex numbers $\mathbb{C}$ and curved spaces (like our universe as described by general relativity) made precise through the notion of a manifold. Rather than starting from scratch each time, a deeper understanding of calculus provides the first step toward unifying these apparently separate domains.

More abstractly, it is possible to work in a "space" consisting of "points" which are objects other than numbers or finite tuples of numbers. To illustrate, consider the following fundamental problem in the history of physics, known as the *brachistochrone problem*. Given two points in space, what is the shape of the curve connecting the two points which minimizes the time it takes a ball to roll along under the influence of gravity? This looks like a max/min problem, but instead of minimizing the value of a function which takes a number to a number, we are minimizing the value of a higher-order function which takes a function (the curve) to a number (the time it takes the ball to roll down the curve). The proper perspective to take in order to tackle this type of problem is in the general setting provided by *functional analysis*, where one meet Banach spaces and Hilbert spaces, places where we can do "calculus" and "geometry" much more generally. The added generality provides a lot more than just insight and unification. For example, the modern formalization of quantum mechanics takes the states of your quantum system to be "points" in a Hilbert space.

Although these examples demonstrate that any applied mathematician should develop an appreciation for modern analysis, the same can be said of any pure mathematician. Of course, the subject is beautiful and intellectually satisfying on its own, but is also incredibly useful to other areas of mathematics. For instance, many parts of modern number theory make use of harmonic analysis, including recent results about the distribution of the primes. Perhaps even more surprisingly, Cantor's work on different sizes of infinity came out of his study of infinite trigonometric series related to Fourier's question!

With the background in mind, it is time to begin our study of analysis. Rather than follow the historical twists and turns, we will take a more modern approach that eases some of the underlying complexity in the early stages. As a result, some of the questions that we ask and the definitions and theorems we develop may appear to be motivated only from a "pure" mathematics perspective. Unfortunately, we will not have the time to tie everything back together to the historical and applied in one semester. To see those topics developed more fully, continue taking mathematics courses!

1.2 Exploring the Real Numbers

Throughout most mathematics courses in high school and early college, such as Calculus and Linear Algebra, the real numbers play a central role. However, it is often not clear what makes the real numbers so special. You may have heard that the real numbers correspond to the points on the "number line", but at this point, you may have little reason to believe that assertion. It is very reasonable to argue that the concept of a number line is vague, grounded in our perception of reality and hence suspect, or possibly even incompatible with reality (if you subscribe to a belief that space, like energy, is fundamentally discrete). In fact, the real numbers are surprisingly subtle, and it will take some effort to describe them precisely. Most people believe that a real number is given by a possibly infinite decimal expansion. As we will see, this is not a great way to define the real numbers, and it is only one possible way to represent the real numbers.

For a toy example of the underlying issues, you might have seen arguments about whether $0.99999\ldots$ represents the same number as $1.00000\ldots$. If we *define* the real numbers as such decimal expansions, we have to define addition, multiplication, and equality in terms of such expansions, so we have to make such choices at the beginning. Now there are very good heuristic arguments (some of which you may have seen) for why we should adopt the equality of these two decimals, but working from such a perspective has problems. First, we have to think about the equality of other decimal expansions. More interestingly, we have to define multiplication of decimals from scratch. It is possible (try it!) to do this properly, but it is not for the faint of heart. Moreover, even if we come up with a proper definition, we would have to check that multiplication is well-defined, i.e. if we change the representation of the factors to equivalent decimals, we need to know that result of our multiplication rule provides the same final answer.

We want to avoid all of these issues. We also want to approach the real numbers from a more refined perspective that is not tied to specific representations. We will carry out this approach in the next section by writing down some axioms for the real numbers, and then we will spend the rest of our time working from these simple assumptions. The axioms will state that we are able to add and multiply real numbers, and that those operations should satisfy the usual arithmetic rules. We also have an ordering $<$ that plays nicely with addition and multiplication. But before diving into the specifics, we should pause to think about what else makes the real numbers special. In the rest of this section, we will explore these ideas from a critical but informal perspective. Since we have not yet laid out the axioms that we will follow, we will instead argue using basic properties of $\mathbb{Q}$ and $\mathbb{R}$ that you certainly believe (all of which will eventually follow from the axioms).

Why do the real numbers play such an important role in Calculus? After all, we can add, multiply, and order the rational numbers $\mathbb{Q}$. And in many ways, $\mathbb{Q}$ feels like it has "enough" numbers. For example, there is no smallest positive element of $\mathbb{Q}$, because if $0 < a$, then $0 < \frac{a}{2} < a$. In fact, between any two elements of $\mathbb{Q}$, there always exists another because if $a, b \in \mathbb{Q}$ and $a < b$, then $\frac{a+b}{2} \in \mathbb{Q}$ and $a < \frac{a+b}{2} < b$. Moreover, by repeatedly using this fact, we conclude that if $a < b$, then the set $\{q \in \mathbb{Q} : a < q < b\}$ is infinite.

So what is it that $\mathbb{Q}$ lacks? In Linear Algebra, you should have seen a proof that there is no $q \in \mathbb{Q}$ with $q^2 = 2$. But why should we care about having a number whose square is 2? On the face of it, there is no a priori reason why such a "number" must exist. After all, the real numbers do not contain an element r with $r^2 = -1$, and you probably do not consider this to be a defect of the real numbers. Historically, the impetus for the existence of a number on the number line whose square is 2 came from the ancient Greeks and the Pythagorean Theorem. If one considers an isosceles right triangle with the two equal sides of length 1, then the hypotenuse should have length c where $c^2 = 1^2 + 1^2 = 2$. Thus, if we want to be able to assign a number representing the length of each geometrically describable line segment, then we must have a number c with $c^2 = 2$.

The ancient Greeks also spent a lot of time thinking about circles. If one draws a circle with radius 1, then what is the area enclosed by the circle? Of course, we denote the area by π, and it turns out that π can not be represented by any rational number. In modern language, we say that π is irrational, but note that this was not proven until the middle of the 18th century. Similarly, the length of the arc enclosing the

unit circle, i.e. the circumference, is 2π, which is also irrational. Thus, by exploring geometry, the ancient Greeks naturally stumbled into lengths and areas that are geometrically describable in terms of the integers, but could not be expressed by elements of $\mathbb{Q}$.

Of course, one way around these problems is to simply throw in new numbers like $\sqrt{2}$ and π. Is that enough? What about $\sqrt{3}$? Can we express $\sqrt{3}$ using just $\sqrt{2}$, π, and elements of $\mathbb{Q}$? These are certainly interesting questions, but I hope that they illustrate the need for a more fundamental and systematic approach to understanding what "holes" and "gaps" exist within the rational numbers.

Toward that end, let's return to an investigation of a number whose square is 2 from the perspective of $\mathbb{Q}$. As mentioned above, there does not exist $q \in \mathbb{Q}$ with $q^2 = 2$. Although this might seem to be the end of the story, there is more to say. We begin by defining the following two sets:

$$A = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 < 2\}$$
$$B = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 > 2\}$$

Since there does not exist $q \in \mathbb{Q}$ with $q^2 = 2$, we have $A \cup B = \{q \in \mathbb{Q} : q > 0\}$. Moreover, using the fact that if $0 < q < r$, then $q^2 < r^2$, it follows that every element of A is less than every element of B.

We claim that A does not have a largest element and that B does not have a smallest element. Let's think through how we would argue the former of these two claims. Let $q \in A$ be arbitrary. We need to show that there exists $r \in \mathbb{Q}$ with $r > q$ and $r^2 < 2$. If one approaches this problem directly by immediately trying to construct such an r, perhaps by breaking up q into a numerator and denominator and playing around with inequalities, things quickly become messy. We instead try to think indirectly and a bit backwards. Let's try to add a tiny rational amount to q so that the result of squaring will still give a value less than 2. Notice that $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ is a sequence of positive rational numbers that becomes arbitrarily small. So we try to argue that there exists an $n \in \mathbb{N}^+$ such that $(q + \frac{1}{n})^2 < 2$. For any $n \in \mathbb{N}^+$, we have

$$\left(q + \frac{1}{n}\right)^2 = q^2 + \frac{2q}{n} + \frac{1}{n^2}.$$

Now we are assuming that $q \in A$, so $q^2 < 2$, and hence $2 - q^2 > 0$. Thus, we want to argue that there exists an $n \in \mathbb{N}^+$ such that

$$\frac{2q}{n} + \frac{1}{n^2} < 2 - q^2.$$

Intuitively, such an n should exist, because the right-hand side is positive, and we can make the left-hand side as small as we would like by choosing n sufficiently large. But that argument is a bit vague. How large do we have to make n before we are sure that it works? If we try to solve the above inequality for n, we have to deal with the n^2 term, which results in an annoying quadratic. But we do not need to find *all* values of n, or even the smallest value of n, that works. We just need to find *some* appropriate value of n. Since we are just trying to find an n that satisfies an inequality, we can make the inequality itself stricter in order to make it easier to solve. In particular, we would love to get rid of the n^2 term. Now of course we can not simply drop it entirely, because

$$\frac{2q}{n} + \frac{1}{n^2} > \frac{2q}{n}$$

for any $n \in \mathbb{N}^+$, which is the wrong direction for the inequality. In other words, if we pick $n \in \mathbb{N}^+$ with $\frac{2q}{n} < 2 - q^2$, then there is no reason to believe that value of n will satisfy

$$\frac{2q}{n} + \frac{1}{n^2} < 2 - q^2.$$

However, notice that for any value of $n \in \mathbb{N}^+$, we certainly have $n \leq n^2$, so $\frac{1}{n^2} \leq \frac{1}{n}$, and hence

$$\frac{2q}{n} + \frac{1}{n^2} \leq \frac{2q}{n} + \frac{1}{n}.$$

Therefore, it suffices to find an $n \in \mathbb{N}^+$ with

$$\frac{2q}{n} + \frac{1}{n} < 2 - q^2,$$

which is the same as finding an $n \in \mathbb{N}^+$ with

$$\frac{2q+1}{n} < 2 - q^2.$$

From this last inequality, it is even more clear that the left-hand side becomes arbitrarily small as we increase n. But to be precise in finding such an n, we can simply solve the inequality by multiplying both sides by $\frac{n}{2-q^2} > 0$ to obtain

$$n > \frac{2q+1}{2-q^2}.$$

Since $\frac{2q+1}{2-q^2} \in \mathbb{Q}$, and we know that given any rational number, we can always find an element of $\mathbb{N}^+$ greater than it, we have figured out how to determine a value of n that will work.

An astute reader will realize that the logic of the above argument is going in the wrong direction. That is, we wanted to find an $n \in \mathbb{N}^+$ with $(q + \frac{1}{n})^2 < 2$, and we started with this goal and worked to derive what n must be. Now in this case it does turn out that each step is reversible, i.e. that any n larger than $\frac{2q+1}{2-q^2}$ will have the property that $(q + \frac{1}{n})^2 < 2$. However, we should really write the argument in the correct logical order, which we now proceed to do.

Proposition 1.2.1. *Let*

$$A = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 < 2\}$$
$$B = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 > 2\}.$$

The set A does not have a maximum element, and the set B does not have a minimum element.

Proof. We first prove that A does not have a maximum element. Let $q \in A$ be arbitrary, so $q^2 < 2$, and hence $2 - q^2 > 0$. Using the fact that for every $r \in \mathbb{Q}$, there exists $m \in \mathbb{N}^+$ with $m > r$, we can fix an $n \in \mathbb{N}^+$ with

$$n > \frac{2q+1}{2-q^2}.$$

Multiplying both sides of this inequality by $\frac{2-q^2}{n} > 0$, we conclude that

$$\frac{2q+1}{n} < 2 - q^2.$$

Now since $n \leq n^2$, we have $\frac{1}{n^2} \leq \frac{1}{n}$, and hence

$$\begin{aligned}
\left(q + \frac{1}{n}\right)^2 &= q^2 + \frac{2q}{n} + \frac{1}{n^2} \\
&\leq q^2 + \frac{2q}{n} + \frac{1}{n} \\
&= q^2 + \frac{2q+1}{n} \\
&< q^2 + (2 - q^2) && \text{(from above)} \\
&= 2.
\end{aligned}$$

Since $q + \frac{1}{n} \in \mathbb{Q}$ and $0 < q < q + \frac{1}{n}$, we have shown that there exists an element of A that is larger than q. Therefore, A does not have a maximum element.

We now show that B does not have a minimum element. Let $q \in B$ be arbitrary, so $q^2 > 2$, and hence $q^2 - 2 > 0$. Fix $n \in \mathbb{N}^+$ with

$$n > \frac{2q}{q^2 - 2}.$$

Multiplying both sides of this inequality by $\frac{q^2-2}{n} > 0$, we conclude that

$$q^2 - 2 > \frac{2q}{n}.$$

Multiplying by sides of this inequality by $-1 < 0$, it follows that

$$2 - q^2 < -\frac{2q}{n},$$

and hence

$$\begin{aligned}
\left(q - \frac{1}{n}\right)^2 &= q^2 - \frac{2q}{n} + \frac{1}{n^2} \\
&\geq q^2 - \frac{2q}{n} && \left(\text{since } \frac{1}{n^2} \geq 0\right) \\
&> q^2 + (2 - q^2) && \text{(from above)} \\
&= 2.
\end{aligned}$$

Thus, to conclude that $q - \frac{1}{n} \in B$, we need only show that $q - \frac{1}{n} > 0$. Notice that we must have $q > 1$, because if $0 < q \leq 1$, then $0 < q^2 \leq 1 < 2$, a contradiction. Now using the fact $\frac{1}{n} \leq 1$, it follows that $q - \frac{1}{n} > 0$, so $q - \frac{1}{n} \in B$, and hence B contains an element smaller than q. Therefore, B does not have a minimum element. □

Before returning to what this result demonstrates about the real numbers, let's pause to learn a few lessons from the above argument, each of which will be important in our study of Analysis.

- Although both of the arguments are straightforward to read and verify line by line, the value of n chosen at the beginning of each seems unmotivated when it first appears. Of course, the value of n in the first argument makes sense given the context that we worked through above. In contrast, I did not provide the same motivating context before diving into the proof that B does not have a smallest element. Now I certainly did work through all of the motivating calculations behind the scenes before writing up the second argument, but they are not logically necessary for the proof, and are often omitted. My position is that they *should* be omitted from the actual proof (which is simply a complete logical argument verifying the statement in question), and that you should do the same in your arguments. However, as a reader, and especially as a novice to the subject, it is helpful to see how somebody thinks through the intuition as well. I will regularly explain how I think through these ideas in the expository paragraphs and in class. However, know that you will often *not* see such motivation in many sources. And realize that you will often have to play around, experiment, and work backwards before you actually find, and eventually write up, such a proof.

- Some of the inequality work in the above arguments may be novel to you. Working with inequalities can be tricky and often requires a new perspective. But inequalities are the lifeblood of Analysis. Know that you can always work with a stricter inequality that might appear more difficult to satisfy, but is easier to work with in practice, in order to accomplish your goal. Figuring out when this is appropriate, and which terms to tweak, is a skill that can only be acquired through a great deal of practice and experimentation.

- You have almost certainly begun to appreciate the amount of creativity and insight that mathematics requires from previous courses. It is very natural to think "I never would have thought of that" when you first read a proof, and it takes time to develop the important insights and ways of thinking that arise in each subject. However, it is essential to spend time working to cultivate these creative instincts. When you encounter the "I never would have thought of that" feeling, read through the proof many times with pencil and paper, playing around with the ideas, in order to think through how somebody might have come up with it. Then try to write up your own motivation. You might not be successful in this endeavor for every proof this semester (I certainly was not when I was an undergraduate), but over time, you will find the arguments easier to understand and motivate. Eventually, with enough time and effort, you will develop expertise, and maybe even find the ideas completely natural. I know of no other way to arrive at this goal.

Let's return to discussing what Proposition 1.2.1 tells us about the structure of the real numbers. Recall that

$$A = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 < 2\}$$
$$B = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 > 2\}$$

and that $A \cup B = \{q \in \mathbb{Q} : q > 0\}$ because there is no $q \in \mathbb{Q}$ with $q^2 = 2$. Now there are many examples of subsets of $\mathbb{R}$ that do not have maximum or minimum elements. The entire set $\mathbb{R}$ has neither a maximum nor minimum, and even some bounded sets like the open interval $(0, 1) = \{r \in \mathbb{R} : 0 < r < 1\}$ also have this property. However, there is something fundamentally different between $(0, 1)$ and A. While $(0, 1)$ does not have a maximum element, the set $[1, \infty) = \{r \in \mathbb{R} : r \geq 1\}$ (i.e. the complement in the set of positive reals) does have a minimum element. Although the set $(0, 1)$ does not include the number 1, or any number greater than 1, it does contain numbers arbitrary close to 1. Thus, although 1 is not the maximum of $(0, 1)$, it plays a very special role in bounding $(0, 1)$ from above. More formally, the number 1 is an upper bound for $(0, 1)$, but no smaller number is an upper bound for $(0, 1)$, and this manifests itself in the fact that the set $[1, \infty)$ of upper bounds for $(0, 1)$ has a minimum element.

The sets A and B are similar to the sets $(0, 1)$ and $[1, \infty)$. In both cases, every element of the former set is strictly less than every element of the latter set, and in both cases, the union is all positive numbers (in the former case, this is all positive rationals, while in the latter, it is all positive reals). However, as shown in Proposition 1.2.1, the set A does not have a maximum and the set B does not have a minimum. Looking at the fact that $\{q \in \mathbb{Q} : q > 0\} = A \cup B$, we see a decomposition that demonstrates a "hole" in the set $\{q \in \mathbb{Q} : q > 0\}$ of positive rational numbers. It feels like there should be a number that serves as the dividing line between A and B, but no such number exists in $\mathbb{Q}$. More formally, the set A is bounded above and does not have a maximum element, but since B does not have a minimum, there is no element of $\mathbb{Q}$ that plays the same role that 1 does to the set $(0, 1)$. Of course, in $\mathbb{R}$, the number $\sqrt{2}$ plays the role of this dividing line between A and B, and does serve as the least upper bound of A. But the key insight is that we can detect a "hole" in $\mathbb{Q}$ without referring to $\mathbb{R}$ at any point.

In the next section, we will use the ideas laid out here to define $\mathbb{R}$ as a place where we can add, multiply, and order elements, but also where there are no "holes" like we see in $\mathbb{Q}$. The interesting and surprising fact is that we can characterize potential "holes" by only referring to the ordering and upper bounds of subsets as described in the above example.

1.3 Ordered Fields and the Real Numbers

We are now ready to write down that axioms of the real numbers, which naturally fall into three categories. We begin with axioms related to basic arithmetic, and then we turn to axioms about the ordering $<$ and how it interacts with the arithmetic. Finally, we will introduce the important Completeness Axiom, which we were building intuition for in the previous section. As we lay out each of these axiom groups, we will give

examples of different models of the axioms, i.e. we will talk about different contexts where they hold. Along the way, we will use $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ in the discussion around the axioms despite the fact that we have not yet completed our description of $\mathbb{R}$ (on which any definition of $\mathbb{C}$ depends). There is no logically circularity here, as these examples only serve an illustrative purposes to build intuition, and play no a prior role in the actual axioms themselves.

We begin with the arithmetical properties of the real numbers. We should be able to add, subtract, multiply, and divide two real numbers (in the last case, we of course require that the second input be nonzero), and these operations should satisfy some simple rules. However, we can define subtraction and division in terms of addition and multiplication, so we only describe those two fundamental operations in the definition. If you have taken Abstract Algebra, then you should be familiar with this collection of axioms. If you have not, then there is no need to worry, as they simply lay out many of the fundamental properties of addition and multiplication that you have used for years.

Definition 1.3.1. *A* field *is a set F equipped with two binary operations $+$ and $\cdot$ (i.e. both $+$ and $\cdot$ are functions from $F \times F$ to F) and two distinct elements $0, 1 \in F$ satisfying the following properties:*

1. *$a + (b + c) = (a + b) + c$ for all $a, b, c \in F$.*
2. *$a + b = b + a$ for all $a, b \in F$.*
3. *$a + 0 = a$ for all $a \in F$.*
4. *For all $a \in F$, there exists $b \in F$ with $a + b = 0$.*
5. *$a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in F$.*
6. *$a \cdot b = b \cdot a$ for all $a, b \in F$.*
7. *$a \cdot 1 = a$ for all $a \in F$.*
8. *For all $a \in F$ with $a \neq 0$, there exists $b \in F$ with $a \cdot b = 1$.*
9. *$a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in F$.*

Each of the sets $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields under the usual operations of addition and multiplication. However, there are many other fields. For example, we can make a field from the set $F = \{0, 1\}$ by defining addition and multiplication as follows:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

If you are familiar with modular arithmetic, then the above operations are simply addition and multiplication modulo 2. In fact, for each prime p, the set $\{0, 1, 2, \ldots, p-1\}$ with operations of addition and multiplication modulo p is a field with finitely many elements. The most surprising axiom that $\{0, 1, 2, \ldots, p-1\}$ satisfies (under the modular operations) is axiom (8), which states that every nonzero number has a multiplicative inverse. See Number Theory or Abstract Algebra for a proof.

There are also many fields "between" $\mathbb{Q}$ and $\mathbb{R}$, such as

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\},$$

with the usual addition and multiplication inherited from $\mathbb{R}$. Since we have simply defined a subset of $\mathbb{R}$ using the operations from $\mathbb{R}$, the only things that we need to check is that the above set is closed under addition, multiplication, and multiplicative inverses. It is straightforward to check that if one adds or multiplies two

elements of the above set, then the result is still in the set. For multiplicative inverses, notice that if at least one of a or b is nonzero (so that $a + b\sqrt{2} \neq 0$), then

$$\begin{aligned}\frac{1}{a+b\sqrt{2}} &= \frac{1}{a+b\sqrt{2}} \cdot \frac{a-b\sqrt{2}}{a-b\sqrt{2}} \\ &= \frac{a-b\sqrt{2}}{a^2-2b^2} \\ &= \frac{a}{a^2-2b^2} + \frac{-b}{a^2-2b^2} \cdot \sqrt{2}.\end{aligned}$$

Although we leave a detailed study of fields to Abstract Algebra, we state a couple of simple facts here. We start by noting that although axioms (4) and (8) simply state the existence of additive and multiplicative inverses, it can be shown using the other axioms that we also have uniqueness, as stated in the next result.

Proposition 1.3.2. *Let F be a field.*

1. *For all $a \in F$, there exists a unique $b \in F$ with $a + b = 0$.*
2. *For all $a \in F$ with $a \neq 0$, there exists a unique $b \in F$ with $a \cdot b = 1$.*

Since these inverses are unique, we can now use them to define subtraction and division in any field.

Notation 1.3.3. *Let F be a field.*

1. *Given $a \in F$, we denote the unique $b \in F$ with $a + b = 0$ by $(-a)$.*
2. *Given $a, b \in F$, we write $a - b$ as shorthand for $a + (-b)$.*
3. *Given $a \in F$ with $a \neq 0$, we denote the unique $b \in F$ with $a \cdot b = 1$ by a^{-1}.*
4. *Given $a, b \in F$ with $b \neq 0$, we write $\frac{a}{b}$ as shorthand for $a \cdot b^{-1}$.*

With all of the notation in hand, it is now possible to prove all of the basic laws of arithmetic. Here is a list of some important examples (see Abstract Algebra for the proofs, or just try them yourself!).

Proposition 1.3.4. *Let F be a field.*

1. *For all $a \in F$, we have $a \cdot 0 = 0$.*
2. *For all $a \in F$, we have $-(-a) = a$.*
3. *For all $a, b \in F$, we have $a \cdot (-b) = -(a \cdot b) = (-a) \cdot b$.*
4. *For all $a, b \in F$, we have $(-a) \cdot (-b) = a \cdot b$.*
5. *For all $a \in F$, we have $-a = (-1) \cdot a$.*
6. *For all $a, b, c \in F$, we have $a \cdot (b - c) = ab - ac$.*
7. *For all $a, b \in F$, if $ab = 0$, then either $a = 0$ or $b = 0$.*
8. *For all $a, b, c, d \in F$ with $b \neq 0$ and $d \neq 0$, we have both*

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} \qquad \text{and} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

The field axioms characterize the important algebraic properties of the real numbers. But beyond the fundamental operations of addition and multiplication, the real numbers are also equipped with a relation $<$. In addition to providing a linear ordering of the elements, this relation also plays nicely with addition and multiplication. Although there are several possible axiomatizations of the core fundamental facts from which the others can be derived, we adopt the following choice of axioms for an *ordered field*.

Definition 1.3.5. *An* ordered field *is a field F equipped with a binary relation $<$ with the following properties:*

1. *If $a < b$ and $b < c$, then $a < c$.*
2. *For all $a \in F$, we have $a \not< a$.*
3. *For all $a, b \in F$, either $a < b$, or $a = b$, or $b < a$.*
4. *If $a < b$ and $c \in F$, then $a + c < b + c$.*
5. *If $0 < a$ and $0 < b$, then $0 < ab$.*

The two most natural ordered fields are $\mathbb{Q}$ and $\mathbb{R}$, each with their usual addition, multiplication, and ordering. However, there are many others. For example, recall from above that

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

is a field that is a subset of $\mathbb{R}$. As such, we can restrict the ordering $<$ on $\mathbb{R}$ to the set $\mathbb{Q}(\sqrt{2})$ in order to make $\mathbb{Q}(\sqrt{2})$ into an ordered field. The same idea works for any subfield of $\mathbb{R}$, i.e. if F is a subfield of $\mathbb{R}$, then F can be viewed as an ordered field by simply restricting the ordering from $\mathbb{R}$.

Can some of the of the other fields that we mentioned above be turned into ordered fields? Consider $F = \{0, 1\}$, where addition and multiplication are modulo 2. The natural idea is to set $0 < 1$. Does this make F into an ordered field? If $0 < 1$, then we must have $1 < 1 + 1$ by the fourth axiom, which would imply $1 < 0$ (since $1 + 1 = 0$ in this field). However, from $0 < 1$ and $1 < 0$, we would be able to conclude that $0 < 0$ by the transitivity axiom (the first one), which would contradict the second axiom. A similar argument shows that $1 < 0$ also does not work. Therefore, it is impossible to turn F into an ordered field. It is also impossible to turn $\{0, 1, 2, \ldots, p-1\}$ under the modulo p operations, and $\mathbb{C}$ under the usual operations, into ordered fields, as we will see below.

We now establish some straightforward notation, as well as some simple properties of ordered fields. Although we do include the proofs of these properties, it is completely acceptable to omit a careful study of them at this point. Each of these properties underlie the fundamental algebraic manipulations of inequalities in high school algebra, so they should be familiar in the world of the real numbers (and hence in any subfield of the real numbers with the inherited ordering).

Notation 1.3.6. *Let F is an ordered field with ordering $<$. We define $a > b$ to mean that $b < a$. We also define a new binary relation $\leq$ by letting $a \leq b$ mean that either $a < b$ or $a = b$. Finally, we define $a \geq b$ to mean that $b \leq a$.*

Proposition 1.3.7. *Let F be an ordered field.*

1. *For all $a, b \in F$, exactly one of $a < b$, $a = b$, or $b < a$ is true.*
2. *For all $a, b \in F$, we have $a < b$ if and only if $0 < b - a$.*
3. *For all $a, b \in F$, we have $a < b$ if and only if $-b < -a$.*
4. *If $a < b$ and $c \leq d$, then $a + c < b + d$.*
5. *If $a < b$ and $0 < c$, then $ac < bc$.*

6. *If $a < b$ and $c < 0$, then $bc < ac$.*
7. *If $a < 0$ and $b > 0$, then $ab < 0$.*
8. *If $a < 0$ and $b < 0$, then $ab > 0$.*
9. *For all $a \in F$, we have $a^2 \geq 0$.*
10. $0 < 1 < 1 + 1 < 1 + 1 + 1 < 1 + 1 + 1 + 1 < \dots$.

Proof.

1. Let $a, b, c \in F$ be arbitrary. We know from the third axiom that at least one of the three is true. We can not have both $a < b$ and $a = b$, as this would violate the second axiom of ordered fields. Similarly, we can not have both $a = b$ and $b < a$. Finally, if we have both $a < b$ and $b < a$, then by transitivity of $<$ (the first axiom), we could conclude that $a < a$, contrary to second axiom.

2. Let $a, b \in F$ be arbitrary. If $a < b$, then we can add $-a$ to both sides to conclude that $a + (-a) < b + (-a)$, from which it follows that $0 < b - a$. Conversely, if $0 < b - a$, then $0 < b + (-a)$, so adding a to both sides we conclude that $a < b + (-a) + a$, from which it follows that $a < b$.

3. Let $a, b \in F$ be arbitrary. If $a < b$, then adding $(-a) + (-b)$ to both sides, it follows that $-b < -a$. Conversely, if $-b < -a$, then adding $a + b$ to both sides, it follows that $a < b$.

4. Suppose that $a < b$ and $c \leq d$. We now have two cases:

   - *Case 1:* Suppose that $c = d$. Since $a < b$, we can add c to both sides to conclude that $a + c < b + c$, and thus $a + c < b + d$.
   - *Case 2:* Suppose that $c < d$. Since $a < b$, we can add c to both sides to conclude that $a + c < b + c$. Since $c < d$, we can add b to both sides to conclude that $b + c < b + d$. Now using the fact that $<$ is transitive (the first axiom), it follows that $a + c < b + d$.

5. Suppose that $a < b$ and $0 < c$. Since $a < b$, we know from part (2) that $0 < b - a$. Using the fifth axiom of ordered fields, it follows that $0 < (b - a) \cdot c$. By part (5) of Proposition 1.3.4, we conclude that $0 < bc - ac$. Using part (2) again, it follows that $ac < bc$.

6. Suppose that $a < b$ and $c < 0$. Since $a < b$, we know from part (2) that $0 < b - a$. Using part (3), it follows that $-0 < -c$, and since $-0 = 0$ (because $0 + 0 = 0$), we conclude that $0 < -c$. By the fifth axiom of ordered fields, it follows that $0 < (b - a) \cdot (-c)$. Using Proposition 1.3.4, we conclude that $0 < ac - bc$. Applying part (2) again, it follows that $bc < ac$.

7. Suppose that $a < 0$ and $b > 0$. Using part (5), we conclude that $ab < 0 \cdot b$. Since $0 \cdot b = 0$, it follows that $ab < 0$.

8. Suppose that $a < 0$ and $b < 0$. Using part (6), we conclude that $0 \cdot b < ab$. Since $0 \cdot b = 0$, it follows that $0 < ab$.

9. Let $a \in F$ be arbitrary. By the third axiom of ordered fields, we know that one of $a < 0$, $a = 0$, or $0 < a$ is true. If $a = 0$, then $a^2 = 0 \cdot 0 = 0 \geq 0$. If $0 < a$, then using the fifth axioms of ordered fields, it follows that $0 < a \cdot a$, so $0 < a^2$. Finally, if $a < 0$, then by part (8), it follows that $a \cdot a > 0$, so $0 < a^2$.

10. We first show that $0 < 1$. Since $1^2 = 1$, we can use part (9) to conclude that $1 \geq 0$. Now note that 0 and 1 are assumed to be distinct in any field (see the definition of a field), so $1 \neq 0$. Therefore, $0 < 1$.

    Since $0 < 1$, we can add 1 to both sides of the inequality to conclude that $0 + 1 < 1 + 1$, and hence that $1 < 1 + 1$. By repeatedly adding 1 to each side, we obtain the later inequalities.

□

Using the last property, together with transitivity of $<$, it follows that $0 < 1 + 1 + \cdots + 1$ for any finite sums of 1's. In particular, we have $1+1+\cdots+1 \neq 0$ for any finite sum of 1's by part (1) of Proposition 1.3.7. If you have taken Abstract Algebra, we can rephrase this as saying that any ordered field has characteristic 0. As an consequence, notice that it is impossible to turn $\{0, 1, 2, \ldots, p-1\}$ under the modulo p operation into an ordered field, because in these fields we obtain 0 by adding 1 to itself a total of p times.

We can also use these properties to argue that $\mathbb{C}$ can not be turned into an ordered field. To see this, suppose instead that we could do it by using an ordering $<$. From above, we know that $0 < 1$. Using part (3) of Proposition 1.3.7, it follows that $-1 < -0$, and hence that $-1 < 0$. Since $-1 = i^2$ in $\mathbb{C}$, so we could also use part (9) of Proposition 1.3.7 to conclude that $0 \leq -1$. However, these two facts together contradict part (1) of Proposition 1.3.7. Therefore, we can not turn $\mathbb{C}$ into an ordered field.

Let F be an ordered field. As stated above, we know that $0 < 1 < 1 + 1 < 1 + 1 + 1 < \ldots$. Moreover, using the associative, commutative, and distributive laws, one can show facts like

$$(1 + 1) \cdot (1 + 1 + 1) = 1 + 1 + 1 + 1 + 1 + 1.$$

In other words, every ordered field contains a copy of $\mathbb{N}$. If we identity 2 with $1 + 1$ in the field, 3 with $1 + 1 + 1$ in the field, etc., then we can view every ordered field as actually containing $\mathbb{N}$ as a subset in such a way that the usual operations on $\mathbb{N}$ correspond with the ordered field operations. From here, a similar argument to the on in part (10) above (or an appeal to part (3)) shows that

$$\cdots < (-1) + (-1) + (-1) < (-1) + (-1) < -1 < 0.$$

Using the associative, commutative, and distributive laws again, it follows that arithmetic on sums of (-1)'s behaves as expected, so every ordered field contains a copy of $\mathbb{Z}$. With a copy of $\mathbb{Z}$ in hand, it is then possible to use multiplicative inverses and the other ordered field axioms to show that every ordered field contains a copy of $\mathbb{Q}$. Henceforth, we will take this fact as given, and use typical notation like $\frac{2}{5}$ for

$$\frac{1 + 1}{1 + 1 + 1 + 1 + 1} = (1 + 1) \cdot (1 + 1 + 1 + 1 + 1)^{-1}$$

in the ordered field.

With all of that interesting, but pedantic, work in hand, we have essentially argued that $\mathbb{Q}$ is the "smallest" ordered field. It might then be natural to suspect that $\mathbb{R}$ is the "biggest" ordered field. However, it turns out that there are much "bigger" ordered fields than $\mathbb{R}$! At first, this might sound really surprising. But let's think about how we could build such an object. The idea is to add an element t to $\mathbb{R}$ and make it bigger than every element of $\mathbb{R}$. Intuitively, t will be an "infinite" element, but don't think of it like the symbol ∞ or anything of the sort. Now if we want to build a field, then we need to be able to add and multiply elements in this new world, so we will also need to include elements like $t + 1$, $-2t + 5$, $t^3 + 3t - 1$, etc. Putting everything together, it looks like we will need to include all polynomials in t with real coefficients. We will adopt Abstract Algebra notation and denote the collection of polynomials with real coefficients as $\mathbb{R}[t]$.

Now how do we order the elements of $\mathbb{R}[t]$? If we intuitively think of t as a really large "infinite" element, then $2t$ should be bigger still, and t^2 should be even bigger still. More formally, if we want $0 < t$, and we want to be in an ordered field, then by adding t to both sides we see that we are forced to accept $t < 2t$. Also, since $1 < t$, we should be able to multiply this inequality on both sides by $t > 0$ to conclude that $t < t^2$. We can work through specific examples in this way, but how should we order things in general? Given a nonzero polynomial $p(t)$, we say that $p(t)$ is *positive* if its leading coefficient is a positive real number. Thus, $5t^2 - 27t - 108$ is positive, but $-t^2 + 3t - 5$ is not. Intuitively, larger powers of t are ever more "infinite" and completely dominate the small powers, so only the leading coefficient matters when determining whether an

element should be greater than 0. Now given two polynomials $p(t)$ and $q(t)$, we define $p(t) < q(t)$ to mean that $q(t) - p(t)$ is positive.

With this definition, we *almost* have an ordered field. All but one axiom is true in this world. The only problem is that not every nonzero element of $\mathbb{R}[t]$ has a multiplicative inverses. In particular, the polynomial t does not have a multiplicative inverse as there does not exist $p(t) \in \mathbb{R}[t]$ with $t \cdot p(t) = 1$. In order to fix this, we need to extend $\mathbb{R}[t]$ to a field. If you have taken Abstract Algebra, then you know that any integral domain can be extended to a field through the field of fractions construction, which is completely analogous to how $\mathbb{Q}$ is built from $\mathbb{Z}$. It is then possible to extend the ordering on $\mathbb{R}[t]$ to this field in order to make it an ordered field. We omit the details, but I encourage you to try to work them out if you have the background.

We need a different approach to understand what makes $\mathbb{R}$ special. We want to say that $\mathbb{R}$ has no "holes" in it, and the example from last section gives us some insight for how to do this while only referring to the ordering $<$. Consider again the sets

$$A = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 < 2\}$$
$$B = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 > 2\}.$$

We showed in Proposition 1.2.1 that A does not have a maximum, and that B does not have a minimum. And since $A \cup B = \{q \in \mathbb{Q} : q > 0\}$, it follows that there is no numbers that serves as a dividing line between the two sets. To express this in terms of the ordering $<$, we noted that although A has an upper bound, it does not have a *smallest* upper bound (the set B of upper bounds for A has no minimum). We now codify these ideas in the following definitions.

Definition 1.3.8. *Let F be an ordered field, and let $A \subseteq F$.*

1. *Given $b \in F$, we say that b is an* upper bound *for A if $a \leq b$ for all $a \in A$.*
2. *Given $b \in F$, we say that b is a* lower bound *for A if $b \leq a$ for all $a \in A$.*
3. *We say that A is* bounded above *if there exists an upper bound (in F) for A.*
4. *We say that A is* bounded below *if there exists a lower bound (in F) for A.*
5. *We say that A is* bounded *if it is both bounded above and bounded below, i.e. if there exists both a lower bound and an upper bound for A.*

Notice that in the definition of bounded above, we are assuming than an upper bound exists *in* F. If we extend $\mathbb{R}$ to a larger ordered field F that has an element t greater than every $r \in \mathbb{R}$ (as roughly outlined above), then the set $\mathbb{R} \subseteq F$ is bounded above when viewed inside F (since t is an upper bound), but of course $\mathbb{R}$ is not bounded above when viewed as a subset of the ordered field $\mathbb{R}$.

We can now use these concepts to express the idea of an ordered field having no "holes" with the following axiom.

Axiom 1.3.9 (Completeness Axiom)**.** *Let F be an ordered field. The* Completeness Axiom *for F is the statement that every nonempty subset of F that is bounded above has a least upper bound.*

The ordered field $\mathbb{Q}$ does not satisfy the Completeness Axiom, because Proposition 1.2.1 gives an example of nonempty subset A of $\mathbb{Q}$ that is bounded above, but which does not have a least upper bound. To see this spelled out carefully, notice that no element of A is an upper bound of A because A does not have a maximum. Next, we use the fact that every element of A is less than every element of B to conclude that every element of B is an upper bound for A. Now since an upper bound for A must be positive, and since $A \cup B = \{q \in \mathbb{Q} : q > 0\}$, we can also conclude that every upper bound of A must be in B. In other words, B is the set of upper bounds of A. Finally, since B does not have a minimum, it follows that A does not have a least upper bound.

In fact, it is not at all clear that *any* ordered field would satisfy the Completeness Axiom. Of course, we might want to think that $\mathbb{R}$ does satisfy the Completeness Axiom, because it intuitively has filled in all of the "holes" of $\mathbb{Q}$. But that is just a vague hope, especially because we have not yet said what $\mathbb{R}$ is! However, there is a truly remarkable result about ordered fields that solves all of these issues.

Theorem 1.3.10. *There exists a unique ordered field F (up to isomorphism) that satisfies the Completeness Axiom, i.e. there exists a unique ordered field F (up to isomorphism) with the property that every nonempty set that is bounded above has an upper bound.*

For the existence part of the proof, one constructs an explicit example of such an ordered field by starting with $\mathbb{Q}$ and suitably extending it. There are many known constructions that accomplish this task, but two stand out as the most standard. The first uses *Dedekind cuts* of the rationals, and the second employs equivalence classes of *Cauchy sequences* of rationals. Each of these constructions is somewhat intricate and requires a nontrivial amount of effort. The uniqueness part of the proof is actually easier, but still requires some time to work through the details. From my perspective, these constructions and arguments are best done later, once we have a better appreciation for the power and utility of the Completeness Axiom. As such, we will defer them to the end of the course, hoping that we have enough time. If you are impatient, see Section 8.6 of Abbott for an outline of the existence part of the argument.

For now, we will take the above theorem as true, and use it as our basis for understanding of the real numbers. Since there is an essentially unique such ordered field, we simply *define* the real numbers to be that object.

Definition 1.3.11. *We define $\mathbb{R}$ be to be the unique ordered field that satisfies the Completeness Axiom.*

At this point, you might be somewhat skeptical of our chosen definition of $\mathbb{R}$. After all, since we have only provided axioms, we have not actually said what a real number *is*. Many mathematicians would respond to such an objection by saying that it does not matter what a mathematical object truly is, because we can always change the names to find isomorphic versions of the object. And those mathematicians would then typically add that the essence of a mathematical object is captured by the properties that it satisfies. Since we are defining the reals in terms of the properties that they satisfy (and using the above theorem to reassure ourselves that such an object exists), we can dispense with the ugly and largely irrelevant details of any underlying construction.

Nonetheless, you might still feel uneasy. After all, does the unique ordered field that satisfies the Completeness Axiom "look like" our intuitive picture of the reals? Centuries of experience tell us that the answer is yes. For a simple example, we know from the discussion above that $\mathbb{R}$ contains a copy of the rationals $\mathbb{Q}$ (since every ordered field does). Of course, you should not just take my word, and the experience of past mathematicians, as strong evidence. Instead, we should begin the process of proving results from the axioms to reassure ourselves that we have found the right definition.

1.4 Supremums and Infimums

In light of Theorem 1.3.10, the most important property of the real numbers is that they satisfy the Completeness Axiom, which asserts that every nonempty $A \subseteq \mathbb{R}$ that is bounded above has a least upper bound. In fact, every nonempty $A \subseteq \mathbb{R}$ that is bounded above has a *unique* least upper bound, because if s and t are both least upper bounds for A, then we must have $s \leq t$ (because s is a least upper bound and t is an upper bound) and also $t \leq s$, so $s = t$. In other words, we can simply say *the* least upper bound instead of *a* least upper bound. But since the phrase "least upper bound" is a bit long, we introduce other notation and terminology for these numbers.

Definition 1.4.1. *If $A \subseteq \mathbb{R}$ is nonempty and bounded above, then we denote the unique least upper bound of A by $\sup A$, and call $\sup A$ the* supremum *of A.*

The key intuition is that a supremum is like a maximum, but in contrast to a maximum, it might not be in the actual set. For example, consider the open interval $A = (0, 1) = \{x \in \mathbb{R} : 0 < x < 1\}$. Clearly, A is nonempty and bounded above (by 1, or by 2, etc). However, A does not have a maximum, because if $x \in A$, then $x < \frac{x+1}{2} < 1$ by the homework, and so $\frac{x+1}{2}$ is a larger element of A. Now we know that A has a least upper bound by the Completeness Axiom. Intuitively, it seems clear that $\sup(A) = 1$. Let's work through how we would prove this.

First, notice that for all $x \in A$, we immediately have $x \leq 1$ by the definition of A, so 1 is an upper bound of A. Now we just need to show that 1 is the *least* such upper bound for A. One way to do this is to argue that no smaller number is an upper bound for A. So let $y \in \mathbb{R}$ with $y < 1$ be arbitrary. If $y \leq 0$, then y is not an upper bound for A, because $\frac{1}{2} \in A$ and $y < \frac{1}{2}$. Suppose then that $y > 0$. We then have $0 < y < \frac{y+1}{2} < 1$ (again by the homework), so $\frac{y+1}{2}$ is an element of A that is larger than y, and hence y is not an upper bound for A. We have shown that every $y \in \mathbb{R}$ with $y < 1$ fails to be an upper bound for A. Therefore, we can now conclude that $\sup(A) = 1$.

Now consider the closed interval $B = [0, 1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$. In this case, 1 is the maximum element of B. Using this fact, it is fairly straightforward to see that 1 is the least upper bound for B (it is certainly an upper bound, and any other upper bound must be greater than or equal to 1, since 1 is actually in the set). Thus, we still have $\sup(B) = 1$.

In general, given a nonempty set $A \subseteq \mathbb{R}$ that is bounded above and also a number $s \in \mathbb{R}$, we need to do two things to argue that $s = \sup A$:

1. Show that s is an upper bound of A, i.e. that $a \leq s$ for all $a \in A$.
2. Show that no number strictly less than s is an upper bound for A. There are at least two natural ways to accomplish this task:
   - Show that given any $t < s$, there exists an $a \in A$ with $t < a$, i.e. no number strictly less than s is an upper bound for A.
   - Show that given any upper bound b of A, we have $s \leq b$, i.e. every upper bound of A is greater than or equal to s.

There are also some other variations on the two methods described that one can employ to show the "least" part. Here is one simple, but important, example.

Proposition 1.4.2. *Suppose that $A \subseteq \mathbb{R}$ is nonempty, and that $s \in \mathbb{R}$ is an upper bound for A. The following are equivalent:*

1. *$s = \sup A$.*
2. *For all $\varepsilon > 0$, there exists $a \in A$ with $a > s - \varepsilon$.*

Proof.

- (1) $\Rightarrow$ (2): Assume (1), i.e. that $s = \sup A$. Let $\varepsilon > 0$ be arbitrary. Multiplying both sides by -1, and then adding s to both sides, we conclude that that $s - \varepsilon < s$. Since s is the least upper bound of A, and $s - \varepsilon < s$, we conclude that $s - \varepsilon$ is not an upper bound of A. Therefore, there exists $a \in A$ with $a > s - \varepsilon$.
- (2) $\Rightarrow$ (1): Assume (2). We are assuming in the problem that that s is an upper bound of A, so we need only show that s is the least upper bound. Let $t \in \mathbb{R}$ with $t < s$ be arbitrary. Letting $\varepsilon = s - t$, notice that $\varepsilon > 0$ and that $s - \varepsilon = t$. By assumption, we can fix $a \in A$ with $a > s - \varepsilon = t$. We have established that there exists an element of A that is greater than t, so t is not an upper bound for A. Since $t < s$ was arbitrary, we conclude that no number smaller than s is an upper bound for A. Therefore, s is the least upper bound of A, which is to say that $s = \sup A$.

□

Notice that the Completeness Axiom is entirely about upper bounds and least upper bounds. What about lower bounds and greatest lower bounds? Why didn't we include the existence of greatest lower bounds (for nonempty sets that are bounded below) in the axiom? The answer is that we can prove their existence from the Completeness Axiom.

Proposition 1.4.3. *If $B \subseteq \mathbb{R}$ is a nonempty set that is bounded below, then there exists a greatest lower bound of B.*

Proof. Let $A = \{-b : b \in B\}$, i.e. A is the set of all negatives of elements of B. We first establish an important lemma.

Lemma: If d is a lower bound of B, then $-d$ is an upper bound of A.
Proof: Suppose d is a lower bound of B. Let $a \in A$ be arbitrary. By definition of A, we can fix $b \in B$ with $a = -b$. Since $b \in B$ and d is a lower bound for B, we have $d \leq b$. Multiplying both sides by $-1 < 0$, we conclude that $-b \leq -d$, and hence $a \leq -d$. Since $a \in A$ was arbitrary, it follows that $-d$ is an upper bound for A.

We now return to the proof. Since B is bounded below by assumption, we know that B has a lower bound, so we can use the lemma to conclude that A is bounded above. Since $A \subseteq \mathbb{R}$ is nonempty (because $B \neq \emptyset$) and bounded above, we know by the Completeness Axiom that A has a least upper bound. So let $c = \sup A$. We claim that $-c$ is a greatest lower bound of B.

- We first show that $-c$ is a lower bound of B. Let $b \in B$ be arbitrary. We then have $-b \in A$ by definition. Since c is an upper bound for A, it follows that $-b \leq c$. Multiplying both sides by $-1 < 0$, we conclude that $-c \leq b$. Since $b \in B$ was arbitrary, it follows that $-c$ is a lower bound of B.

- Let $d \in \mathbb{R}$ be an arbitrary lower bound of B. From above, we know that $-d$ is an upper bound of A. Since $c = \sup A$, it follows that $c \leq -d$. Multiplying both sides by $-1 < 0$, we conclude that $d \leq -c$. Thus, every lower bound of B is less than $-c$.

Therefore, $-c$ is indeed a greatest lower bound of B. □

Although the above proof showed the existence of greatest lower bounds for sets that are nonempty and bounded below, we also immediately obtain uniqueness (as we did for least upper bounds).

Definition 1.4.4. *If $A \subseteq \mathbb{R}$ is nonempty and bounded below, then we denote the unique greatest lower bound of A by* $\inf A$*, and call* $\inf A$ *the* infimum *of A.*

We now start to prove some fundamental consequences of the Completeness Axiom. Our first result will be about how $\mathbb{N}$ sits inside $\mathbb{R}$. The set $\mathbb{N}$ is not bounded above in our intuitive picture of the real numbers. That is, given any real number, we feel like we can "round up" to a strictly larger natural number. Although this fact is "obvious" in our intuitive understanding of the reals, recall that we defined $\mathbb{R}$ in a seemingly strange way as the unique ordered field that satisfies the Completeness Axiom. Thus, we have to prove it from the axioms.

Before diving into the argument, it is interesting and important to note that there exist ordered fields where $\mathbb{N}$ is bounded above. For example, consider the polynomials with real coefficients $\mathbb{R}[t]$, where we made t strictly greater than every real number. Recall that although this is not a field, it can be extended to an ordered field F. In F, the set $\mathbb{N}$ is bounded! To see this, simply notice that $n < t$ for all $n \in \mathbb{N}$, so t is an upper bound of $\mathbb{N}$. However, it turns out that this field does not satisfy the Completeness Axiom, and in fact although the set $\mathbb{N}$ is bounded above, there is no least upper bound of $\mathbb{N}$ in F. For example, although t is an upper bound of $\mathbb{N}$, it is not the least upper bound, because $t - 1$ is also an upper bound. Moreover,

$t - k$ is an upper bound of $\mathbb{N}$ for all k. In fact, it can be shown that whenever $a \in F$ is an upper bound of $\mathbb{N}$, we then have that $a - 1$ is an upper bound as well.

Thus, if we want to argue that $\mathbb{N}$ is unbounded in $\mathbb{R}$, then we will have to make essential use of the Completeness Axiom, which we now proceed to do.

Proposition 1.4.5. *For all $x \in \mathbb{R}$, there exists $n \in \mathbb{N}$ with $n > x$.*

Proof. Let $x \in \mathbb{R}$ be arbitrary. First note that if $x < 0$, then we can take $n = 0$. Assume then that $x \geq 0$. Let $A = \{n \in \mathbb{N} : n \leq x\}$. Notice that $A \neq \emptyset$ (because $0 \in A$) and that A is bounded above (by x). Let $s = \sup A$, and notice that $s \geq 0$ because $0 \in A$. Since s is the least upper bound of A and $s - 1 < s$, it follows that $s - 1$ is not an upper bound of A, and hence we can fix $m \in A$ with $m > s - 1$. Adding 1 to both sides, we conclude that $m + 1 > s$. As s is an upper bound of A, it follows that $m + 1 \notin A$. But notice that $m + 1 \in \mathbb{N}$ because $m \in \mathbb{N}$, so we must have $m + 1 \not\leq x$. Therefore, $x < m + 1$, completing the proof. □

We now obtain a couple of simple consequences of this result. The first is typically called the Archimedean Property of $\mathbb{R}$. Intuitively, it says that given any two positive real numbers x and y, if we keep adding x to itself over and over to form values like $x + x + x + \cdots + x$, then we will eventually exceed the value y. If we think of x and y as representing the lengths of line segments, then no matter how small x is, or how large y is, we can eventually put together enough line segments of length x until they, taken together, exceed the length of y.

Corollary 1.4.6 (Archimedean Property of $\mathbb{R}$)**.** *For all $x, y \in \mathbb{R}$ with $x > 0$ and $y > 0$, there exists $n \in \mathbb{N}$ with $nx > y$.*

Proof. Let $x, y \in \mathbb{R}$ be arbitrary with both $x > 0$ and $y > 0$. By Proposition 1.4.5, we can fix $n \in \mathbb{N}$ with $n > \frac{y}{x}$. Multiplying both sides by $x > 0$, we conclude that $nx > y$. □

We can also use Proposition 1.4.5 to argue that the values of $\frac{1}{n}$ for $n \in \mathbb{N}^+$ become arbitrarily small within the positive real numbers.

Corollary 1.4.7. *For all $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$, there exists $n \in \mathbb{N}^+$ with $\frac{1}{n} < \varepsilon$.*

Proof. Let $\varepsilon > 0$ be arbitrary. Using Proposition 1.4.5, we can fix $n \in \mathbb{N}$ with $n > \frac{1}{\varepsilon}$. Notice that $n > 0$ because $\frac{1}{\varepsilon} > 0$. Multiplying both sides of the inequality $n > \frac{1}{\varepsilon}$ by $\frac{\varepsilon}{n} > 0$, it follows that $\varepsilon > \frac{1}{n}$. □

With all of the groundwork in place, we can finally prove a theorem that is not intuitively obvious. We know that between any two rationals numbers we can always find a rational number, and between any two real numbers we can always find a real number (in both cases by simply taking the average). We will now show that between any two real numbers, we can always find a rational number. Before diving into the proof, we first sketch the idea. Let $x, y \in \mathbb{R}$ with $x < y$. For simplicity in visualizing the argument, let's assume that $0 < x < y$. We are trying to build a rational number $\frac{m}{n}$ where $m, n \in \mathbb{N}^+$ and $x < \frac{m}{n} < y$. Here is an outline of the argument.

- First we pick a really large denominator n so that $\frac{1}{n}$ is really small. How small? If we want to consider the rationals $\frac{1}{n}$, $\frac{2}{n}$, etc., then we should pick n so that this sequence of rational numbers does not "jump over" the potentially small gap between x and y. So we should pick n big enough so that $\frac{1}{n} < y - x$.
- Next we want to argue that if we keep adding $\frac{1}{n}$ to itself, then we eventually surpass x. The key to knowing the we will eventually reach this point is the Archimedean Property of $\mathbb{R}$.
- Once we know that we can add $\frac{1}{n}$ to itself enough times to get past x, the idea is to do this just enough times so that we barely surpass x. That is, we want to take the *least* $m \in \mathbb{N}$ with $\frac{m}{n} > x$. To that end, recall the well-ordering property of the natural numbers, which states that every nonempty subset of $\mathbb{N}$ has a minimum.

With those ideas in place, we now turn to the proof.

Theorem 1.4.8 (Density of $\mathbb{Q}$ in $\mathbb{R}$)**.** *For all $x, y \in \mathbb{R}$ with $x < y$, there exists $q \in \mathbb{Q}$ with $x < q < y$.*

Proof. We first prove the claim in the case where we also have $x > 0$. So let $x, y \in \mathbb{R}$ be arbitrary with $x < y$ and $x > 0$. Since $y - x > 0$, we can apply Corollary 1.4.7 to fix an $n \in \mathbb{N}^+$ with $\frac{1}{n} < y - x$. Let

$$A = \left\{ a \in \mathbb{N} : \frac{a}{n} > x \right\}.$$

First notice that $A \neq \emptyset$ by the Archimedean Property of the reals, as we know that there exists $a \in \mathbb{N}$ with $a \cdot \frac{1}{n} > x$. Since $A \subseteq \mathbb{N}$, we can use the well-ordering of $\mathbb{N}$ to conclude that A has a minimum. Let $m = \min(A)$. Since $m \in A$, we trivially have $x < \frac{m}{n}$.

We now show that $\frac{m}{n} < y$. First notice that $m > 0$ because $m \in A$ and $x > 0$. Thus, $m - 1 \in \mathbb{N}$. Since $m = \min(A)$ and $m - 1 < m$, it follows that $m - 1 \notin A$, and hence $\frac{m-1}{n} \leq x$. Adding $\frac{1}{n}$ to both sides, we conclude that $\frac{m}{n} < x + \frac{1}{n}$. Now $\frac{1}{n} < y - x$, so $x + \frac{1}{n} < y$. By transitivity of $<$, we conclude that $\frac{m}{n} < y$.

We have shown that $x < \frac{m}{n} < y$. Since $m, n \in \mathbb{N}$ and $n > 0$, we have shown the existence of a $q \in \mathbb{Q}$ with $x < q < y$ in the case where $x > 0$.

Suppose now that $x \leq 0$. Using Proposition 1.4.5, we can fix $k \in \mathbb{N}^+$ with $k > -x$. We then have $x + k < y + k$, and also $x + k > 0$. By the result we just proved, we can fix $q \in \mathbb{Q}$ with $x + k < q < y + k$. We then have $q - k \in \mathbb{Q}$, and $x < q - k < y$. □

We now show how we can use the Completeness Axiom to prove that every nonnegative number has a square root. The heart of the argument resembles the proof of Proposition 1.2.1. In broad outline, let $a \in \mathbb{R}$ be arbitrary with $a > 0$. We want to show the existence of a number $c \geq 0$ with $c^2 = a$. The idea is to look at the set $A = \{x \in \mathbb{R} : x > 0 \text{ and } x^2 < a\}$ of all nonnegative numbers whose square is strictly less than a. It is reasonably straightforward to show that this set is nonempty and bounded above, so it has a supremum. Let $c = \sup A$. To show that $c^2 = a$, we rule out the other two possibilities. Here is the intuition:

- If $c^2 < a$, then we can use the argument in the proof of Proposition 1.2.1 to show that there exists $n \in \mathbb{N}^+$ with $(c + \frac{1}{n})^2 < a$. It then follows that $c + \frac{1}{n} \in A$, contradicting the fact that c is an upper bound of A.
- If $c^2 > a$, then we can use the argument in the proof of Proposition 1.2.1 to show that there exists $n \in \mathbb{N}^+$ with $(c - \frac{1}{n})^2 > a$. Now $c - \frac{1}{n} < c$, so $c - \frac{1}{n}$ is not an upper bound for A. In other words, we can find an element of A that is bigger than $c - \frac{1}{n}$, from which it is natural to conclude that the square of this element is bigger than a, which would be a contradiction. There is a small issue here in that in order to make the last part of the argument work, we need to know that $c - \frac{1}{n} > 0$. Intuitively, we can always do this by making n larger still, but it will requires a small tweak in the choice of n at the beginning.

We now turn to the details.

Theorem 1.4.9. *For all $a \in \mathbb{R}$ with $a \geq 0$, there exists $c \in \mathbb{R}$ with $c \geq 0$ and $c^2 = a$.*

Proof. Let $a \in \mathbb{R}$ with $a \geq 0$ be arbitrary. Notice that if $a = 0$, then we can take $c = 0$. Also, if $a = 1$, then we can take $c = 1$. So assume that $a > 0$ and $a \neq 1$. Let

$$A = \{x \in \mathbb{R} : x > 0 \text{ and } x^2 < a\}.$$

We first claim that A is nonempty and bounded above. We prove this by considering two cases:

- Suppose first that $0 < a < 1$. Multiplying these inequalities by $a > 0$, we see that $0 < a^2 < a$, so $a \in A$, and hence A is nonempty. Also, notice that for any $x \in A$, we must have $x \leq 1$, because if $x > 1$, then $x^2 > 1 > a$. Therefore, A is bounded above by 1.

- Suppose now that $a > 1$. We then have $1^2 = 1 < a$, so $1 \in A$, and hence A is nonempty. Also, notice that for any $x \in A$, we must have $x < a$, because if $x \geq a$, then $x^2 \geq a^2 > a$. Therefore, A is bounded above by a.

Thus, in either case, we know that A is nonempty and bounded above. Let $c = \sup A$, and notice that $c > 0$ because A contains a positive element, and c is greater than or equal to every element of A. We show that $c^2 = a$ by showing that the other two options lead to contradictions.

- Suppose that $c^2 < a$. By Proposition 1.4.5, we can fix an $n \in \mathbb{N}^+$ with

$$n > \frac{2c+1}{a-c^2}.$$

Multiplying both sides of this inequality by $\frac{a-c^2}{n} > 0$, we conclude that

$$\frac{2c+1}{n} < a - c^2.$$

Now since $n \leq n^2$, we have $\frac{1}{n^2} \leq \frac{1}{n}$, and hence

$$\begin{aligned}
\left(c + \frac{1}{n}\right)^2 &= c^2 + \frac{2c}{n} + \frac{1}{n^2} \\
&\leq c^2 + \frac{2c}{n} + \frac{1}{n} \\
&= c^2 + \frac{2c+1}{n} \\
&< c^2 + (a - c^2) && \text{(from above)} \\
&= a.
\end{aligned}$$

Therefore, $c + \frac{1}{n} \in A$, which contradicts the fact that c is an upper bound for A.

- Suppose that $a < c^2$, so $c^2 - a > 0$. By Proposition 1.4.5, we can fix an $n \in \mathbb{N}^+$ with

$$n > \max\left\{\frac{2c}{c^2-a}, \frac{1}{c}\right\}.$$

Since $n > \frac{2c}{c^2-a}$, we can multiply both sides by $\frac{c^2-a}{n} > 0$ to conclude that

$$c^2 - a > \frac{2c}{n}.$$

Multiplying by sides of this inequality by $-1 < 0$, it follows that

$$a - c^2 < -\frac{2c}{n},$$

and hence

$$\begin{aligned}
\left(c - \frac{1}{n}\right)^2 &= c^2 - \frac{2c}{n} + \frac{1}{n^2} \\
&\geq c^2 - \frac{2c}{n} && \left(\text{since } \frac{1}{n^2} \geq 0\right) \\
&> c^2 + (a - c^2) && \text{(from above)} \\
&= a.
\end{aligned}$$

We also know that $n > \frac{1}{c}$, so multiplying both sides by $\frac{c}{n} > 0$, we see that $c > \frac{1}{n}$, and hence $c - \frac{1}{n} > 0$. Now c is the least upper bound of A, and $c - \frac{1}{n} < c$, so $c - \frac{1}{n}$ is not an upper bound of A. Thus, we can fix $x \in A$ with $x > c - \frac{1}{n}$. Since $0 < c - \frac{1}{n} < x$, it follows that $(c - \frac{1}{n})^2 < x^2$, and hence $x^2 > a$. However, this contradicts the fact that $x \in A$.

Since both cases lead to a contradiction, we conclude that $c^2 = a$. □

Since $(-c)^2 = c^2$ for all $c \in \mathbb{R}$, we can immediately conclude that for all $a \in \mathbb{R}$ with $a \geq 0$, there exists $c \in \mathbb{R}$ with $c \leq 0$ and $c^2 = a$. Now by Problem 5d on Homework 1, we know that if $c, d \in \mathbb{R}$ and $c^2 = d^2$, then either $c = d$ or $c = -d$, so for any $a \in \mathbb{R}$, the set $\{x \in \mathbb{R} : x^2 = a\}$ has at most two elements. Combining these facts, we conclude that for each $a \in \mathbb{R}$ with $a \neq 0$, there a unique positive $c \in \mathbb{R}$ with $c^2 = a$, and there exists a unique negative $c \in \mathbb{R}$ with $c^2 = a$.

Definition 1.4.10. *Let $a \in \mathbb{R}$ with $a \geq 0$. We denote the unique nonnegative $c \in \mathbb{R}$ with $c^2 = a$ by $\sqrt{a}$.*

Before moving on to a another fundamental manifestation of Completeness Axiom in the reals, we pause to talk about unions and intersections. Of course, given two sets A and B, the union $A \cup B$ is the set of elements that are in at least one of A or B, and the intersection $A \cap B$ is the set of elements that are in both A and B. Given three sets A, B, and C, we can form the "union" by either looking at $(A \cup B) \cup C$ or $A \cup (B \cup C)$. However, it is straightforward to see that each of these is just the set of elements that are in at least one of A, B, or C. Thus, we do not need parentheses, and can simply write $A \cup B \cup C$. Similarly, we can just write $A \cap B \cap C$, as this set consists of those elements that are in all three of A, B, and C.

We can then generalize this construction to any finite collection of sets $A_1, A_2, \ldots, A_m$. Although we could (and sometimes do) write $A_1 \cup A_2 \cup \cdots \cup A_m$ and also $A_1 \cap A_2 \cap \ldots A_m$, we also use the notation

$$\bigcup_{n=1}^{m} A_n \qquad \text{and} \qquad \bigcap_{n=1}^{m} A_n$$

to denote the union and intersection, respectively. Notice that this notation matches up with the summation notation

$$\sum_{n=1}^{m} a_n = a_1 + a_2 + \cdots + a_m$$

in that we start the index n at the bottom number, increment it repeatedly by 1 until we equal the top element, and then perform the corresponding operation (union, intersection, sum) on the various terms.

What if we have an infinite collection of objects, rather than a finite collection? For unions and intersections, there is no issue in generalizing the above construction. For example, suppose that we have a set A_n for each $n \in \mathbb{N}^+$, so that we have a sequence $A_1, A_2, A_3, \ldots$ of sets. We can still form the union of this collection of sets by simply taking the set of elements that are in at least one of the A_n. In other words, the union $A_1 \cup A_2 \cup A_3 \cup \ldots$ is the set of elements x such that *there exists* $n \in \mathbb{N}^+$ with $x \in A_n$. Similarly, the intersection $A_1 \cap A_2 \cap A_3 \cap \ldots$ is the set of elements x such that *for all* $n \in \mathbb{N}^+$, we have $x \in A_n$. Taking a cue from the above notation, we write

$$\bigcup_{n=1}^{\infty} A_n = \{x : \text{There exists } n \in \mathbb{N}^+ \text{ with } x \in A_n\}$$

$$\bigcap_{n=1}^{\infty} A_n = \{x : \text{For all } n \in \mathbb{N}^+, \text{ we have } x \in A_n\}$$

Although this is a reasonably natural extension of the above notation, notice one key difference: we do not ever "reach" the upper bound ∞ when incrementing n. In other words, you should *not* take the symbol ∞

literally in the symbolism. Rather, it just serves to signify that we repeatedly increment n forever without end. Eventually, we will also give a careful definition of

$$\sum_{n=1}^{\infty} a_n,$$

but there is more subtlety in that situation. But let's consider a few examples of infinite unions and intersections.

- Let $A_n = \{n\}$ for each $n \in \mathbb{N}^+$. We then have

$$\bigcup_{n=1}^{\infty} A_n = \mathbb{N}^+ \qquad \text{and} \qquad \bigcap_{n=1}^{\infty} A_n = \emptyset.$$

- Let $A_n = (-n, n) = \{x \in \mathbb{R} : -n < x < n\}$ for each $n \in \mathbb{N}^+$. We then have

$$\bigcup_{n=1}^{\infty} A_n = \mathbb{R} \qquad \text{and} \qquad \bigcap_{n=1}^{\infty} A_n = (-1, 1).$$

  The former of these equalities relies on Proposition 1.4.5, and the latter follows from the fact that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, so the intersection is just the first set.

- Let $A_n = \left(0, \frac{1}{n}\right) = \left\{x \in \mathbb{R} : 0 < x < \frac{1}{n}\right\}$ for each $n \in \mathbb{N}^+$. We then have

$$\bigcup_{n=1}^{\infty} A_n = (0, 1) \qquad \text{and} \qquad \bigcap_{n=1}^{\infty} A_n = \emptyset.$$

  The former of these equalities follows from the $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, and the latter relies on Proposition 1.4.7.

Let's examine the last example a little more closely. If we had a *finite* nested decreasing sequence of nonempty sets $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_m$, then the intersection is $A_m \neq \emptyset$. However, as the above example demonstrates, we can have an *infinite* nested decreasing sequence of nonempty sets $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$ with empty intersection. Notice that if instead we have used closed intervals $A_n = \left[0, \frac{1}{n}\right] = \left\{x \in \mathbb{R} : 0 \leq x \leq \frac{1}{n}\right\}$, then the infinite intersection would be $\{0\}$, which is nonempty. We now argue that we *always* obtain a nonempty intersection when we have a nested sequence of nonempty closed intervals.

Theorem 1.4.11 (Nested Interval Theorem)**.** *Suppose that for each $n \in \mathbb{N}^+$, we have a closed interval $I_n = [a_n, b_n] = \{x \in \mathbb{R} : a_n \leq x \leq b_n\}$ with $a_n \leq b_n$. Assume further that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \ldots$, i.e. that $a_1 \leq a_2 \leq a_3 \leq \ldots$ and $b_1 \geq b_2 \geq b_3 \geq \ldots$. We then have that $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$, i.e. there exists $c \in \mathbb{R}$ such that $c \in I_n$ for all $n \in \mathbb{N}^+$.*

Since we have a nice sequence $a_1 \leq a_2 \leq a_3 \leq \ldots$ of left-hand endpoints, a natural guess for an element of the intersection is $c = \sup\{a_n : n \in \mathbb{N}^+\}$. The interesting part of the argument is showing that $c \leq b_n$ for each $b_n \in \mathbb{N}^+$. The task becomes much easier if we notice that c is the least upper bound of $\{a_n : n \in \mathbb{N}^+\}$, so it suffices to argue that each b_n is an upper bound of this set (as the least upper bound will then be less than or equal to each b_n).

Proof. Let $A = \{a_n : n \in \mathbb{N}^+\}$. Notice that for any $n \in \mathbb{N}^+$, we have $a_n \leq b_n \leq b_1$, so b_1 is an upper bound for A. Since A is clearly nonempty, we can let $c = \sup A$. Since c is an upper bound of A, we certainly have $a_n \leq c$ for all $n \in \mathbb{N}^+$.

We next claim that each b_n is an upper bound for A. Let $n \in \mathbb{N}^+$ be arbitrary, and consider b_n. To show that b_n is an upper bound for A, we need to show that $a_k \leq b_n$ for all $k \in \mathbb{N}^+$. So let $k \in \mathbb{N}^+$ be arbitrary. We have two cases:

- *Case 1:* Suppose that $k \leq n$. We then have $a_k \leq a_n \leq b_n$ (the former because $a_1 \leq a_2 \leq \dots$), hence $a_k \leq b_n$.
- *Case 2:* Suppose that $k > n$. We then have $a_k \leq b_k \leq b_n$ (the latter because $b_1 \geq b_2 \geq \dots$), hence $a_k \leq b_n$.

Thus, $a_k \leq b_n$ for all $k \in \mathbb{N}^+$, and hence b_n is an upper bound for A.

We have shown that each b_n is an upper bound for A, so as $c = \sup A$ is the least upper bound for A, it follows that $c \leq b_n$ for all $n \in \mathbb{N}^+$. Therefore, $c \in I_n$ for all $n \in \mathbb{N}^+$. □

1.5 Countability and Uncountability

Given a finite set A, we let $|A|$ denote the cardinality of A, which is just the number of elements of A. Now given any two finite sets A and B, one way to determine whether A and B have the same cardinality is to determine $|A|$ and $|B|$ individually, and then check whether these two numbers are equal. Although this might seem like the most natural way to establish that two finite sets have the same number of elements, there are other powerful methods. For example, consider the following simple result.

Proposition 1.5.1. *Let A and B be finite sets. The following are equivalent:*

1. $|A| = |B|$.
2. *There exists a bijection $f : A \to B$.*

Intuitively, two finite sets have the same number of elements if and only if we can "pair off" the elements of A and B so that everybody on one side has exactly one "buddy" on the other sides. Although the bijection approach to characterizing cardinality might seem a bit more abstract, it is a fundamental tool in Combinatorics, where it is sometimes possibly to build an explicit bijection between two sets without having a nice formula for $|A|$ or $|B|$ directly. Although we will not follow this thread here, we can use the above characterization as a way to define when two *infinite* sets have the same "size". That is, we can simply *define* two sets A and B as having the same size exactly when there is a bijection between them.

With this in mind, think about $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ and the subset $\mathbb{N}^+ = \{1, 2, 3, 4, \dots\}$. Although $\mathbb{N}^+$ is a proper subset of $\mathbb{N}$ and "obviously" has one fewer element, the function $f : \mathbb{N} \to \mathbb{N}^+$ given by $f(n) = n + 1$ is a bijection, and so $\mathbb{N}$ and $\mathbb{N}^+$ have the same "size". For another even more surprising example, let $A = \{2n : n \in \mathbb{N}\} = \{0, 2, 4, 6, \dots\}$ be the set even natural numbers, and notice that the function $f : \mathbb{N} \to A$ given by $f(n) = 2n$ is a bijection from $\mathbb{N}$ to A. Hence, even though A intuitively seems to only have "half" of the elements of $\mathbb{N}$, there is still a bijection between $\mathbb{N}$ and A.

The next proposition shows that $\mathbb{N}$ is the "smallest" infinite set, in the sense that we can injectively embed it into any infinite set.

Proposition 1.5.2. *If A is an infinite set, then there is an injective function $f : \mathbb{N} \to A$.*

Proof. We define $f : \mathbb{N} \to A$ recursively. Fix some $a_0 \in A$, and define $f(0) = a_0$. Suppose that $n \in \mathbb{N}$ and we have defined the values $f(0), f(1), \dots, f(n)$, all of which are elements of A. Since A is infinite, we have that $\{f(0), f(1), \dots, f(n)\} \neq A$. Thus, we can fix some $a_{n+1} \in A \backslash \{f(0), f(1), \dots, f(n)\}$, and define $f(n+1) = a_{n+1}$. With this recursive definition, we have defined a function $f : \mathbb{N} \to A$. Notice that if $m < n$, then $f(n)$ was chosen to be distinct from $f(m)$ by definition, so $f(m) \neq f(n)$. Therefore, f is injective. □

With this in mind, we introduce a name for those infinite sets for which we can find a bijection with $\mathbb{N}$, and think of them as the "smallest" types of infinite sets.

Definition 1.5.3. *Let A be a set.*

- *We say that A is* countably infinite *if there exists a bijection $f\colon \mathbb{N} \to A$.*
- *We say that A is* countable *if it is either finite or countably infinite.*
- *If A is not countable, we say that A is* uncountable.

Suppose that A is countably infinite. We then have a bijection $f\colon \mathbb{N} \to A$, so we can arrange its elements in a list without repetition by listing out $f(0), f(1), f(2), f(3), \ldots$ to get:

$$a_0 \quad a_1 \quad a_2 \quad a_3 \quad \cdots$$

Conversely, writing out such a list without repetition shows how to build a bijection $f\colon \mathbb{N} \to A$. Since working with such lists is more intuitively natural (although perhaps a little less rigorous), we'll work with countable sets in this way. What about lists that allow repetitions?

Proposition 1.5.4. *Let A be a set. The following are equivalent.*

1. *It is possible to list A, possibly with repetitions, as $a_0, a_1, a_2, a_3, \ldots$.*
2. *There is a surjection $g\colon \mathbb{N} \to A$.*
3. *A is countable, i.e. either finite or countably infinite.*

Proof. (1) $\Leftrightarrow$ (2): This is essentially the same as the argument just given. If we can list A, possibly with repetitions, as $a_0, a_1, a_2, a_3, \ldots$, then the function $g\colon \mathbb{N} \to A$ given by $g(n) = a_n$ is a surjection. Conversely, if there is a surjection $g\colon \mathbb{N} \to A$, then $g(0), g(1), g(2), g(3), \ldots$ is a listing of A.

(1) $\Rightarrow$ (3): Suppose that there is a surjection $g\colon \mathbb{N} \to A$. If A is finite, then A is countable by definition, so we may assume that A is infinite. We define a new list as follows. Let $b_0 = a_0$. If we have defined $b_0, b_1, \ldots, b_n$, let $b_{n+1} = a_k$, where k is chosen as the least value such that $a_k \notin \{b_0, b_1, \ldots, b_n\}$ (such a k exists because A is infinite). Then

$$b_0 \quad b_1 \quad b_2 \quad b_3 \quad \cdots$$

is a listing of A without repetitions. Therefore, A is countably infinite.

(3) $\Rightarrow$ (1): Suppose that A is countable. If A is countably infinite, then there is a bijection $f\colon \mathbb{N} \to A$, in which case

$$f(0) \quad f(1) \quad f(2) \quad f(3) \quad \cdots$$

is a listing of A (even without repetition). On the other hand, if A is finite, say $A = \{a_0, a_1, a_2, \ldots, a_n\}$, then

$$a_0 \quad a_1 \quad a_2 \quad \cdots \quad a_n \quad a_n \quad a_n \quad \cdots$$

is a listing of A (with repetition). □

Our first really interesting result is that $\mathbb{Z}$, the set of integers, is countable. Of course, some insight is required because if we simply start to list the integers as

$$0 \quad 1 \quad 2 \quad 3 \quad 4 \quad \cdots$$

we won't ever get to the negative numbers. We thus use the sneaky strategy of bouncing back-and-forth between positive and negative integers.

Proposition 1.5.5. *$\mathbb{Z}$ is countable.*

Proof. We can list $\mathbb{Z}$ as

$$0 \quad 1 \quad -1 \quad 2 \quad -2 \quad \cdots$$

More formally, we could define $f\colon \mathbb{N} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} -\frac{n}{2} & \text{if } n \text{ is even} \\ \frac{n+1}{2} & \text{if } n \text{ is odd} \end{cases}$$

and check that f is a bijection. □

The key idea used in previous proof can be abstracted into the following result.

Proposition 1.5.6. *If A and B are countable, then $A \cup B$ is countable.*

Proof. Since A is countable, we may list it as $a_0, a_1, a_2, a_3, \ldots$. Since B is countable, we may list it as $b_0, b_1, b_2, b_3, \ldots$. We therefore have the following two lists:

$$\begin{array}{ccccc} a_0 & a_1 & a_2 & a_3 & \cdots \\ b_0 & b_1 & b_2 & b_3 & \cdots \end{array}$$

We can list $A \cup B$ by going back-and-forth between the above lists as

$$a_0 \quad b_0 \quad a_1 \quad b_1 \quad a_2 \quad b_2 \quad \cdots$$

□

A slightly stronger result is now immediate.

Corollary 1.5.7. *If $A_0, A_1, \ldots, A_n$ are countable, then $A_0 \cup A_1 \cup \cdots \cup A_n$ is countable.*

Proof. This follows from Proposition 1.5.6 by induction. Alternatively, we can argue as follows. For each fixed k with $0 \leq k \leq n$, we know that A_k is countable, so we may list it as $a_{k,0}, a_{k,1}, a_{k,2}, \ldots$. We can visualize the situation with the following table.

$$\begin{array}{ccccc} a_{0,0} & a_{0,1} & a_{0,2} & a_{0,3} & \cdots \\ a_{1,0} & a_{1,1} & a_{1,2} & a_{1,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ a_{n,0} & a_{n,1} & a_{n,2} & a_{n,3} & \cdots \end{array}$$

We now list $A_0 \cup A_1 \cup \cdots \cup A_n$ by moving down each column in order, to obtain:

$$a_{0,0} \quad a_{1,0} \quad \cdots \quad a_{n,0} \quad a_{0,1} \quad a_{1,1} \quad \cdots \quad a_{n,1} \quad \cdots \quad \cdots$$

□

In fact, we can prove quite a significant extension of the above results. The next proposition is usually referred to by saying that "the countable union of countable sets is countable".

Proposition 1.5.8. *If $A_0, A_1, A_2, \ldots$ are all countable, then $\bigcup_{k=0}^{\infty} A_k = A_0 \cup A_1 \cup A_2 \cup \cdots$ is countable.*

Proof. For each $n \in \mathbb{N}$, we know that A_n is countable, so we may list it as $a_{k,0}, a_{k,1}, a_{k,2}, a_{k,3}, \ldots$. We now have the following table.

$$\begin{array}{ccccc} a_{0,0} & a_{0,1} & a_{0,2} & a_{0,3} & \cdots \\ a_{1,0} & a_{1,1} & a_{1,2} & a_{1,3} & \cdots \\ a_{2,0} & a_{2,1} & a_{2,2} & a_{2,3} & \cdots \\ a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Now we can't list this by blindly walking down the rows or columns. We thus need a new, much more clever, strategy. The idea is to list the elements of the table by moving between rows and columns. One nice approach which works is to step along certain diagonals and obtain the following listing of $\bigcup_{n=0}^{\infty} A_n$:

$$a_{0,0} \quad a_{0,1} \quad a_{1,0} \quad a_{0,2} \quad a_{1,1} \quad a_{2,0} \quad \cdots$$

The pattern here is that we are walking along the diagonals in turn, each of which is finite. Alternatively, we can describe this list as follows. For each $m \in \mathbb{N}$, there are only finitely many pairs $(i, j) \in \mathbb{N} \times \mathbb{N}$ with $i + j = m$. We first list the finitely many $a_{i,j}$ with $i + j = 0$, followed by those finitely many $a_{i,j}$ with $i + j = 1$, then those finitely many $a_{i,j}$ with $i + j = 2$, etc. This gives a listing of $\bigcup_{k=0}^{\infty} A_k$. □

Theorem 1.5.9. $\mathbb{Q}$ *is countable.*

Proof. For each $k \in \mathbb{N}^+$, let $A_k = \left\{\frac{a}{k} : a \in \mathbb{Z}\right\}$. Notice that each A_k is countable because we can list it as

$$\frac{0}{k} \quad \frac{1}{k} \quad \frac{-1}{k} \quad \frac{2}{k} \quad \frac{-2}{k} \quad \cdots$$

Since

$$\mathbb{Q} = \bigcup_{k=1}^{\infty} A_k = A_1 \cup A_2 \cup A_3 \cup \cdots$$

we can use Proposition 1.5.8 to conclude that $\mathbb{Q}$ is countable. □

With all of this in hand, it is natural to ask whether uncountable sets exist. Cantor answered this question by by proving the following amazing result.

Theorem 1.5.10 (Cantor)**.** $\mathbb{R}$ *is uncountable.*

At first, it appears difficult to prove such a result. We need to argue that there is no possible way to put the real numbers into an infinite list. After seeing the above clever ways of listing $\mathbb{Z}$ and $\mathbb{Q}$, how could we possibly rule out every conceivable approach to forming such a list. Thinking through the logic behind such a task, we need to take an arbitrary list $r_1, r_2, r_3, \ldots$ of real numbers, and somehow argue that there exists a real number c that is missing from the list, i.e. such that $c \neq r_n$ for all $n \in \mathbb{N}^+$.

Suppose then that we have an arbitrary list $r_1, r_2, r_3, \ldots$ of real numbers. The idea is to use the Nested Interval Theorem to produce a suitable value of c. With this in mind, we will create a sequence $I_n = [a_n, b_n]$ of closed bounded intervals with $a_n < b_n$, and we will build the intervals in such a way that $r_n \notin [a_n, b_n]$ for all $n \in \mathbb{N}^+$. In other words, we will dodge the elements of the list $r_1, r_2, r_3, \ldots$ one by one as we continue to create new intervals. Since we only have to avoid one real at a time, we can just shrink a given interval to avoid the next element of the list. We then use the Nested Interval Theorem to get an element that is in each I_n, and hence must be distinct from each r_n.

With the intuition in mind, the only thing we need to do is to convince ourselves that we can take an interval $[a, b]$, and find two disjoint closed intervals inside of it (so that we pick one that does contain the next value of the list). Although it is intuitively clear that it is possible to do this, we can outline a systemic

process to build two such subintervals. Suppose that we have $a, b \in \mathbb{R}$ with $a < b$. We then know from the homework that

$$a < \frac{a+b}{2} < b.$$

We can apply this result again to conclude that

$$a < \frac{a + \frac{a+b}{2}}{2}.$$

Now the right-hand side simplifies to $\frac{3a+b}{4}$. If we instead average the midpoint with b, we obtain $\frac{a+3b}{4}$, and hence we have

$$a < \frac{3a+b}{4} < \frac{a+3b}{4} < b.$$

In other words, within the interval $[a, b]$, we can always form the two disjoint subintervals $[a, \frac{3a+b}{4}]$ and $[\frac{a+3b}{4}, b]$, which are just the first and last quarter of the interval in question. We now turn to the proof.

Proof of 1.5.10. Let $r_1, r_2, r_3, \ldots$ be an arbitrary listing of real numbers. We show that there exists $c \in \mathbb{R}$ with $c \neq r_i$ for all i. To do this, we define two sequences a_n and b_n recursively with the following properties:

- $a_0 \leq a_1 \leq a_2 \leq \ldots$
- $b_0 \geq b_1 \geq b_2 \geq \ldots$
- $a_n < b_n$ for all n.
- $r_n \notin [a_n, b_n]$ for all n.

The construction is as follows. We start be letting $a_0 = 0$ and $b_0 = 1$. Now assume that we have defined a_n and b_n with the above properties. We now have three cases:

- *Case 1:* If $r_{n+1} \notin [a_n, b_n]$, then we define $a_{n+1} = a_n$ and $b_{n+1} = b_n$.
- *Case 2:* If $a_n \leq r_{n+1} \leq \frac{3a_n+b_n}{4}$, then we define $a_{n+1} = \frac{a_n+3b_n}{4}$ and $b_{n+1} = b_n$.
- *Case 3:* Otherwise, i.e. if $\frac{3a_n+b_n}{4} < r_{n+1} \leq b_n$, then we define $a_{n+1} = a_n$ and $b_{n+1} = \frac{a_n+3b_n}{4}$.

Notice that in all cases we have $a_n \leq a_{n+1} < b_{n+1} \leq b_n$, and $r_{n+1} \notin [a_{n+1}, b_{n+1}]$. This completes the recursive construction of a_n and b_n.

By the Nested Interval Theorem, we can fix $c \in \mathbb{R}$ such that $c \in [a_n, b_n]$ for all n. Since $r_n \notin [a_n, b_n]$ for all $n \in \mathbb{N}^+$, we conclude that $c \neq r_n$ for all $n \in \mathbb{N}^+$. Thus, c is real number that does not appear in the list $r_1, r_2, r_3, \ldots$. □

Notice that we arbitrarily chose $a_0 = 0$ and $b_0 = 1$ in the above proof. We could have started with any values of a_0 and b_0 with $a_0 < b_0$, and it would not have affected anything in the proof. We can use that flexibility to prove the much stronger result that any nonempty open interval is uncountable.

Corollary 1.5.11. *For any $c, d \in \mathbb{R}$ with $c < d$, the open interval $(c, d) = \{x \in \mathbb{R} : c < x < d\}$ is uncountable.*

Proof. Follow the proof of Theorem 1.5.10, but now start by letting $a_0 = \frac{3c+d}{4}$ and $b_0 = \frac{c+3d}{4}$. □

It is also possible (and a good exercise!) to explicitly construct a bijection $f : \mathbb{R} \to (c, d)$, which is another way to prove that the open interval (c, d) is uncountable.

Recall from Theorem 1.4.8 that between any two real numbers, we can always find a rational. Can we also find an irrational between any two real numbers? We can use our work on countability to immediately answer this question affirmatively.

Corollary 1.5.12. *For any $c, d \in \mathbb{R}$ with $c < d$, there exists $z \in \mathbb{R} \backslash \mathbb{Q}$ with $c < z < d$.*

Proof. Using Corollary 1.5.11, we know that (c, d) is uncountable. However, the set $\mathbb{Q} \cap (c, d)$ is countable because it is a subset of the countable set $\mathbb{Q}$. Therefore, $(c, d) \backslash \mathbb{Q}$ must be nonempty. □

1.6 The Absolute Value Function

The following function should be familiar from Calculus.

Definition 1.6.1. *Define a function $f : \mathbb{R} \to \mathbb{R}$ by letting*

$$f(a) = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{otherwise.} \end{cases}$$

We call f the absolute value function*, and we use $|a|$ to denote $f(a)$.*

In Calculus, this function is often used as a simple example of a continuous function that is not differentiable at 0, but it is often not used widely beyond that. However, we will make extensive use of the absolute value function. One reason why it plays such an important role is that $|a-b|$ expresses the distance between the points a and b. We will spend this section establishing some fundamental inequalities involving absolute values that we will use throughout our study. We begin with a simple result.

Lemma 1.6.2. *For all $a \in \mathbb{R}$, we have both $-a \leq |a|$ and also $a \leq |a|$.*

Proof. Let $a \in \mathbb{R}$ be arbitrary. We have two cases:

- *Case 1:* Suppose that $a \geq 0$. We then have $-a \leq 0 \leq |a|$, so $-a \leq |a|$. Also, since $|a| = a$, we trivially have $a \leq |a|$.
- *Case 2:* Suppose that $a < 0$. We then have $a < 0 \leq |a|$, so $a \leq |a|$. Also, since $|a| = -a$, we trivially have $-a \leq |a|$.

Therefore, in either case, we have both $-a \leq |a|$ and also $a \leq |a|$. □

Our next result says that we can turn a simple inequality involving an absolute value into a two inequalities.

Proposition 1.6.3. *For all $a \in \mathbb{R}$ and $\varepsilon > 0$, we have $|a| < \varepsilon$ if and only if $-\varepsilon < a < \varepsilon$. Similarly, for all $a \in \mathbb{R}$ and all $r \geq 0$, we have $|a| \leq r$ if and only if $-r \leq a \leq r$*

Proof. Let $a \in \mathbb{R}$ be arbitrary.

- Suppose first that $|a| < \varepsilon$. By Lemma 1.6.2, we have $a \leq |a|$, so $a < \varepsilon$. Since $-a \leq |a|$ by Lemma 1.6.2 again, we also have $-a < \varepsilon$. Multiplying both sides by -1, we conclude that $-\varepsilon < a$. Therefore, $-\varepsilon < a < \varepsilon$.
- Suppose conversely that $-\varepsilon < a < \varepsilon$. We have two cases:
  - *Case 1:* Suppose that $a \geq 0$. We then have $|a| = a < \varepsilon$.
  - *Case 2:* Suppose that $a < 0$. Since $-\varepsilon < a$, we can multiply both sides by -1 to conclude that $-a < \varepsilon$. We then have $|a| = -a < \varepsilon$.

  Thus, in either case, we have $|a| < \varepsilon$.

For the last claim, we can simply use what we just proved, and handle the equality cases and the $r = 0$ case separately. Alternatively, we can just mimic the above proof with $\leq$ everywhere in place of $<$. The details are left as an exercise. □

Although the next result might seem like a very elaborate way to express the equality between two numbers, we will find some uses for it in the next chapter.

Proposition 1.6.4. *Let $a, b \in \mathbb{R}$. If $|a - b| < \varepsilon$ for all $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$, then $a = b$.*

Proof. We prove the contrapositive. That is, we show that if $a \neq b$, then there exists $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$ such that $|a - b| \geq \varepsilon$. Suppose then that $a \neq b$. We then have $a - b \neq 0$, so $|a - b| \neq 0$ (since if $-x = 0$, then $x = 0$). Notice then that $|a - b| > 0$, so we can take $\varepsilon = |a - b|$. □

There are number of other basic properties of the absolute value function that we list here for reference. As usual, they can be proven using a straightforward case analysis.

Proposition 1.6.5.

1. *For all $a \in \mathbb{R}$, we have $|-a| = |a|$.*
2. *For all $a \in \mathbb{R}$, we have $|a|^2 = a^2$.*
3. *For all $a, b \in \mathbb{R}$, we have $|ab| = |a| \cdot |b|$.*

Proof. Exercise. □

Our next result is one of the most widely used inequalities in Analysis.

Theorem 1.6.6 (Triangle Inequality)**.** *For all $a, b \in \mathbb{R}$, we have $|a + b| \leq |a| + |b|$.*

In the above results, we handled proofs involving the absolute value function by breaking into cases based on the sign of the input We have three different inputs here. So it might seem that a natural approach would be to consider 8 different cases, based on the signs of a, b, and $a + b$. Now a couple of these are unnecessary (if $a \geq 0$ and $b \geq 0$, then $a + b \geq 0$), but some of the necessary cases end up having some subcases. To avoid such a messy argument, we employ a bit of cleverness by only doing cases on the sign of $a + b$, and then leaning heavily on Lemma 1.6.2.

Proof. Let $a, b \in \mathbb{R}$ be arbitrary. We have two cases:

- *Case 1:* Suppose that $a + b \geq 0$. We then have

$$\begin{aligned} |a + b| &= a + b \\ &\leq |a| + b && \text{(by Lemma 1.6.2)} \\ &\leq |a| + |b| && \text{(by Lemma 1.6.2).} \end{aligned}$$

- *Case 2:* Suppose that $a + b < 0$. We then have

$$\begin{aligned} |a + b| &= -(a + b) \\ &= (-a) + (-b) \\ &\leq |a| + (-b) && \text{(by Lemma 1.6.2)} \\ &\leq |a| + |b| && \text{(by Lemma 1.6.2).} \end{aligned}$$

Therefore, in either case, we have $|a + b| \leq |a| + |b|$. □

What about if we use $-$ rather than $+$? It is straightforward to see that we do *not* in general have $|a - b| \leq |a| - |b|$, because the left side is always nonnegative while the right-hand side is negative when $|b| > |a|$. How about the reverse inequality? Would we expect $|a| - |b| \leq |a - b|$ to always hold? Now we get some easy wins, because the left-hand side is sometimes negative. Surprisingly, if we force the left-hand side to be positive by taking an absolute value of it, then we still have a valid inequality. Even more surprisingly, we can obtain a relatively straightforward proof by using the Triangle Inequality.

Proposition 1.6.7. *For all $a, b \in \mathbb{R}$, we have $||a| - |b|| \leq |a - b|$.*

Proof. Let $a, b \in \mathbb{R}$ be arbitrary. Notice that

$$\begin{aligned} |a| &= |a - b + b| \\ &\leq |a - b| + |b| && \text{(by the Triangle Inequality).} \end{aligned}$$

Adding $-|b|$ to both sides, it follows that $|a| - |b| \leq |a - b|$. Now we also have

$$\begin{aligned} |b| &= |b - a + a| \\ &\leq |b - a| + |a| && \text{(by the Triangle Inequality)} \\ &= |-(b - a)| + |a| && \text{(by Proposition 1.6.5)} \\ &= |a - b| + |a|. \end{aligned}$$

Adding $-|a - b| - |b|$ to both sides, it follows that $-|a - b| \leq |a| - |b|$. Putting these inequalities together, we conclude that

$$-|a - b| \leq |a| - |b| \leq |a - b|.$$

Using Proposition 1.6.3, it follows that $||a| - |b|| \leq |a - b|$. □

Corollary 1.6.8. *For all $a, b \in \mathbb{R}$, we have $|a| - |b| \leq |a - b|$.*

Proof. Immediate from Lemma 1.6.2 and Proposition 1.6.7. □

Before moving on, we use the inequalities in this section to establish a couple of results. Although they are interesting on their own, their proofs illustrate the kinds of arguments that we will employ often, but in a simpler setting. Suppose that we have two numbers $x_0, y_0 \in \mathbb{R}$. Suppose also that we have two other numbers $x, y \in \mathbb{R}$ that we think of as being "close" to x_0 and y_0, respectively. Do we expect $x + y$ to be "close" to $x_0 + y_0$, and do we expect xy to be "close" to $x_0 y_0$?

To give reasonable answers to these questions, we need to codify what we mean by "close". Suppose that we have two error ranges coded by numbers ε_x and ε_y. Notice that $|x - x_0| < \varepsilon_x$ is a simple way to say x is within ε_x of x_0. So suppose that $|x - x_0| < \varepsilon_x$ and $|y - y_0| < \varepsilon_y$. Now under these assumption, we want to answer the question of how close $x + y$ is to $x_0 + y_0$, which is to say that we want to establish a good upper bound for $|(x + y) - (x_0 + y_0)|$.

For example, suppose that $x_0 = 4$, $y_0 = 7$, $\varepsilon_x = .01$, and $\varepsilon_y = .02$. Now if $|x - x_0| < \varepsilon_x$ and $|y - y_0| < \varepsilon_y$, then $3.99 < x < 4.01$ and $6.98 < y < 7.02$. When we think about adding the values of x and y, if both numbers are close to the right-hand endpoints of these open intervals, then the errors will add. In particular, notice that $4.01 + 7.02 = 11.03$, which is $.03 = \varepsilon_x + \varepsilon_y$ away (of course, the right-hand endpoint 4.01 is not a possible value of x, but it does serve as an upper bound).

Multiplication is more interesting. Can we get a good upper bound on $|xy - x_0 y_0|$ in terms of ε_x and ε_y? Suppose we have the same values as the previous paragraph. Notice that

$$\begin{aligned} 4.01 \cdot 7.02 &= (4 + .01) \cdot (7 + .02) \\ &= 28 + 4 \cdot .02 + 7 \cdot .01 + .01 \cdot .02 \\ &= 28.1502 \end{aligned}$$

Pay more attention to the line before the final value. When thinking about values near the end of the intervals, we see that the values of $x_0 = 4$ and $y_0 = 7$ appear in the error estimate. If we think about keeping $\varepsilon_x = .01$ and $\varepsilon_y = .02$, but making x and y very large, then the resulting error between xy and $x_0 y_0$ seems like it could get much larger. In particular, it appears that we can not express the error estimate on $|xy - x_0 y_0|$ in terms of ε_x and ε_y alone. Instead, we are forced to include x_0 and y_0 in the final error estimate. Seeing the distributive law in action in the above example, we can generalize to the following result.

Proposition 1.6.9. *Let $x_0, y_0, x, y, \varepsilon_x, \varepsilon_y \in \mathbb{R}$ with both $\varepsilon_x > 0$ and $\varepsilon_y > 0$.*

1. *If $|x - x_0| < \varepsilon_x$ and $|y - y_0| < \varepsilon_y$, then $|(x + y) - (x_0 + y_0)| < \varepsilon_x + \varepsilon_y$.*
2. *If $|x - x_0| < \varepsilon_x$ and $|y - y_0| < \varepsilon_y$, then $|x \cdot y - x_0 \cdot y_0| < |x_0| \cdot \varepsilon_y + |y_0| \cdot \varepsilon_x + \varepsilon_x \cdot \varepsilon_y$.*

Proof.

1. We have

$$\begin{aligned} |(x + y) - (x_0 + y_0)| &= |(x - x_0) + (y - y_0)| \\ &\leq |x - x_0| + |y - y_0| && \text{(by the Triangle Inequality)} \\ &< \varepsilon_x + \varepsilon_y. \end{aligned}$$

2. First notice that

$$\begin{aligned} |x| - |x_0| &\leq |x - x_0| && \text{(by Corollary 1.6.8)} \\ &< \varepsilon_x. \end{aligned}$$

Adding $|x_0|$ to both sides, it follows that $|x| < |x_0| + \varepsilon_x$. Therefore,

$$\begin{aligned} |x \cdot y - x_0 \cdot y_0| &= |x \cdot y - x \cdot y_0 + x \cdot y_0 - x_0 \cdot y_0| \\ &\leq |x \cdot y - x \cdot y_0| + |x \cdot y_0 - x_0 \cdot y_0| && \text{(by the Triangle Inequality)} \\ &= |x \cdot (y - y_0)| + |y_0 \cdot (x - x_0)| \\ &= |x| \cdot |y - y_0| + |y_0| \cdot |x - x_0| && \text{(by Proposition 1.6.5)} \\ &\leq |x| \cdot \varepsilon_y + |y_0| \cdot \varepsilon_x \\ &< (|x_0| + \varepsilon_x) \cdot \varepsilon_y + |y_0| \cdot \varepsilon_x && \text{(from above)} \\ &= |x_0| \cdot \varepsilon_y + |y_0| \cdot \varepsilon_x + \varepsilon_x \cdot \varepsilon_y. \end{aligned}$$

□

We pause to notice a couple of interesting details in the proof. First, we inserted some terms into $|x \cdot y - x_0 \cdot y_0|$ without affecting the value so that we could use the estimates given in the assumptions. This is a clever trick that is common in Analysis. There is another subtle point in the string of inequalities. When going form the fourth line to the fifth in the second part of the argument, we used $\leq$ rather $<$ despite the fact that $|x - x_0| < \varepsilon_x$ and $|y - y_0| < \varepsilon_y$. Why? Thinking through the logic, to use these inequalities in this step, we first have to multiply the former by $|y_0|$ and the latter by $|x|$. Now both of these values are nonnegative, but they might be 0. In other words, we only know that $|y_0| \geq 0$ and $|x| \geq 0$. Thus, when we multiply both sides of $|x - x_0| < \varepsilon_x$ by $|y_0| \geq 0$, we only obtain the weak inequality, not the strong one. However, in the step from the fifth to the sixth line, we do get a strong inequality, because we know that both $|x| < |x_0| + \varepsilon_x$ and that $\varepsilon_y > 0$, so we know that $|x| \cdot \varepsilon_y < (|x_0| + \varepsilon_x) \cdot \varepsilon_y$, and hence that

$$|x| \cdot \varepsilon_y + |y_0| \cdot \varepsilon_x < (|x_0| + \varepsilon_x) \cdot \varepsilon_y + |y_0| \cdot \varepsilon_x.$$

in order to employ our assumptions to estimate $|x \cdot y - x_0 \cdot y_0|$, we used a clever trick of adding and subtracting a common value.

Finally, we might ask a similar question about division. In the simplest case, this is just taking reciprocals. So assume that we know that $|y - y_0| < \varepsilon_y$, and that y and y_0 are both nonzero. How close must y be to y_0? In general, the answer is not at all. The primary issue is that y and y_0 might be on opposite sides of 0. For example, consider the case where $y_0 = .00001$ and $y = -.00001$. Obviously y and y_0 are close, but their reciprocals are very far away from each other. We could get some traction by forcing y_0 and y to have the same sign. However, to avoid too long a digression before getting to the heart of Analysis, we avoid a full exploration of these ideas until we need them later.

Chapter 2

Sequences of Real Numbers

Calculus is based around the idea of approximation. The recurring hope is that we can make a precise definition of a subtle concept by employing simpler approximations and taking some kind of "limit". For instance, think back about how we define the derivative of a function f at a point. We want to capture the intuitive but elusive idea of the slope of a tangent line at the point. To get around the fact that we only have one point, we begin by approximating the answer that we seek by first by using the slope of a secant lines (which is easier because we get two points), then we refine the approximation by bringing the second point closer to the original point, and then we finally take a limit. Also, think back about how we defined the integral of a (nonnegative) function f on an interval $[a, b]$. We want to capture the intuitive but elusive idea of the area bounded by the curve and the x-axis. Since it is difficult to find the area of a curved shape, we instead approximate the area by using rectangles, then we refine the approximations by making the rectangles thinner, and then we finally take a limit.

Instead of beginning our study of Analysis with derivatives and integrals, we will first explore infinite sequences of real numbers. An infinite sequence is a fancy name for a big long list. For instance, $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ is an infinite sequence of real numbers. Although you likely have some experience with such sequences, a careful study reveals many subtleties and surprises. Furthermore, whereas you have already learned techniques for calculating derivatives and integrals of many reasonable functions, you probably only feel comfortable with very "nice" sequences. Instead of being a hindrance, this discomfort can be helpful, because you likely have fewer preconceived notions about how sequences behave. Moreover, as we will see, sequences are more fundamental to Analysis than derivatives and integrals are. One of our first tasks will be to understand how to carefully define the notion of a limit in this context. For the example of $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ above, although this sequence never actually reaches 0, it's pretty clear that it is somehow "approaching" zero, and our first task is to make this idea of "approaching" precise.

How is a study of sequences and their limiting behavior interesting and useful to us? We will see many applications of sequences in time, but here is one very simple example where they arise. Suppose we have a mathematical object that we want to understand. If the object is very complicated, perhaps we can understand it better if we can approximate it by a sequence of simpler objects that "approaches" the one we care about. For instance, if our mathematical object is a particular real number r (think of π), perhaps we can find a sequence of simple rationals numbers that approach r, and somehow combine knowledge of the rational approximations and the properties of "limits" of sequences to get information about r. As we will see, this is actually what a decimal expansion of a real number is all about, because if we truncate an infinite decimal expansion, we get a rational number. In fact, sequences will provide a way to make decimal expansions precise.

However, we will eventually see that the real power of sequences comes not from approximating a real number by a sequence of simpler numbers, but instead by approximating a function by a sequence of simpler functions. If you have encountered Taylor series before, then this idea should feel familiar. In that context,

we attempt to approximate a complicated function by a sequence of polynomials. If we can make sense of the "limit" of this sequence of polynomials, perhaps we can use our knowledge of polynomials (how to compute their values, how to differentiate them, how to integrate them, etc.) and our soon-to-be accumulated knowledge of limits of sequences to perform the same tasks on f. In the case of Fourier's question, rather that using polynomials as our "simple" objects, we are trying to approximate a complicated by finite linear combinations of sines and cosines.

Another intriguing possibility is to run this idea backwards and use simpler functions (like polynomials or simple trigonometric functions) and limits to *define* new and interesting functions that resist definitions in other ways.

2.1 Definitions and Basic Results

We start with the formal definition of an infinite sequence of real numbers.

Definition 2.1.1. *An* infinite sequence of real numbers*, often just abbreviated* sequence *when there is no ambiguity, is a function* $f\colon \mathbb{N}^+ \to \mathbb{R}$.

Of course, we don't usually think of a sequence as a function. We think of it as an infinite list of numbers. Thus, instead of using the notation of functions (and thus refer to the fifth element of a sequence f as $f(5)$), we will adopt special notation when discussing sequences. Since we think of a sequence as a list of numbers, perhaps notation such as $a_1, a_2, a_3, \dots$ be appropriate. Since repetition is allowed, i.e. we are not assuming that f is injective, it is natural to think of a sequence as an infinite tuple of numbers. As such, notation such as $(a_1, a_2, a_3, \dots)$ would be reasonable. However, regular parentheses are overused in mathematics, so we will use angle brackets like $\langle a_1, a_2, a_3, \dots \rangle$ instead. Of course, it is tiresome to write all of those symbols whenever we want to refer to a sequence, so we instead simply write $\langle a_n \rangle$ to refer to a generic sequence. For a particular example, we can refer to the sequence $\langle 1, \frac{1}{2}, \frac{1}{3}, \dots \rangle$ by just writing $\langle \frac{1}{n} \rangle$. With the general notation $\langle a_n \rangle$ for a sequence, we can then thereafter use a_n to denote the n^{th} element of the sequence.

Before moving on, you should know that other sources use different notation for sequences. For example, since the notation $\langle a_n \rangle$ might give the impression that there is one one "term" a_n in the sequence, it is also common to denote a sequence by $\langle a_n \rangle_{n=1}^{\infty}$. Truth be told, this is much better notation, but it becomes tiresome to write it constantly. Some authors use (a_n) or the more adorned $(a_n)_{n=1}^{\infty}$, but as much mentioned above, parentheses are overused in mathematics and (a_n) alone looks a bit strange. Another common notation is $\{a_n\}$, or the more adorned $\{a_n\}_{n=1}^{\infty}$, but I find that notation confusing and unnatural because the set notation $\{\dots\}$ suggests the order of elements does not matter.

Sequences arise in many ways. Of course, the most straightforward way to define a sequence is by giving a formula for the n^{th} term a_n. For example, we can let $a_n = \frac{1}{n}$ to obtain the sequence above, or we can consider $a_n = \frac{n^2-5n+1}{2n^2+7}$. Recursion is another common way to define a sequence. For example, on the first homework, we defined a sequence a_n by letting $a_1 = 0$ and then defining $a_{n+1} = \frac{1}{3}(a_n + 1)$ for all $n \in \mathbb{N}^+$.

We can also describe a sequence somewhat indirectly. Here are two examples:

- For each $n \in \mathbb{N}^+$, let a_n be the number of primes less than or equal to n. For example, we have $a_{10} = 4$, $a_{11} = 5$, $a_{12} = 5$, etc.
- For each $n \in \mathbb{N}^+$, let b_n be the number of subsets of $\{1, 2, 3, \dots, n\}$ with 3 elements. For example, we have $b_2 = 0$, $b_3 = 1$, $b_4 = 4$, $b_5 = 10$, etc.

Although you may have learned a simple formula for b_n (it turns out that $b_n = \frac{n(n-1)(n-2)}{6}$), there is no simple formula for a_n.

We are interested in the long term behavior of sequences. That is, we don't care about the "early" terms of the sequence. And although trying to find a nice formula for the n^{th} term of a sequence can be a very

worthwhile endeavor, we care more about what happens over the long run. Consider the example where a_n is the number of primes less than or equal to n. We know that a_n gets arbitrarily large because there are infinitely many primes. But how fast does it grow? Over the long run, does it eventually grow faster or slower than $\frac{n}{100}$?

We will eventually be able to make all of these questions precise by using the notion of a limit of a sequence. Let's return to our simple example of the sequence $\langle \frac{1}{n} \rangle$, i.e. the sequence of numbers $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$. The first thing to notice is that this sequence becomes strictly smaller and smaller as we proceed onward. More formally, if $m < n$, then $\frac{1}{m} > \frac{1}{n}$. Although the sequence does proceed downward in this way, it does not get arbitrarily small. For example, it never dips below the value 0, let alone below the value -4. However, it does seem to marching ever toward the number 0, with the caveat that it never actually reaches this endpoint.

We need to give a precise definition for "the sequence $\langle a_n \rangle$ approaches ℓ". Of course, the key word here is *approaches*. A sequence can approach a number ℓ despite the fact that no term actually equals ℓ, as the example $\langle \frac{1}{n} \rangle$ illustrates. Intuitively, we want to say that $\langle a_n \rangle$ approaches ℓ if we can get arbitrarily close to ℓ, provided we only look at terms sufficiently far out in the sequence. But how do we turn the *arbitrarily close* and *sufficiently far out* into precise mathematical language?

The key idea is to think about an evil adversary handing you a very small error $\varepsilon > 0$. If you are claiming that $\langle a_n \rangle$ is approaching ℓ, then sufficiently far out terms of the sequence should be within ε of ℓ. To convince the evil adversary that this does eventually happen, you need to respond with a value $N \in \mathbb{N}^+$ such that every term of the sequence starting from the N^{th} term onward is indeed within ε of ℓ. For example, maybe the evil adversary hands you a small error $\varepsilon = \frac{1}{666}$. You respond with $N = 3000$, and then successfully convince them that for every $n \geq 3000$, the term a_n is within $\frac{1}{666}$ of ℓ. Bah, lucky!, responds the evil adversary. They then challenge you with a small error $\varepsilon = \frac{1}{6666}$. After some hard work, you respond with $N = 748141$, and then successfully convince them that for every $n \geq 748141$, the term a_n is within $\frac{1}{6666}$ of ℓ. Undeterred, the evil adversary picks an even smaller $\varepsilon > 0$...

Now if you are claiming that $\langle a_n \rangle$ approaches ℓ, then you should be able to handle *every* challenge from the evil adversary. That is, no matter what value of $\varepsilon > 0$ is given to you, you should be able to respond with some $N \in \mathbb{N}^+$, and be able to demonstrate that for every $n \geq N$, the term a_n is within ε of ℓ. Our last step in turning this idea into a precise definition is to replace the phrase "a_n is within ε of ℓ" by the more precise $|a_n - \ell| < \varepsilon$. With that in mind, we adopt the following definition. We use the word *converges* rather than *approaches* because we want to sound sophisticated.

Definition 2.1.2. *Let $\langle a_n \rangle$ be a sequence and let $\ell \in \mathbb{R}$. We say that $\langle a_n \rangle$* converges to ℓ *if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, we have $|a_n - \ell| < \varepsilon$.*

Notice that the above definition has three quantifiers, and they alternate from *for all*, to *there exists*, and then back to *for all*. Most of the challenge in working with the definition lies in complexity of this alternation of quantifiers. In pure symbols, we can write the definition of "$\langle a_n \rangle$ converges to ℓ" as

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N}^+)(\forall n \geq N)[|a_n - \ell| < \varepsilon].$$

Although it is useful to examine this symbolic statement carefully and think about it often (as it clearly lays out the quantifiers), it is better to avoid excessive symbolism within proofs in the interest of human readability.

Given a sequence $\langle a_n \rangle$ and a number ℓ, let's outline the structure of a direct proof of the statement that $\langle a_n \rangle$ converges to ℓ.

- Since the first quantifier is *for all*, start by taking an arbitrary $\varepsilon > 0$. You have no control over it! As illustrated above, you can think about it as being handed to you by an evil adversary. And you should typically think about it as being really really tiny.

- Since the next quantifier is *there exists*, respond to the challenge ε by building a particular N. You have complete control over the construction of N! However, you will need to argue that your N has special properties related to the ε chosen by the enemy (see the next step). Typically, when responding to ever smaller ε, the N that you will answer with will become larger and larger. Note that you do *not* have to explain the process you used to find the N that you choose.

- Argue that your N works. To do this, start by taking an arbitrary $n \geq N$ (since the third quantifier is *for all*), and then prove that $|a_n - \ell| < \varepsilon$.

For example, consider the sequence $\langle \frac{1}{n} \rangle$, i.e. the sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$. Suppose that we want to argue that $\langle \frac{1}{n} \rangle$ converges to 0. Before jumping into the general argument, let's examine a few special cases. When challenged with $\varepsilon = 1$, we can take $N = 2$. Why? For any $n \geq 2$, we have $n > 1$, so $\frac{1}{n} < 1$, and hence $|\frac{1}{n} - 0| = \frac{1}{n} < 1$. What if we are challenged with $\varepsilon = \frac{1}{10}$? We claim that $N = 11$ works. To see this, notice that for any $n \geq 11$, we have $n > 10$, so $\frac{1}{n} < \frac{1}{10}$, and hence $|\frac{1}{n} - 0| = \frac{1}{n} < \frac{1}{10}$. How do we handle a general $\varepsilon > 0$, especially one that is not of the form $\frac{1}{k}$ for some $k \in \mathbb{N}^+$? We simply need to pick an $N \in \mathbb{N}^+$ with $N > \frac{1}{\varepsilon}$, as we now demonstrate.

Proposition 2.1.3. *The sequence $\langle \frac{1}{n} \rangle$ converges to 0.*

Proof. Let $\varepsilon > 0$ be arbitrary. By Proposition 1.4.5, we can fix $N \in \mathbb{N}$ with $N > \frac{1}{\varepsilon}$. Now let $n \in \mathbb{N}^+$ with $n \geq N$ be arbitrary. We then have $n > \frac{1}{\varepsilon}$. Multiplying both sides of this inequality by $\frac{\varepsilon}{n} > 0$, we conclude that $\frac{1}{n} < \varepsilon$, so

$$\begin{aligned} \left| \frac{1}{n} - 0 \right| &= \frac{1}{n} \\ &\leq \frac{1}{N} \\ &< \varepsilon. \end{aligned}$$

Therefore, $\langle \frac{1}{n} \rangle$ converges to 0. □

We can also prove the intuitively obvious claim that a constant sequence converges to that constant value.

Proposition 2.1.4. *Let $c \in \mathbb{R}$ and let $a_n = c$ for all $n \in \mathbb{N}^+$. We then have that $\langle a_n \rangle$ converges to c.*

Proof. Let $\varepsilon > 0$ be arbitrary. Consider $N = 1$ (any element of $\mathbb{N}$ will do). For any $n \in \mathbb{N}^+$ with $n \geq N$, we have

$$\begin{aligned} |a_n - c| &= |c - c| \\ &= 0 \\ &< \varepsilon. \end{aligned}$$

Therefore, $\langle a_n \rangle$ converges to c. □

For a more interesting example, let's show that $\langle \frac{\sqrt{n} + \sin n}{\sqrt{n}} \rangle$ converges to 1. Let $\varepsilon > 0$ be arbitrary. By Proposition 1.4.5, we can fix $N \in \mathbb{N}^+$ with $N > \frac{1}{\varepsilon^2}$. Now let $n \in \mathbb{N}^+$ with $n \geq N$ be arbitrary. We then have $n > \frac{1}{\varepsilon^2}$. Using the contrapositive Problem 5c on Homework 1, it follows that $\sqrt{n} > \frac{1}{\varepsilon}$ (if $\sqrt{n} \leq \frac{1}{\varepsilon}$, then we could square both sides to conclude that $n \leq \frac{1}{\varepsilon^2}$). Multiplying both sides by $\frac{\varepsilon}{\sqrt{n}} > 0$ we conclude that

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

$$\begin{aligned}\left|\frac{\sqrt{n}+\sin n}{\sqrt{n}} - 1\right| &= \left|1 + \frac{\sin n}{\sqrt{n}} - 1\right| \\ &= \left|\frac{\sin n}{\sqrt{n}}\right| \\ &= \frac{|\sin n|}{\sqrt{n}} \\ &\leq \frac{1}{\sqrt{n}} \\ &< \varepsilon.\end{aligned}$$

Therefore, $\langle \frac{\sqrt{n}+\sin n}{\sqrt{n}} \rangle$ converges to 1.

Suppose that we want to argue that a sequence $\langle a_n \rangle$ does *not* converge to a given number ℓ. We can naturally form the negation symbolically as

$$\neg(\forall \varepsilon > 0)(\exists N \in \mathbb{N}^+)(\forall n \geq N)[|a_n - \ell| < \varepsilon].$$

Now we can push the negation inside one quantifier at a time, while flipping the quantifiers. Thus, the statement "$\langle a_n \rangle$ does not converge to ℓ" becomes

$$(\exists \varepsilon > 0)(\forall N \in \mathbb{N}^+)(\exists n \geq N)[|a_n - \ell| \geq \varepsilon].$$

In words, we can express the statement "$\langle a_n \rangle$ does not converge to ℓ" as saying that there exists $\varepsilon > 0$ such that for all $N \in \mathbb{N}^+$, there exists $n \geq N$ such that $|a_n - \ell| \geq \varepsilon$.

Let's use this to argue that $\langle \frac{1}{n} \rangle$ does not converge to $\frac{1}{2}$. In this case, the first quantifier is a *there exists* quantifier, so we have to pick an $\varepsilon > 0$ that has the required property. In other words, we get to play the role of the evil adversary. Intuitively, as long as we pick some $\varepsilon < \frac{1}{2}$, then eventually the sequences will be more than ε away from $\frac{1}{2}$. But we just need to pick *one* such value of $\varepsilon > 0$ and argue that it works. A natural example is $\varepsilon = \frac{1}{4}$.

So consider $\varepsilon = \frac{1}{4}$. Let $N \in \mathbb{N}^+$ be arbitrary. Next let $n = \max\{N, 8\}$. Notice that we have $n \geq N$. We also have $n \geq 8$, so $\frac{1}{n} \leq \frac{1}{8}$, and hence $\frac{1}{n} - \frac{1}{8} \leq 0$. Therefore

$$\begin{aligned}\left|\frac{1}{n} - \frac{1}{2}\right| &= -\left(\frac{1}{n} - \frac{1}{2}\right) \\ &= \frac{1}{2} - \frac{1}{n} \\ &\geq \frac{1}{2} - \frac{1}{8} \\ &= \frac{3}{8} \\ &> \varepsilon.\end{aligned}$$

Notice in the above argument that we took $n = \max\{N, 8\}$. We needed to ensure that $n \geq N$, which explains its appearance. What did we choose 8? In fact we could have chosen 5 here, be the arithmetic was easier with a slightly larger n, so why not make the argument a little cleaner?

Our next proposition says that a sequence can not converge to two different limits.

Proposition 2.1.5. *Let $\langle a_n \rangle$ be a sequence and let $\ell, m \in \mathbb{R}$. If $\langle a_n \rangle$ converges to both ℓ and m, then $\ell = m$.*

We give two proofs. The first is by contradiction. Suppose that $\ell \neq m$, and assume that $\ell < m$, say. Since $\langle a_n \rangle$ converges to ℓ, we know that the terms of the sequence will eventually be as close to ℓ as we would like. Similarly, the terms of the sequence will eventually be as close to m as we would like. But if we make the "windows" around ℓ and m small enough so that they do not overlap, this seems to be problem. How small? We can take the distance between ℓ and m, and divide it by 2. That is, consider $\varepsilon = \frac{m-\ell}{2}$.

Proof 1 of Proposition 2.1.5. Suppose that $\ell \neq m$. Without loss of generality, we may assume that $\ell < m$ (otherwise, switch what we are calling ℓ and m). Notice that $\frac{m-\ell}{2} > 0$. Since $\langle a_n \rangle$ converges to ℓ, we can fix $N_1 \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{m-\ell}{2}$ for all $n \geq N_1$. Since $\langle a_n \rangle$ converges to m, we can fix $N_2 \in \mathbb{N}^+$ such that $|a_n - m| < \frac{m-l}{2}$ for all $n \geq N_2$. Let $n = \max\{N_1, N_2\}$. We then have $n \geq N_1$, so $|a_n - \ell| < \frac{m-\ell}{2}$, hence $a_n - \ell < \frac{m-\ell}{2}$ by Proposition 1.6.3, and thus

$$a_n < \ell + \frac{m-\ell}{2} = \frac{\ell + m}{2}.$$

We also have $n \geq N_2$, so $|a_n - m| < \frac{m-\ell}{2}$, hence $a_n - m > -\frac{m-\ell}{2}$ by Proposition 1.6.3, and thus

$$a_n > m - \frac{m-\ell}{2} = \frac{\ell + m}{2}.$$

It follows that $\frac{\ell+m}{2} < a_n < \frac{\ell+m}{2}$, a contradiction. Therefore, $\ell = m$. □

We can also give a direct proof by showing that $|\ell - m| < \varepsilon$ for all $\varepsilon > 0$. Here is the idea. Suppose we want to show that $|\ell - m| < 1$. Since $\langle a_n \rangle$ converges to ℓ, we can go far enough out so that the terms are within $\frac{1}{2}$ of ℓ. Similarly, since $\langle a_n \rangle$ converges to m, we can go far enough out so that the terms are within $\frac{1}{2}$ of m. Once we go beyond both of these points, we will be able to use the Triangle Inequality to conclude that $|\ell - m| < 1$. Notice that we had to divide 1 in half because the errors might add. In general, if we want to show that $|\ell - m| < \varepsilon$, we should go far enough out so that the terms are within $\frac{\varepsilon}{2}$ of each of ℓ and m. Here is the argument.

Proof 2 of Proposition 2.1.5. We show that $|\ell - m| < \varepsilon$ for all $\varepsilon > 0$. Let $\varepsilon > 0$ be arbitrary. Since $\langle a_n \rangle$ converges to ℓ, we can fix $N_1 \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{\varepsilon}{2}$ for all $n \geq N_1$. Since $\langle a_n \rangle$ converges to m, we can fix $N_2 \in \mathbb{N}^+$ such that $|a_n - m| < \frac{\varepsilon}{2}$ for all $n \geq N_2$. Consider $n = \max\{N_1, N_2\}$. We then have

$$\begin{aligned}
|\ell - m| &= |\ell - a_n + a_n - m| \\
&\leq |\ell - a_n| + |a_n - m| && \text{(by the Triangle Inequality)} \\
&= |a_n - \ell| + |a_n - m| \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } n \geq N_1 \text{ and } n \geq N_2\text{)} \\
&= \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that $|\ell - m| < \varepsilon$ for all $\varepsilon > 0$. Using Proposition 1.6.4, we conclude that $\ell = m$. □

Definition 2.1.6. *Let $\langle a_n \rangle$ be a sequence. We say that $\langle a_n \rangle$* converges *if there exists an $\ell \in \mathbb{R}$ such that $\langle a_n \rangle$ converges to ℓ. Otherwise, we say that $\langle a_n \rangle$* diverges.

Notice that the statement "$\langle a_n \rangle$ converges" adds yet another quantifier to the above definition. That is, when written purely symbolically, the statement "$\langle a_n \rangle$ converges" becomes

$$(\exists \ell \in \mathbb{R})(\forall \varepsilon > 0)(\exists N \in \mathbb{N})(\forall n > N)[|a_n - \ell| < \varepsilon].$$

Let's take a look at the statement "$\langle a_n \rangle$ diverges". We need to negate the above, and push the negation inside one quantifier at a time, while flipping the quantifiers. Thus, the statement "$\langle a_n \rangle$ diverges" becomes

$$(\forall \ell \in \mathbb{R})(\exists \varepsilon > 0)(\forall N \in \mathbb{N})(\exists n > N)[|a_n - \ell| \geq \varepsilon].$$

Here is our first example of a natural sequence that diverges.

Proposition 2.1.7. *The sequence $\langle (-1)^n \rangle$ diverges.*

Proof. Let $\ell \in \mathbb{R}$ be arbitrary. We show that $\langle (-1)^n \rangle$ does not converge to ℓ. In order to do this, we need to show that there exists $\varepsilon > 0$ such that for all $N \in \mathbb{N}$, there exists $n \geq N$ with $|(-1)^n - \ell| \geq \varepsilon$.

- *Case 1:* Suppose that $\ell \geq 0$. Consider $\varepsilon = 1$, and let $N \in \mathbb{N}$ be arbitrary. Fix the odd $n \in \{N, N+1\}$ (note that if N is not odd, then $N+1$ is odd). We then have

$$\begin{aligned} |(-1)^n - \ell| &= |-1-\ell| && \text{(since } n \text{ is odd)} \\ &= -(-1-\ell) && \text{(since } -1-\ell \leq 0 \text{ as } \ell \geq 0) \\ &= 1+\ell \\ &\geq 1 && \text{(since } \ell \geq 0) \\ &= \varepsilon. \end{aligned}$$

- *Case 2:* Suppose that $\ell < 0$. Consider $\varepsilon = 1$, and let $N \in \mathbb{N}$ be arbitrary. Fix the even $n \in \{N, N+1\}$ (note that if N is not even, then $N+1$ is even). We then have

$$\begin{aligned} |(-1)^n - \ell| &= |1-\ell| && \text{(since } n \text{ is even)} \\ &= 1-\ell && \text{(since } 1-\ell \geq 0 \text{ as } \ell < 0) \\ &\geq 1 && \text{(since } \ell < 0 \text{ so } -\ell > 0) \\ &= \varepsilon. \end{aligned}$$

In all cases, we have shown that $\langle (-1)^n \rangle$ does not converge to ℓ. Since $\ell \in \mathbb{R}$ was arbitrary, it follows that $\langle (-1)^n \rangle$ diverges. □

There are other simple examples of sequences that diverge, such as $\langle n \rangle$. Before moving on, we introduce some standard notation.

Notation 2.1.8. *If $\langle a_n \rangle$ is a sequence and $\ell \in \mathbb{R}$, we write $\lim_{n\to\infty} a_n = \ell$ to mean that $\langle a_n \rangle$ converges to ℓ.*

Notice that this notation only makes sense in light if Proposition 2.1.5, which says that a limit is unique (when it exists). However, be careful *not* to write $\lim_{n\to\infty} a_n$ in isolation before knowing a sequence converges, since this notation does not make sense for a general divergent sequence.

2.2 Algebraic and Order-Theoretic Properties of Limits

When we introduced sequences, we said that our focus will be on what happens over the long run. Our next two results express two ways to make precise the idea that the beginnings or "origin" of a sequence has no impact on whether it converges, or what the limit is. We start with a proposition that says that if two sequences agree from some point onward, then they have the same convergence properties.

Proposition 2.2.1. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences, and assume that there exists $M \in \mathbb{N}^+$ such that $a_n = b_n$ for all $n \geq M$.*

1. *If $\langle a_n \rangle$ converges, then $\langle b_n \rangle$ converges, and in this case we have $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.*

2. *If $\langle a_n \rangle$ diverges, then $\langle b_n \rangle$ diverges.*

Proof.

1. Suppose that $\langle a_n \rangle$ converges, and let $\ell = \lim_{n\to\infty} a_n$. By assumption, we can fix $M \in \mathbb{N}^+$ such that $a_n = b_n$ for all $n \geq M$. We show that $\langle b_n \rangle$ converges to ℓ. Let $\varepsilon > 0$ be arbitrary. Since $\lim_{n\to\infty} a_n = \ell$, we can fix $K \in \mathbb{N}^+$ such that $|a_n - \ell| < \varepsilon$ for all $n \geq K$. Let $N = \max\{K, M\}$. For any $n \geq N$, we then have

$$\begin{aligned} |b_n - \ell| &= |a_n - \ell| && \text{(because } a_n = b_n \text{ as } n \geq N \geq M\text{)} \\ &< \varepsilon && \text{(since } n \geq N \geq K\text{)}. \end{aligned}$$

   Therefore, $\lim_{n\to\infty} b_n = \ell$.

2. The contrapositive, i.e. that if $\langle b_n \rangle$ converges then $\langle a_n \rangle$ converges, is immediate from part (1). In more detail, suppose that $\langle b_n \rangle$ converges and fix $\ell \in \mathbb{R}$ with $\lim_{n\to\infty} b_n = \ell$. Since there exists $M \in \mathbb{N}$ such that $b_n = a_n$ for all $n \geq M$ and $\langle b_n \rangle$ converges, we conclude from part 1 that $\langle a_n \rangle$ converges.

□

Proposition 2.2.1 allows us to deal with sequences that may not be defined at every natural number. For instance, as it currently stands, $\langle \frac{n+1}{n-3} \rangle$ is not actually a sequence because it doesn't make sense when $n = 3$. Barring such "sequences" from consideration seems drastic because they arise very naturally. However, if all that we really care about is the limit of the sequences under consideration, the previous proposition says that it doesn't matter how we fill in the value of the "sequence" $\langle \frac{n+1}{n-3} \rangle$ at $n = 3$ to make it an honest sequence. Thus, we will allow the possibility that the sequences we discuss will not be defined at finitely many points. If you really want to force them to be defined at such places, feel free to make any such sequence have the value 0 at each place where it's not defined.

Our next result says the we do not affect the convergence properties of a sequence if we "shift" it.

Proposition 2.2.2. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences and suppose that there exists $m \in \mathbb{N}$ such that $b_n = a_{n+m}$ for all $n \in \mathbb{N}^+$.*

1. *If $\langle a_n \rangle$ converges, then $\langle b_n \rangle$ converges, and in this case we have $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.*

2. *If $\langle b_n \rangle$ converges, then $\langle a_n \rangle$ converges, and in this case we have $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.*

3. *If one of $\langle a_n \rangle$ or $\langle b_n \rangle$ diverges, then so does the other.*

Proof.

1. Suppose that $\langle a_n \rangle$ converges, and let $\ell = \lim_{n\to\infty} a_n$. By assumption, we can fix $m \in \mathbb{N}^+$ such that $b_n = a_{n+m}$ for all $n \geq M$. We show that $\langle b_n \rangle$ converges to ℓ. Let $\varepsilon > 0$ be arbitrary. Since $\lim_{n\to\infty} a_n = \ell$, we can fix $N \in \mathbb{N}^+$ such that $|a_n - \ell| < \varepsilon$ for all $n \geq N$. For any $n \geq N$, we then have $n + m \geq n \geq N$, hence

$$\begin{aligned} |b_n - \ell| &= |a_{n+m} - \ell| \\ &< \varepsilon. \end{aligned}$$

   Therefore, $\lim_{n\to\infty} b_n = \ell$.

2. Suppose that $\langle b_n \rangle$ converges, and let $\ell = \lim_{n\to\infty} b_n$. By assumption, we can fix $m \in \mathbb{N}^+$ such that $b_n = a_{n+m}$ for all $n \geq M$. We show that $\langle a_n \rangle$ converges to ℓ. Let $\varepsilon > 0$ be arbitrary. Since $\lim_{n\to\infty} b_n = \ell$, we can fix $K \in \mathbb{N}^+$ such that $|b_n - \ell| < \varepsilon$ for all $n \geq K$. Let $N = m + K$. For any $n \geq N$, we then have $n \geq m + K$, hence $n - m \geq K$, and so

$$\begin{aligned} |a_n - \ell| &= |b_{n-m} - \ell| \\ &< \varepsilon. \end{aligned}$$

Therefore, $\lim_{n\to\infty} a_n = \ell$.

3. This follows immediately from (1) and (2).

□

As we've seen, proving that a particular sequence converges or diverges can be a somewhat elaborate process. It would be helpful to establish some fundamental limit laws that will make understanding a given sequence easier. Our first major result in this direction will be Theorem 2.2.8, but we start with a special case of that theorem to illustrate the key ideas. Suppose that we have a sequence $\langle a_n \rangle$ that converges to ℓ. If we have a $c \in \mathbb{R}$, and we scale the sequence by multiplying each individual term by c, then we might expect that the new sequence converges to $c\ell$.

Let's think through how we would argue such a result. Suppose that $\langle a_n \rangle$ converges to ℓ, and that $c = 3$. Let's examine the sequence $\langle 3a_n \rangle$. Suppose that we are challenged with $\varepsilon = \frac{1}{10}$, i.e. we are asked to find an $N \in \mathbb{N}^+$ such that $\langle 3a_n \rangle$ is within $\frac{1}{10}$ of $c\ell$ for all $n \geq N$. We start to compute

$$\begin{aligned} |3a_n - 3\ell| &= |3 \cdot (a_n - \ell)| \\ &= 3 \cdot |a_n - \ell|. \end{aligned}$$

Now we are assuming that $\langle a_n \rangle$ converges to ℓ, so we know that we can make $|a_n - \ell|$ as small as we would like by taking n to be sufficiently large. How small do we need to ensure $|a_n - \ell|$ is in order to achieve our goal? If we want $|3a_n - 3\ell|$ to be less that $\frac{1}{10}$, then the last line above tell us that we should go out to a point where $|a_n - \ell|$ is less than $\frac{1}{30}$. In other words, when challenged with $\varepsilon = \frac{1}{10}$, we turn around and use the assumed convergence of $\langle a_n \rangle$ on the challenged value $\frac{1}{30}$, rather than the ε given to us, in order to establish an N that works.

Using this idea, we turn to the general proof.

Proposition 2.2.3. *Let $\langle a_n \rangle$ be a convergent sequence with $\lim_{n\to\infty} a_n = \ell$ and let $c \in \mathbb{R}$. We then have that $\langle c \cdot a_n \rangle$ converges with $\lim_{n\to\infty} c \cdot a_n = c \cdot \ell$.*

Proof. If $c = 0$, we then have $c \cdot a_n = 0$ for all $n \in \mathbb{N}$, so $\lim_{n\to\infty} c \cdot a_n = 0 = c \cdot \ell$ by Proposition 2.1.4. Suppose that $c \neq 0$. Let $\varepsilon > 0$ be arbitrary. Since $\lim_{n\to\infty} a_n = \ell$, we can fix $N \in \mathbb{N}$ such that $|a_n - \ell| < \frac{\varepsilon}{|c|}$ for all $n \geq N$. Let $n \geq N$ be arbitrary. We then have

$$\begin{aligned} |c \cdot a_n - c \cdot \ell| &= |c \cdot (a_n - \ell)| \\ &= |c| \cdot |a_n - \ell| \\ &< |c| \cdot \frac{\varepsilon}{|c|} && \text{(since } n \geq N\text{)} \\ &= \varepsilon. \end{aligned}$$

Therefore, $\lim_{n\to\infty} c \cdot a_n = c \cdot \ell$. □

We immediately obtain the following corollary.

Corollary 2.2.4. *Let $\langle a_n \rangle$ be a convergent sequence with $\lim_{n\to\infty} a_n = \ell$. We then have that $\langle -a_n \rangle$ converges with $\lim_{n\to\infty} (-a_n) = -\ell$.*

Proof. This follows from Proposition 2.2.3 using $c = -1$ and the fact that $(-1) \cdot x = -x$ for every x. □

Before getting our important theorem, we establish a helpful result that is interesting in its own right. We first extend our definition of bounded to sequences.

Definition 2.2.5. *Let $\langle a_n \rangle$ be a sequence.*

1. *$\langle a_n \rangle$ is* bounded above *if the set $\{a_n : n \in \mathbb{N}\}$ is bounded above, i.e. if there exists $d \in \mathbb{R}$ with $a_n \leq d$ for all $n \in \mathbb{N}$.*
2. *$\langle a_n \rangle$ is* bounded below *if the set $\{a_n : n \in \mathbb{N}\}$ is bounded below, i.e. if there exists $d \in \mathbb{R}$ with $a_n \geq d$ for all $n \in \mathbb{N}$.*
3. *$\langle a_n \rangle$ is* bounded *if it is both bounded above and bounded below.*

We can rephrase the definition of bounded in the following way.

Proposition 2.2.6. *Let $\langle a_n \rangle$ be a sequence. We then have that $\langle a_n \rangle$ is bounded if and only if there exists $d \in \mathbb{R}$ such that $|a_n| \leq d$ for all $n \in \mathbb{N}^+$*

Proof. Exercise. □

Proposition 2.2.7. *Every convergent sequence is bounded.*

Here is the key idea of the argument. Suppose that we have a convergent sequence $\langle a_n \rangle$, and let ℓ be the limit. We know that the terms of $\langle a_n \rangle$ are all eventually close to ℓ. Now the word "eventually" in the previous sentence might cause you to worry, but there are only finitely many exceptions. Thus, in order to get an appropriate bound on the sequence, we simply take the make of the finitely many terms of the sequence that might be exceptions, together with a value just bigger than $|\ell|$.

Proof. Let $\langle a_n \rangle$ be an arbitrary convergent sequence. By definition, we can fix $\ell \in \mathbb{R}$ with $\lim_{n\to\infty} a_n = \ell$. Using $\varepsilon = 1$ in the definition of convergence to ℓ, we can fix $N \in \mathbb{N}$ such that $|a_n - \ell| < 1$ for all $n \geq N$. Let $d = \max\{|a_1|, |a_2|, \ldots, |a_{N-1}|, |\ell| + 1\}$. We claim that $|a_n| \leq d$ for all $n \in \mathbb{N}$. To see this, let $n \in \mathbb{N}$ be arbitrary. We have two cases:

- *Case 1:* Suppose that $n < N$. We then clearly have $|a_n| \leq d$ by definition of d.
- *Case 2:* Suppose then that $n \geq N$. We then have

$$\begin{aligned} |a_n| &= |\ell + a_n - \ell| \\ &\leq |\ell| + |a_n - \ell| && \text{(by the Triangle Inequality)} \\ &< |\ell| + 1 && \text{(since } n \geq N\text{)} \\ &\leq d. \end{aligned}$$

Therefore, $|a_n| \leq d$ for all $n \in \mathbb{N}$. It follows that $\langle a_n \rangle$ is bounded. □

We now prove that limits of sequences behave well with respect to arithmetic operations.

Theorem 2.2.8. *Suppose that $\langle a_n \rangle$ and $\langle b_n \rangle$ are convergent sequences with $\lim_{n\to\infty} a_n = \ell$ and $\lim_{n\to\infty} b_n = m$.*

1. *The sequence* $\langle a_n + b_n \rangle$ *converges with* $\lim_{n\to\infty}(a_n + b_n) = \ell + m$.

2. *The sequence* $\langle a_n - b_n \rangle$ *converges with* $\lim_{n\to\infty}(a_n - b_n) = \ell - m$.

3. *The sequence* $\langle a_n \cdot b_n \rangle$ *converges with* $\lim_{n\to\infty}(a_n \cdot b_n) = \ell \cdot m$.

4. *If* $m \neq 0$, *the sequence* $\langle \frac{1}{b_n} \rangle$ *is eventually well-defined, and converges with* $\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{m}$.

5. *If* $m \neq 0$, *the sequence* $\langle \frac{a_n}{b_n} \rangle$ *is eventually well-defined, and converges with* $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\ell}{m}$.

Before jumping into the proof, let's examine the key ideas behind the proof of parts (1) and (3). Suppose that we are trying to prove (1). Thus, we are assuming that $\langle a_n \rangle$ converges to ℓ and that $\langle b_n \rangle$ converges to m. We are trying to argue that $\langle a_n + b_n \rangle$ converges to $\ell + m$. So suppose that we are challenged with an $\varepsilon > 0$, and we want to argue that $|(a_n + b_n) - (\ell + m)| < \varepsilon$ for large values of n. We compute

$$\begin{aligned} |(a_n + b_n) - (\ell + m)| &= |a_n - \ell + b_n - m| \\ &\leq |a_n - \ell| + |b_n - m|. \end{aligned}$$

Now we know that we can make $|a_n - \ell|$ as small as we would like by taking n sufficiently large. We also know that we can make $|b_n - m|$ as we small as we would like by taking m sufficiently large. How small should we make each of these? We have ε worth of wiggle room in the end, so the idea is to use half of our budget on each term. That is, we should go out far enough so that $|a_n - \ell| < \frac{\varepsilon}{2}$, and we should go out far enough so that $|b_n - m| < \frac{\varepsilon}{2}$. If we are out far enough that *both* of these are true, then we can argue that $|(a_n + b_n) - (\ell + m)| < \varepsilon$. See the proof of part (1) below for the details.

Multiplication is more interesting. Suppose that we are challenged with an $\varepsilon > 0$, and we want to argue that $|a_n b_n - \ell m| < \varepsilon$ for large values of n. We compute

$$\begin{aligned} |a_n b_n - \ell m| &= |a_n b_n - a_n m + a_n m - \ell m| \\ &\leq |a_n b_n - a_n m| + |a_n m - \ell m| \\ &= |a_n| \cdot |b_n - m| + |m| \cdot |a_n - \ell|. \end{aligned}$$

Since $\langle a_n \rangle$ converges to ℓ, we can make $|a_n - \ell|$ as small as we would like by taking n sufficiently large. If we use the same trick as above, then we want to use half of our ε budget on each term. So since we want the right term $|m| \cdot |a_n - \ell|$ to be at most $\frac{\varepsilon}{2}$, we should go out far enough so that $|a_n - \ell| < \frac{\varepsilon}{2|m|}$. This works, so long as $m \neq 0$. Now we can either handle the $m = 0$ case separately, or we can just simply make the denominator a little bigger to avoid this issue. That is, we can instead just go out far enough so that $|a_n - \ell| < \frac{\varepsilon}{2|m|+1}$.

The first term is a bit harder to handle, because we have $|a_n|$ in front of $|b_n - m|$ rather than a constant. Here is where Proposition 2.2.7 comes to the rescue. We know that $\langle a_n \rangle$ is bounded, so we can fix $d \in \mathbb{R}$ with $|a_n| \leq d$ for all $n \in \mathbb{N}^+$. We can then go out far enough so that $|b_n - m| < \frac{\varepsilon}{2d}$ to make the inequalities work. As above, we can simply add a bit to the denominator to avoid worrying about the case where $d = 0$.

Proof of Theorem 2.2.8.

1. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} a_n = \ell$, we can fix $N_1 \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{\varepsilon}{2}$ for all $n \geq N_1$. Since $\lim_{n\to\infty} b_n = m$, we can fix $N_2 \in \mathbb{N}^+$ such that $|b_n - m| < \frac{\varepsilon}{2}$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. For

any $n \geq N$, we have

$$\begin{aligned} |(a_n + b_n) - (\ell + m)| &= |a_n - \ell + b_n - m| \\ &\leq |a_n - \ell| + |b_n - m| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } n \geq N_1 \text{ and } n \geq N_2\text{)} \\ &= \varepsilon. \end{aligned}$$

Therefore, $\lim\limits_{n\to\infty} (a_n + b_n) = \ell + m$.

2. Using Corollary 2.2.4, we know that $\langle -b_n \rangle$ converges with $\lim\limits_{n\to\infty} (-b_n) = -m$. Now notice that $\langle a_n - b_n \rangle$ is just $\langle a_n + (-b_n) \rangle$, and apply part (1).

3. Let $\varepsilon > 0$. Since $\langle a_n \rangle$ converges, we know from Proposition 2.2.7 that $\langle a_n \rangle$ is bounded, so we can fix $d \in \mathbb{R}$ with $|a_n| \leq d$ for all $n \in \mathbb{N}^+$. Notice that we must have $d \geq 0$. Since $\lim\limits_{n\to\infty} a_n = \ell$, we can fix $N_1 \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{\varepsilon}{2|m|+1}$ for all $n \geq N_1$. Since $\lim\limits_{n\to\infty} b_n = m$, we can fix $N_2 \in \mathbb{N}^+$ such that $|b_n - m| < \frac{\varepsilon}{2d+1}$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. For any $n \geq N$, we then have

$$\begin{aligned} |a_n b_n - \ell m| &= |a_n b_n - a_n m + a_n m - \ell m| \\ &\leq |a_n b_n - a_n m| + |a_n m - \ell m| \\ &= |a_n| \cdot |b_n - m| + |m| \cdot |a_n - \ell| \\ &\leq d \cdot |b_n - m| + |m| \cdot |a_n - \ell| \\ &\leq d \cdot \frac{\varepsilon}{2d+1} + |m| \cdot \frac{\varepsilon}{2|m|+1} && \text{(since } n \geq N_1 \text{ and } n \geq N_2\text{)} \\ &= \varepsilon \cdot \left(\frac{d}{2d+1} + \frac{|m|}{2|m|+1} \right) \\ &< \varepsilon \cdot \left(\frac{1}{2} + \frac{1}{2} \right) \\ &= \varepsilon. \end{aligned}$$

Therefore, $\lim\limits_{n\to\infty} (a_n \cdot b_n) = \ell \cdot m$.

4. Assume that $m \neq 0$. Let $\varepsilon > 0$. Since $\lim\limits_{n\to\infty} b_n = m$, we can fix $N \in \mathbb{N}^+$ such that $|b_n - m| < \min\left\{ \frac{|m|}{2}, \frac{\varepsilon m^2}{2} \right\}$ for all $n \geq N$. Notice that for any $n \geq N$, we have

$$\begin{aligned} |m| &= |m - b_n + b_n| \\ &\leq |m - b_n| + |b_n| \\ &= |b_n - m| + |b_n| \\ &< \frac{|m|}{2} + |b_n|. \end{aligned}$$

Subtracting $\frac{|m|}{2}$ from both sides, we conclude that

$$|b_n| > \frac{|m|}{2}$$

for all $n \geq N$. In particular, for any $n \geq N$, we have $b_n \neq 0$, so it makes sense to talk about $\frac{1}{b_n}$. Moreover, multiplying both sides of the above inequality by $\frac{2}{|m| \cdot |b_n|} > 0$, it follows that

$$\frac{1}{|b_n|} < \frac{2}{|m|}$$

for all $n \geq N$. Now for any $n \geq N$, we have

$$\begin{aligned}
\left| \frac{1}{b_n} - \frac{1}{m} \right| &= \left| \frac{m - b_n}{m b_n} \right| \\
&= \frac{|m - b_n|}{|m| \cdot |b_n|} \\
&= |b_n - m| \cdot \frac{1}{|m|} \cdot \frac{1}{|b_n|} \\
&< \frac{\varepsilon m^2}{2} \cdot \frac{1}{|m|} \cdot \frac{1}{|b_n|} && \text{(since } n \geq N\text{)} \\
&< \frac{\varepsilon m^2}{2} \cdot \frac{1}{|m|} \cdot \frac{2}{|m|} && \text{(from above)} \\
&= \varepsilon.
\end{aligned}$$

Therefore, $\lim_{n \to \infty} \frac{1}{b_n} = \frac{1}{m}$.

5. Notice that $\frac{a_n}{b_n} = a_n \cdot \frac{1}{b_n}$ for all $n \in \mathbb{N}$. Since $\lim_{n \to \infty} b_n = m \neq 0$, it follows from part (4) that $\langle \frac{1}{b_n} \rangle$ is eventually well-defined, and converges with $\lim_{n \to \infty} \frac{1}{b_n} = \frac{1}{m}$. Now we can use part (3) to conclude that $\lim_{n \to \infty} \frac{a_n}{b_n} = \lim_{n \to \infty} a_n \cdot \frac{1}{b_n} = \ell \cdot \frac{1}{m} = \frac{\ell}{m}$.

□

The next proposition is again intuitively reasonable, but we will see that it is quite useful. The key idea in the argument was used in the proof of part (4) of the above theorem to ensure that $\frac{1}{b_n}$ made sense, and that we could bound its values away from 0.

Proposition 2.2.9. *Suppose that $\langle a_n \rangle$ is a sequence that converges to ℓ.*

1. *If $\ell > 0$, then there exists $N \in \mathbb{N}$ such that $a_n > 0$ for all $n \geq N$.*
2. *If $\ell < 0$, then there exists $N \in \mathbb{N}$ such that $a_n < 0$ for all $n \geq N$.*

Proof.

1. Suppose that $\ell > 0$. Since $\lim_{n \to \infty} a_n = \ell$, we can fix $N \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{\ell}{2}$ for all $n \geq N$. Let $n \geq N$ be arbitrary. We then have $|a_n - \ell| < \frac{\ell}{2}$, so we know from Proposition 1.6.3 that $a_n - \ell > -\frac{\ell}{2}$. Adding ℓ to both sides, it follows that $a_n > \frac{\ell}{2}$. Since $\frac{\ell}{2} > 0$, we conclude that $a_n > 0$.

2. Suppose that $\ell < 0$. Since $\lim_{n \to \infty} a_n = \ell$, we can fix $N \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{-\ell}{2}$ for all $n \geq N$. Let $n \geq N$ be arbitrary. We then have have $|a_n - \ell| < \frac{-\ell}{2}$, so we know from Proposition 1.6.3 that $a_n - \ell < -\frac{\ell}{2}$. Adding ℓ to both sides, it follows that $a_n < \frac{\ell}{2}$. Since $\frac{\ell}{2} < 0$, we conclude that $a_n < 0$.

□

We can now use this result in conjunction with Theorem 2.2.8 to prove that limits behave reasonably with respect to the ordering of the real numbers.

Theorem 2.2.10. *Suppose that $\langle a_n \rangle$ and $\langle b_n \rangle$ are both convergent sequences. Let $\ell = \lim_{n\to\infty} a_n$ and let $m = \lim_{n\to\infty} b_n$.*

1. *If there exists $M \in \mathbb{N}^+$ such that $a_n \geq 0$ for all $n \geq M$, then $\ell \geq 0$.*
2. *If there exists $M \in \mathbb{N}^+$ such that $a_n \leq 0$ for all $n \geq M$, then $\ell \leq 0$.*
3. *If there exists $M \in \mathbb{N}^+$ such that $a_n \leq b_n$ for all $n \geq M$, then $\ell \leq m$.*

Proof.

1. Fix $M \in \mathbb{N}^+$ such that $a_n \geq 0$ for all $n \geq M$. Suppose instead that $\ell < 0$. By Proposition 2.2.9, we can then fix $N \in \mathbb{N}^+$ such that $a_n < 0$ for all $n \geq N$. Letting $n = \max\{M, N\}$, we then have both $a_n \geq 0$ and $a_n < 0$, a contradiction. Therefore, we must have $\ell \geq 0$.
2. Completely analogous to (1).
3. Consider the sequence $\langle c_n \rangle$ defined by letting $c_n = b_n - a_n$ for all $n \in \mathbb{N}^+$. Using Theorem 2.2.8, we know that $\langle c_n \rangle$ converges and that $\lim_{n\to\infty} c_n = m - \ell$. Fix $M \in \mathbb{N}^+$ with $a_n \leq b_n$ for all $n \geq M$. We then have $b_n - a_n \geq 0$ for all $n \geq M$, and hence $c_n \geq 0$ for all $n \geq M$. Using part (1), we conclude that $m - \ell \geq 0$, and hence $\ell \leq m$.

□

We now prove an exceedingly useful little theorem which says that if we can sandwich a sequence $\langle b_n \rangle$ between two sequences $\langle a_n \rangle$ and $\langle c_n \rangle$ which have the same limit, we can conclude that the middle sequence $\langle b_n \rangle$ actually converges to same limit. Notice that Theorem 2.2.10 implies that *if* $\langle b_n \rangle$ converges, then it does converge to the right limit. The novelty here is we can actually prove that $\langle b_n \rangle$ does converge.

Theorem 2.2.11 (The Squeeze Theorem)**.** *Suppose that $\langle a_n \rangle$, $\langle b_n \rangle$, and $\langle c_n \rangle$ are sequences with*

$$\lim_{n\to\infty} a_n = \ell = \lim_{n\to\infty} c_n.$$

Suppose also that there exists $M \in \mathbb{N}^+$ such that $a_n \leq b_n \leq c_n$ for all $n \geq M$. We then have that $\langle b_n \rangle$ converges, and that $\lim_{n\to\infty} b_n = \ell$.

Proof. Let $\varepsilon > 0$. By assumption, we can fix $M \in \mathbb{N}^+$ such that $a_n \leq b_n \leq c_n$ for all $n \geq M$. Since $\lim_{n\to\infty} a_n = \ell$, we can fix $N_1 \in \mathbb{N}^+$ such that $|a_n - \ell| < \varepsilon$ for all $n \geq N_1$. Since $\lim_{n\to\infty} c_n = \ell$, we can fix $N_2 \in \mathbb{N}^+$ such that $|c_n - \ell| < \varepsilon$ for all $n \geq N_2$. Let $N = \max\{M, N_1, N_2\}$. Now let $n \geq N$ be arbitrary. Since $n \geq N_1$, we have $|a_n - \ell| < \varepsilon$. Since $n \geq N_2$, we have $|c_n - \ell| < \varepsilon$. Using Proposition 1.6.3, it follows that

$$-\varepsilon < a_n - \ell < \varepsilon \qquad \text{and} \qquad -\varepsilon < c_n - \ell < \varepsilon.$$

In particular, we have both

$$\ell - \varepsilon < a_n \qquad \text{and} \qquad c_n < \ell + \varepsilon.$$

Now since $n \geq M$, we know that $a_n \leq b_n \leq c_n$, and therefore we have

$$\ell - \varepsilon < b_n < \ell + \varepsilon.$$

It follows that $-\varepsilon < b_n - \ell < \varepsilon$, and so using Proposition 1.6.3 again, we conclude that $|b_n - \ell| < \varepsilon$. Therefore, $\lim_{n\to\infty} b_n = \ell$. □

For example, suppose we can now argue that $\langle \frac{1}{n^2} \rangle$ converges to 0, by simply noticing that since $0 < n \leq n^2$ for all $n \in \mathbb{N}^+$, it follows that

$$0 < \frac{1}{n^2} \leq \frac{1}{n}$$

for all $n \in \mathbb{N}^+$. Now $\lim\limits_{n\to\infty} 0 = 0 = \lim\limits_{n\to\infty} \frac{1}{n}$ by Proposition 2.1.4 and Proposition 2.1.3. Therefore, by the Squeeze Theorem, we conclude that $\langle \frac{1}{n^2} \rangle$ converges to 0.

Sometimes, we have to do a bit more inequality work. For example, consider the sequence $\langle \frac{1}{n^2-3n-1} \rangle$. Notice that for any $n \geq 5$, we have

$$\begin{aligned} n^2 - 3n - 1 &= n(n-3) - 1 \\ &\geq 2n - 5 \\ &= n + (n-5) \\ &\geq n. \end{aligned}$$

Thus, for all $n \geq 5$, we have

$$0 < \frac{1}{n^2 - 3n - 1} \leq \frac{1}{n}.$$

Using the Squeeze Theorem again, it follows that $\langle \frac{1}{n^2-3n-1} \rangle$ converges to 0.

We now turn our attention to some special types of sequences that are relatively straightforward to analyze.

Definition 2.2.12. *Let $\langle a_n \rangle$ be a sequence.*

1. *$\langle a_n \rangle$ is* increasing *if $a_m \leq a_n$ for all $m, n \in \mathbb{N}^+$ with $m < n$.*
2. *$\langle a_n \rangle$ is* strictly increasing *if $a_m < a_n$ for all $m, n \in \mathbb{N}^+$ with $m < n$.*
3. *$\langle a_n \rangle$ is* decreasing *if $a_m \geq a_n$ for all $m, n \in \mathbb{N}^+$ with $m < n$.*
4. *$\langle a_n \rangle$ is* strictly decreasing *if $a_m > a_n$ for all $m, n \in \mathbb{N}^+$ with $m < n$.*
5. *$\langle a_n \rangle$ is* monotonic *if it is either increasing or decreasing.*

The nice thing about the discrete nature of $\mathbb{N}^+$ is that there is always an immediate "next" element of a sequence, so we can determine whether or not a sequence is increasing (or decreasing) by simply checking adjacent elements of the sequence. More precisely, we have the following result which is proved by a straightforward induction.

Proposition 2.2.13. *Let $\langle a_n \rangle$ be a sequence.*

1. *$\langle a_n \rangle$ is increasing if and only if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}^+$.*
2. *$\langle a_n \rangle$ is strictly increasing if and only if $a_n < a_{n+1}$ for all $n \in \mathbb{N}^+$.*
3. *$\langle a_n \rangle$ is decreasing if and only if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}^+$.*
4. *$\langle a_n \rangle$ is strictly decreasing if and only if $a_n > a_{n+1}$ for all $n \in \mathbb{N}^+$.*

We now establish a simple criterion for understanding when a monotonic sequence converges. This result is another manifestation of $\mathbb{R}$ not having any "holes".

Theorem 2.2.14 (Monotone Convergence Theorem)**.** *Let $\langle a_n \rangle$ be a sequence.*

1. *If $\langle a_n \rangle$ is increasing and bounded above, then $\langle a_n \rangle$ converges, and $\lim\limits_{n\to\infty} a_n = \sup\{a_n : n \in \mathbb{N}\}$.*

2. *If $\langle a_n \rangle$ is decreasing and bounded below, then $\langle a_n \rangle$ converges, and $\lim_{n\to\infty} a_n = \inf\{a_n : n \in \mathbb{N}\}$.*

Proof. We prove (1), and leave the completely analogous proof of (2) as an exercise. Suppose then that $\langle a_n \rangle$ is increasing and bounded above. By definition, it follows that the set $\{a_n : n \in \mathbb{N}\}$ is bounded above. Since this set is clearly nonempty, we can let $\ell = \sup\{a_n : n \in \mathbb{N}\}$. We show that $\lim_{n\to\infty} a_n = \ell$.

Let $\varepsilon > 0$. Since $\ell - \varepsilon < \ell$, we know that $\ell - \varepsilon$ is not an upper bound for $\{a_n : n \in \mathbb{N}\}$, hence we can fix $N \in \mathbb{N}^+$ with $a_N > \ell - \varepsilon$. Let $n \geq N$ be arbitrary. Since $\langle a_n \rangle$ is increasing and $n \geq N$, we have $a_n \geq a_N > \ell - \varepsilon$. Also, since ℓ is an upper bound for $\{a_n : n \in \mathbb{N}\}$, we have $a_n \leq \ell < \ell + \varepsilon$. Therefore, $\ell - \varepsilon < a_n < \ell + \varepsilon$, so $-\varepsilon < a_n - \ell < \varepsilon$, and hence $|a_n - \ell| < \varepsilon$ by Proposition 1.6.3. It follows that $\lim_{n\to\infty} a_n = \ell$. □

We can now finish off something we started on Homework 1. On that assignment, we defined a sequence $\langle a_n \rangle$ recursively by letting $a_1 = 0$ and letting $a_{n+1} = \frac{1}{3}(a_n + 1)$ for all $n \in \mathbb{N}^+$, and you proved that $\langle a_n \rangle$ was (strictly) increasing and bounded above. By the Monotone Convergence Theorem, it follows that $\langle a_n \rangle$ converges. Let $\ell = \lim_{n\to\infty} a_n$. We can use the results in this section to determine ℓ with very little work.

Define a new sequence $\langle b_n \rangle$ by shifting the sequence $\langle a_n \rangle$ by 1, i.e. by letting $b_n = a_{n+1}$ for all $n \in \mathbb{N}^+$. Notice that $b_n = \frac{1}{3}(a_n + 1)$ for all $n \in \mathbb{N}^+$. By Proposition 2.2.2, we know that $\langle b_n \rangle$ converges to ℓ as well. Using Theorem 2.2.8, we also know that $\langle b_n \rangle$ converges to $\frac{1}{3}(\ell + 1)$. Thus, we have $\ell = \frac{1}{3}(\ell + 1)$. Solving for ℓ, we see that $3\ell = \ell + 1$, and hence $\ell = \frac{1}{2}$.

Before moving on, we pause to note that we needed to know that $\langle a_n \rangle$ converged before we could carry out the calculation in the previous paragraph. To see what could go wrong, suppose that we define a sequence $\langle a_n \rangle$ recursively by letting $a_1 = 0$ and letting $a_{n+1} = 1 - a_n$ for all $n \in \mathbb{N}^+$. Notice then that $a_2 = 1$, $a_3 = 0$, $a_4 = 1$, etc. Thus, the sequence $\langle a_n \rangle$ diverges. If we try to follow the above argument and let $\ell = \lim_{n\to\infty} a_n$, then we find that $\ell = 1 - \ell$, and hence $\ell = \frac{1}{2}$. All that this calculation shows is that *if* $\langle a_n \rangle$ converges, then it must converge to $\frac{1}{2}$. But as mentioned above, the sequence $\langle a_n \rangle$ does not converge.

In hindsight, these examples demonstrate the power of the Monotone Convergence Theorem. Typically, if we want to prove that a sequence $\langle a_n \rangle$ converges, we need to have a candidate ℓ in mind, and then show that $\langle a_n \rangle$ converges to ℓ. But the Monotone Convergence Theorem provides a way to guarantee convergence without foreknowledge about what ℓ should be. In a few sections, we will see another very nice criterion for determining convergence of a sequence that does not require having a potential limit in hand.

2.3 Infinite Limits

Given a sequence $\langle a_n \rangle$, we defined what it means to say that $\langle a_n \rangle$ converges to ℓ. All other sequences, i.e. sequences that did not converge to any $\ell \in \mathbb{R}$, were thrown into the same category of divergent sequences. However, not all divergent sequences are created equal. We now define two special subclasses of divergent sequences.

Definition 2.3.1. *Let $\langle a_n \rangle$ be a sequence.*

1. *We say that $\langle a_n \rangle$* diverges to ∞ *if for every $z \in \mathbb{R}$, there exists $N \in \mathbb{N}^+$ such that $a_n > z$ for all $n \geq N$. We denote this by writing $\lim_{n\to\infty} a_n = \infty$.*

2. *We say that $\langle a_n \rangle$* diverges to $-\infty$ *if for every $z \in \mathbb{R}$, there exists $N \in \mathbb{N}^+$ such that $a_n < z$ for all $n \geq N$. We denote this by writing $\lim_{n\to\infty} a_n = -\infty$.*

Intuitively, a sequence $\langle a_n \rangle$ diverges to ∞ if we can make the terms arbitrarily large, provided we only look far enough out at a tail of the sequence. For a simple example, let's show that $\langle n \rangle$ diverges to ∞. Let

$z \in \mathbb{R}$. By Proposition 1.4.5, we can fix $N \in \mathbb{N}^+$ with $N > z$. For any $n \geq N$, we then have $n \geq N > z$ and hence $n > z$. It follows that $\lim_{n\to\infty} n = \infty$.

To familiarize ourselves with the definition, we prove the following simple result.

Proposition 2.3.2. *Let $\langle a_n \rangle$ be a sequence.*

1. *If $\lim_{n\to\infty} a_n = \infty$, then $\langle a_n \rangle$ is not bounded above.*
2. *If $\lim_{n\to\infty} a_n = -\infty$, then $\langle a_n \rangle$ is not bounded below.*

Proof.

1. Suppose that $\lim_{n\to\infty} a_n = \infty$. Let $d \in \mathbb{R}$ be arbitrary. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N \in \mathbb{R}$ such that $a_n > d$ for all $n \geq N$. We then have $a_N > d$, so d is not an upper bound for $\langle a_n \rangle$. It follows that $\langle a_n \rangle$ is not bounded above.
2. Suppose that $\lim_{n\to\infty} a_n = -\infty$. Let $d \in \mathbb{R}$ be arbitrary. Since $\lim_{n\to\infty} a_n = -\infty$, we can fix $N \in \mathbb{R}$ such that $a_n < d$ for all $n \geq N$. We then have $a_N < d$, so d is not a lower bound for $\langle a_n \rangle$. It follows that $\langle a_n \rangle$ is not bounded below.

□

We can now justify the word "diverges" in "diverges to ∞" and "diverges to $-\infty$".

Corollary 2.3.3. *Let $\langle a_n \rangle$ be a sequence. If either $\lim_{n\to\infty} a_n = \infty$ or $\lim_{n\to\infty} a_n = -\infty$, then $\langle a_n \rangle$ diverges.*

Proof. Suppose that $\langle a_n \rangle$ diverges to either ∞ or $-\infty$. By Proposition 2.3.2, the sequence $\langle a_n \rangle$ is not bounded, hence $\langle a_n \rangle$ does not converge by Proposition 2.2.7. Therefore, $\langle a_n \rangle$ diverges. □

Be careful to note that the converse to Proposition 2.3.2 is *not* true. That is, it is possible that a sequence is not bounded above, but also does not diverge to ∞. For example, consider the sequence $\langle a_n \rangle$ defined by letting

$$a_n = \begin{cases} n & \text{if } n \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

Also, although it is straightforward to see that an increasing sequence that is not bounded above must diverge to ∞, it is *not* true that every sequence that diverges to ∞ must be increasing. For example, consider the sequence $\langle a_n \rangle$ defined by letting

$$a_n = \begin{cases} n-2 & \text{if } n \text{ is even} \\ n & \text{otherwise.} \end{cases}$$

The next simple proposition will cut much of our later work in half.

Proposition 2.3.4. *Let $\langle a_n \rangle$ be a sequence.*

1. *If $\lim_{n\to\infty} a_n = \infty$, then $\lim_{n\to\infty} (-a_n) = -\infty$.*
2. *If $\lim_{n\to\infty} a_n = -\infty$, then $\lim_{n\to\infty} (-a_n) = \infty$.*

Proof.

1. Suppose that $\lim_{n\to\infty} a_n = \infty$. Let $z \in \mathbb{R}$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N \in \mathbb{N}^+$ such that $a_n > -z$ for all $n \geq N$. We then have $-a_n < z$ for all $n \geq N$. It follows that $\lim_{n\to\infty} (-a_n) = -\infty$.

2. Suppose that $\lim_{n\to\infty} a_n = -\infty$. Let $z \in \mathbb{R}$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N \in \mathbb{N}^+$ such that $a_n < -z$ for all $n \geq N$. We then have $-a_n > z$ for all $n \geq N$. It follows that $\lim_{n\to\infty} (-a_n) = \infty$.

□

We now examine how these limits behave with respect to addition.

Proposition 2.3.5. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences.*

1. *If $\lim_{n\to\infty} a_n = \infty$, and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$, then $\lim_{n\to\infty} (a_n + b_n) = \infty$.*
2. *If $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = \infty$, then $\lim_{n\to\infty} (a_n + b_n) = \infty$.*
3. *If $\lim_{n\to\infty} a_n = -\infty$, and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$, then $\lim_{n\to\infty} (a_n + b_n) = -\infty$.*
4. *If $\lim_{n\to\infty} a_n = -\infty$ and $\lim_{n\to\infty} b_n = -\infty$, then $\lim_{n\to\infty} (a_n + b_n) = -\infty$.*

Proof.

1. Let $z \in \mathbb{R}$. Since $\lim_{n\to\infty} b_n = \ell$, we can fix $N_1 \in \mathbb{N}^+$ such that $|b_n - \ell| < 1$ for all $n \geq N_1$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N_2 \in \mathbb{N}^+$ such that $a_n > z - \ell + 1$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. Let $n \geq N$ be arbitrary. Since $n \geq N_1$, we then have $|b_n - \ell| < 1$, so $-1 < b_n - \ell$, and hence $b_n > \ell - 1$. It follows that

$$\begin{aligned} a_n + b_n &> a_n + (\ell - 1) && \text{(from above)} \\ &> (z - \ell + 1) + (\ell - 1) && \text{(since } n \geq N_2\text{)} \\ &= z. \end{aligned}$$

   Therefore, $\lim_{n\to\infty} (a_n + b_n) = \infty$.

2. Let $z \in \mathbb{R}$. Since $\lim_{n\to\infty} b_n = \infty$, we can fix $N_1 \in \mathbb{N}^+$ such that $b_n > 0$ for all $n \geq N_1$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N_2 \in \mathbb{N}^+$ such that $a_n > z$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. For any $n \geq N$, we then have $a_n + b_n > z + 0 = z$ since $n \geq N_1$ and $n \geq N_2$. Therefore, $\lim_{n\to\infty} (a_n + b_n) = \infty$.

3. We know that $\lim_{n\to\infty} (-a_n) = \infty$ by Proposition 2.3.4 and that $\lim_{n\to\infty} (-b_n) = -\ell$ by Proposition 2.2.3. Therefore, $\lim_{n\to\infty} -(a_n + b_n) = \infty$ by part (1), so $\lim_{n\to\infty} (a_n + b_n) = -\infty$ by Proposition 2.3.4.

4. We know that $\lim_{n\to\infty} (-a_n) = \infty$ and $\lim_{n\to\infty} (-b_n) = \infty$ by Proposition 2.3.4. Therefore, $\lim_{n\to\infty} -(a_n + b_n) = \infty$ by part (2), so $\lim_{n\to\infty} (a_n + b_n) = -\infty$ by Proposition 2.3.4.

□

One important thing to notice in the above proposition is that we said nothing about the sequence $\langle a_n + b_n \rangle$ when $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = -\infty$. There's good reason for this. Simply from the knowledge that $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = -\infty$, we can conclude absolutely nothing about the convergence or divergence of $\langle a_n + b_n \rangle$. In fact, anything at all can happen. Here are some examples:

- If $a_n = n$ and $b_n = -n$, then $a_n + b_n = 0$ for all n, so $\lim_{n\to\infty} (a_n + b_n) = 0$.
- If $a_n = n + 7$ and $b_n = -n$, then $a_n + b_n = 7$ for all n, so $\lim_{n\to\infty} (a_n + b_n) = 7$.

- If $a_n = 2n$ and $b_n = -n$, then $a_n + b_n = n$ for all n, so $\lim_{n\to\infty}(a_n + b_n) = \infty$.
- If $a_n = n$ and $b_n = -2n$, then $a_n + b_n = -n$ for all n, so $\lim_{n\to\infty}(a_n + b_n) = -\infty$.
- If $a_n = n + (-1)^n$ and $b_n = -n$, then $a_n + b_n = (-1)^n$ for all n, so $\langle a_n + b_n\rangle$ diverges.

Multiplication is a bit more complicated.

Proposition 2.3.6. *Let $\langle a_n\rangle$ and $\langle b_n\rangle$ be sequences.*

1. *If $\lim_{n\to\infty} a_n = \infty$, and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell > 0$, then $\lim_{n\to\infty}(a_n \cdot b_n) = \infty$.*
2. *If $\lim_{n\to\infty} a_n = \infty$, and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell < 0$, then $\lim_{n\to\infty}(a_n \cdot b_n) = -\infty$.*
3. *If $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = \infty$, then $\lim_{n\to\infty}(a_n \cdot b_n) = \infty$.*
4. *If $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = -\infty$, then $\lim_{n\to\infty}(a_n \cdot b_n) = -\infty$.*
5. *If $\lim_{n\to\infty} a_n = -\infty$, and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell > 0$, then $\lim_{n\to\infty}(a_n \cdot b_n) = -\infty$.*
6. *If $\lim_{n\to\infty} a_n = -\infty$, and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell < 0$, then $\lim_{n\to\infty}(a_n \cdot b_n) = \infty$.*
7. *If $\lim_{n\to\infty} a_n = -\infty$ and $\lim_{n\to\infty} b_n = -\infty$, then $\lim_{n\to\infty}(a_n \cdot b_n) = \infty$.*

Proof.

1. Let $z \in \mathbb{R}$. We may assume that $z > 0$ (if we can find an N which works for all $z > 0$, then we're done because we can use the N that works for $z = 1$ for all $z \leq 0$). Since $\lim_{n\to\infty} b_n = \ell > 0$, we can fix $N_1 \in \mathbb{N}^+$ such that $|b_n - \ell| < \frac{\ell}{2}$ for all $n \geq N_1$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N_2 \in \mathbb{N}^+$ such that $a_n > \frac{2z}{\ell}$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. Let $n \geq N$ be arbitrary. Since $n \geq N_1$, e then have $|b_n - \ell| < \frac{\ell}{2}$, so $-\frac{\ell}{2} < b_n - \ell$, and hence $b_n > \ell - \frac{\ell}{2} = \frac{\ell}{2} > 0$. Now we also have $n \geq N_2$, so $a_n > \frac{2z}{\ell} > 0$. It follows that
$$\begin{aligned} a_n \cdot b_n &> \frac{2z}{\ell} \cdot \frac{\ell}{2} \\ &= z. \end{aligned}$$
Therefore, $\lim_{n\to\infty}(a_n \cdot b_n) = \infty$.
2. Since $\lim_{n\to\infty} b_n = \ell$, it follows from Proposition 2.2.3 that $\lim_{n\to\infty}(-b_n) = -\ell$. Since $-\ell > 0$, we may use part (1) to conclude that $\lim_{n\to\infty} -(a_n \cdot b_n) = \infty$. Therefore, $\lim_{n\to\infty}(a_n \cdot b_n) = -\infty$ by Proposition 2.3.4.
3. Let $z \in \mathbb{R}$. As in part (1), we may assume that $z > 0$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $N_1 \in \mathbb{N}^+$ such that $a_n > z$ for all $n \geq N_1$. Since $\lim_{n\to\infty} b_n = \infty$, we can fix $N_2 \in \mathbb{N}^+$ such that $b_n > 1$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. For any $n \geq N$, we then have $a_n \cdot b_n > z \cdot 1 = z$. Therefore, $\lim_{n\to\infty}(a_n \cdot b_n) = \infty$.
4. By Proposition 2.3.4, we have $\lim_{n\to\infty} -b_n = \infty$, hence $\lim_{n\to\infty} -(a_n \cdot b_n) = \infty$ by part (3), and so $\lim_{n\to\infty}(a_n \cdot b_n) = -\infty$ by Proposition 2.3.4.
5. By Proposition 2.3.4, we have $\lim_{n\to\infty} -a_n = \infty$, hence $\lim_{n\to\infty} -(a_n \cdot b_n) = \infty$ by part (1), and so $\lim_{n\to\infty}(a_n \cdot b_n) = -\infty$ by Proposition 2.3.4.

6. By Proposition 2.3.4, we have $\lim_{n\to\infty} -a_n = \infty$, hence $\lim_{n\to\infty} -(a_n \cdot b_n) = -\infty$ by part (1), and so $\lim_{n\to\infty} (a_n \cdot b_n) = \infty$ by Proposition 2.3.4.

7. By Proposition 2.3.4, we have $\lim_{n\to\infty} -a_n = \infty$ and $\lim_{n\to\infty} -b_n = \infty$, hence $\lim_{n\to\infty} (a_n \cdot b_n) = \infty$ by part (3).

□

Again, notice in the in this proposition we said nothing about the sequence $\langle a_n \cdot b_n \rangle$ when one of $\langle a_n \rangle$ or $\langle b_n \rangle$ has limit ∞ or $-\infty$, while the other has limit 0. Once again, we can conclude nothing about the convergence or divergence of the sequence $\langle a_n \cdot b_n \rangle$ from this information. It's a good exercise to come up with examples of sequences $\langle a_n \rangle$ and $\langle b_n \rangle$ where $\langle a_n \rangle$ diverges to ∞ and $\langle b_n \rangle$ converges to 0, but $\langle a_n \cdot b_n \rangle$ converges or diverges to anything you would like.

Before handling the general division rules, we first examine reciprocals. Notice the need for the extra positive/negative conditions in parts (2) and (4) of the next result. First, we need to ensure that $b_n = 0$ for only finitely many n or else the sequence $\langle \frac{1}{b_n} \rangle$ doesn't make sense even if we allow finitely many undefined points. More importantly, the sequence $\langle \frac{(-1)^n}{n} \rangle$ converges to 0, but the reciprocal sequence $\langle \frac{n}{(-1)^n} \rangle$, which is the sequence $\langle (-1)^n n \rangle$, neither diverges to ∞ nor diverges to $-\infty$ (intuitively, it bounces back and forth between huge positive values and huge negative values).

Although we state parts (2) and (4) in terms of $\langle b_n \rangle$ always being positive (resp. negative) for simplicity, it suffices to assume that $\langle b_n \rangle$ is eventually positive (resp. negative).

Proposition 2.3.7. *Let $\langle b_n \rangle$ be a sequence.*

1. *If $\lim_{n\to\infty} b_n = \infty$, then $\lim_{n\to\infty} \frac{1}{b_n} = 0$.*

2. *If $\lim_{n\to\infty} b_n = 0$ and $b_n > 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{1}{b_n} = \infty$.*

3. *If $\lim_{n\to\infty} b_n = -\infty$, then $\lim_{n\to\infty} \frac{1}{b_n} = 0$.*

4. *If $\lim_{n\to\infty} b_n = 0$ and $b_n < 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{1}{b_n} = -\infty$.*

Proof.

1. Suppose that $\lim_{n\to\infty} b_n = \infty$. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} b_n = \infty$, we can fix $N \in \mathbb{N}^+$ such that $b_n > \frac{1}{\varepsilon}$ for all $n \geq N$. For any $n \geq N$, we then have $|\frac{1}{b_n} - 0| = \frac{1}{b_n} < \varepsilon$ because $b_n > \frac{1}{\varepsilon} > 0$. Therefore, $\lim_{n\to\infty} \frac{1}{b_n} = 0$.

2. Suppose that $\lim_{n\to\infty} b_n = 0$ and $b_n > 0$ for all $n \in \mathbb{N}^+$. Let $z \in \mathbb{R}$. As above, we may assume that $z > 0$. Since $\lim_{n\to\infty} b_n = 0$, we can fix $N \in \mathbb{N}^+$ such that $|b_n - 0| < \frac{1}{z}$ for all $n \geq N$. For any $n \geq N$, we then have $0 < b_n < \frac{1}{z}$ since $b_n > 0$, hence $\frac{1}{b_n} > z$. Therefore, $\lim_{n\to\infty} \frac{1}{b_n} = \infty$.

3. We have $\lim_{n\to\infty} (-b_n) = \infty$ by Proposition 2.3.4, so $\lim_{n\to\infty} \frac{1}{-b_n} = 0$ by part (1), and hence $\lim_{n\to\infty} \frac{1}{b_n} = 0$.

4. We have $\lim_{n\to\infty} (-b_n) = 0$ and $-b_n > 0$ for all $n \in \mathbb{N}^+$, so $\lim_{n\to\infty} \frac{1}{-b_n} = \infty$ by part (2), and hence $\lim_{n\to\infty} \frac{1}{b_n} = -\infty$ by Proposition 2.3.4.

□

By combining the two previous propositions, we easily arrive at the limit rules for the quotient of two sequences. It turns out that there are many cases, but each follows immediately from the corresponding cases for inversion and multiplication given above.

Proposition 2.3.8. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences.*

1. *If $\lim_{n\to\infty} a_n = \ell$ where $\ell \in \mathbb{R}$ and $\lim_{n\to\infty} b_n = \infty$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$.*
2. *If $\lim_{n\to\infty} a_n = \ell$ where $\ell \in \mathbb{R}$ and $\lim_{n\to\infty} b_n = -\infty$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$.*
3. *If $\lim_{n\to\infty} a_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell > 0$, and $\lim_{n\to\infty} b_n = 0$ where $b_n > 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$.*
4. *If $\lim_{n\to\infty} a_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell > 0$, and $\lim_{n\to\infty} b_n = 0$ where $b_n < 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = -\infty$.*
5. *If $\lim_{n\to\infty} a_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell < 0$, and $\lim_{n\to\infty} b_n = 0$ where $b_n > 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = -\infty$.*
6. *If $\lim_{n\to\infty} a_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell < 0$, and $\lim_{n\to\infty} b_n = 0$ where $b_n < 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$.*
7. *If $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell > 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$.*
8. *If $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell < 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = -\infty$.*
9. *If $\lim_{n\to\infty} a_n = -\infty$ and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell > 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = -\infty$.*
10. *If $\lim_{n\to\infty} a_n = -\infty$ and $\lim_{n\to\infty} b_n = \ell$ where $\ell \in \mathbb{R}$ and $\ell < 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$.*
11. *If $\lim_{n\to\infty} a_n = \infty$, and $\lim_{n\to\infty} b_n = 0$ where $b_n > 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$.*
12. *If $\lim_{n\to\infty} a_n = \infty$, and $\lim_{n\to\infty} b_n = 0$ where $b_n < 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = -\infty$.*
13. *If $\lim_{n\to\infty} a_n = -\infty$, and $\lim_{n\to\infty} b_n = 0$ where $b_n > 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = -\infty$.*
14. *If $\lim_{n\to\infty} a_n = -\infty$, and $\lim_{n\to\infty} b_n = 0$ where $b_n < 0$ for all $n \in \mathbb{N}^+$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$.*

The next proposition is a kind of Squeeze Theorem for sequences which diverge to either ∞ or $-\infty$. It's much simpler the normal Squeeze Theorem.

Proposition 2.3.9. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences.*

1. *If $\lim_{n\to\infty} a_n = \infty$ and there exists an $M \in \mathbb{N}^+$ such that $b_n \geq a_n$ for all $n \geq M$, then $\lim_{n\to\infty} b_n = \infty$.*
2. *If $\lim_{n\to\infty} a_n = -\infty$ and there exists an $M \in \mathbb{N}^+$ such that $b_n \leq a_n$ for all $n \geq M$, then $\lim_{n\to\infty} b_n = -\infty$.*

Proof.

1. Let $z \in \mathbb{R}$. Fix $M \in \mathbb{N}^+$ such that $b_n \geq a_n$ for all $n \geq M$. Since $\lim_{n\to\infty} a_n = \infty$, we can fix $K \in \mathbb{N}^+$ such that $a_n > z$ for all $n \geq K$. Let $N = \max\{K, M\}$. For any $n > N$, we have $b_n \geq a_n > z$ (where the first inequality follows because $n \geq M$, and the second because $n \geq K$). Therefore, $\lim_{n\to\infty} b_n = \infty$.

2. Let $z \in \mathbb{R}$. Fix $M \in \mathbb{N}^+$ such that $b_n \leq a_n$ for all $n \geq M$. Since $\lim_{n\to\infty} a_n = -\infty$, we can fix $K \in \mathbb{N}^+$ such that $a_n < z$ for all $n \geq K$. Let $N = \max\{K, M\}$. For any $n \geq N$, we have $b_n \leq a_n < z$ (where the first inequality follows because $n > M$, and the second because $n > K$). Therefore, $\lim_{n\to\infty} b_n = -\infty$.

□

With these fundamental results in hand, it's time to determine convergence and divergence (and if convergence, find the limit) of an important class of sequences that will arise many times throughout our study. Take an $x \in \mathbb{R}$, and consider the sequence $\langle x^n \rangle$, i.e. the sequence $x, x^2, x^3, x^4, \ldots$. If $x > 1$, it is reasonable to believe that this sequence diverges to ∞. Similarly, if $0 \leq x < 1$, then it appears likely that the sequence converges to 0. In order to handle negative values of x easily, we will employ the following lemma.

Lemma 2.3.10. *Let $\langle a_n \rangle$ be a sequence. We then have $\lim_{n\to\infty} |a_n| = 0$ if and only if $\lim_{n\to\infty} a_n = 0$.*

Proof. Exercise. The key fact is that $||a_n| - 0| = ||a_n|| = |a_n| = |a_n - 0|$ for any $n \in \mathbb{N}^+$. □

Proposition 2.3.11. *Let $x \in \mathbb{R}$.*

1. *If $x > 1$, then $\lim_{n\to\infty} x^n = \infty$.*
2. *If $x = 1$, then $\lim_{n\to\infty} x^n = 1$.*
3. *If $|x| < 1$, then $\lim_{n\to\infty} x^n = 0$.*
4. *If $x \leq -1$, then $\langle x^n \rangle$ diverges.*

Proof.

1. Assume that $x > 1$. On Homework 1, you showed that $(1 + y)^n \geq 1 + ny$ for all $y > -1$ and $n \in \mathbb{N}$. Since $x > 1$, we have $x - 1 > 0$, so we can apply this result to conclude that

$$\begin{aligned} x^n &= (1 + (x - 1))^n \\ &\geq 1 + n(x - 1) \end{aligned}$$

for all $n \in \mathbb{N}^+$. Now we know that $\lim_{n\to\infty} n = \infty$. Since $x - 1 > 0$, we can use Proposition 2.3.6 to conclude that $\lim_{n\to\infty} (x - 1)n = \infty$. Using Proposition 2.3.5, it follows that $\lim_{n\to\infty} (1 + n(x - 1)) = \infty$. Finally, since $x^n \geq 1 + n(x - 1)$ for all $n \in \mathbb{N}^+$, we can apply Proposition 2.3.9 to conclude that $\lim_{n\to\infty} x^n = \infty$.

2. Immediate from Proposition 2.1.4 since $1^n = 1$ for every $n \in \mathbb{N}$.

3. Assume that $|x| < 1$. If $x = 0$, then the conclusion is clear, so we may assume that $x \neq 0$. We now give two proofs that $\lim_{n\to\infty} x^n = 0$:

   - Since $x \neq 0$, we can multiply both sides of the inequality $|x| < 1$ by $\frac{1}{|x|} > 0$ to conclude that $\frac{1}{|x|} > 1$. By part (1), it follows that $\lim_{n\to\infty} (\frac{1}{|x|})^n = \infty$, and hence $\lim_{n\to\infty} (\frac{1}{|x^n|}) = \infty$. Since $|x^n| = \frac{1}{1/|x^n|}$, we can apply Proposition 2.3.7 to conclude that $\lim_{n\to\infty} |x^n| = 0$. Using Lemma 2.3.10, it follows that $\lim_{n\to\infty} x^n = 0$.

- Let $a_n = |x|^n = |x^n|$ for all $n \in \mathbb{N}^+$. Since $|x| < 1$, we have

$$\begin{aligned} a_{n+1} &= |x|^{n+1} \\ &= |x| \cdot |x|^n \\ &< |x|^n \\ &= a_n \end{aligned}$$

  for all $n \in \mathbb{N}^+$, so $\langle a_n \rangle$ is strictly decreasing. Since $a_n > 0$ for all $n \in \mathbb{N}^+$, also know that $\langle a_n \rangle$ is bounded below. Thus, we can use the Monotone Convergence Theorem to conclude that $\langle a_n \rangle$ converges. Let $\ell = \lim\limits_{n\to\infty} a_n$. Since $a_{n+1} = |x| \cdot a_n$ for all $n \in \mathbb{N}^+$, it follows that $\ell = |x| \cdot \ell$, and hence $(1 - |x|) \cdot \ell = 0$. Since $|x| \neq 1$, we have $1 - |x| \neq 0$, and thus we must have $\ell = 0$.

4. If $x = -1$, this follows from Proposition 2.1.7. Suppose that $x < -1$. We then have $|x| > 1$, so by part (1) we know that $\lim\limits_{n\to\infty} |x|^n = \infty$. In particular, the sequence $\langle |x|^n \rangle$ is not bounded, hence the sequence $\langle x^n \rangle$ is not bounded. Using Proposition 2.2.7, it follows that $\langle x^n \rangle$ diverges.

□

Before moving on, we pause to ask a simple question. How fast does $n!$ grow? Of course, it grows very quickly, but can we say more? Can we compare it to some easier sequences? First, notice that for any $n \in \mathbb{N}^+$, we have

$$2^{n-1} \leq n! \leq n^n$$

because every term in the product $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$ is at most n, and at least 2 (with the exception of the first). However, these bounds are very loose. To formalize this idea, one way to compare the growth rate of two sequences $\langle a_n \rangle$ and $\langle b_n \rangle$ is to look at the quotient $\langle \frac{a_n}{b_n} \rangle$. If $\langle a_n \rangle$ grows much faster than $\langle b_n \rangle$, then we expect that the quotient will diverge to ∞. If $\langle b_n \rangle$ grows much faster than $\langle a_n \rangle$, then we expect that the quotient converges to 0. But what if the quotient converges to a real number strictly greater than 0? If $\langle \frac{a_n}{b_n} \rangle$ converges to 2, then for large values of n, we expect that $\langle a_n \rangle$ is roughly twice as big as $\langle b_n \rangle$. The gold standard is when the quotient converges to 1, which is a formal way to say that $\langle a_n \rangle$ and $\langle b_n \rangle$ grow at roughly the same rate.

With that in mind, let's return to $n!$. Can we find a simple sequence $\langle b_n \rangle$ that grows at roughly the same rate? Since $n!$ is the product of the first n natural numbers, it seems that $(\frac{n}{2})^n$ is a plausible guess (since we are somehow "averaging out" the n factors $1, 2, \ldots, n$ and replacing them all by $\frac{n}{2}$). However, this doesn't work. It turns out that

$$\lim_{n\to\infty} \frac{(n/2)^n}{n!} = \infty.$$

To get a sense for why this happens, consider comparing $100!$ to 50^{100}. We can look at $100!$ and form the 50 pairs $1 \cdot 100, 2 \cdot 99, 3 \cdot 98, \ldots, 50 \cdot 51$. Notice that essentially all of these products are much less that $50 \cdot 50$. The problem is that to maximize xy subject to the constraint $x + y = 101$, we should take $x = y = \frac{101}{2} = 50.5$.

Perhaps the next thing to try is bring down the denominator in $\frac{n}{2}$ a bit to compensate. However, it turns out that

$$\lim_{n\to\infty} \frac{(n/3)^n}{n!} = 0,$$

so $(\frac{n}{3})^n$ grows too slow. Perhaps shockingly, the number e (the most *obvious* number between 2 and 3) is the right choice. Although

$$\lim_{n\to\infty} \frac{(n/e)^n}{n!} = 0,$$

if we add just a little bit to help out the numerator, then things turn around. In fact, the sequence

$$\left\langle \frac{\sqrt{n} \cdot (n/e)^n}{n!} \right\rangle$$

does converge, but it just so happens that the limit is $\frac{1}{\sqrt{2\pi}}$.

Theorem 2.3.12 (Stirling's Approximation to $n!$)**.** *We have*

$$\lim_{n\to\infty} \frac{\sqrt{2\pi n} \cdot (n/e)^n}{n!} = 1,$$

Unfortunately, a complete proof of this fact will take us too far afield. However, we will eventually be able to prove some bounds in this direction once we study integration.

2.4 Subsequences and Cauchy Sequences

The Monotone Convergence Theorem tells us that every bounded monotonic sequence converges. However, many important sequences fail to be monotonic, so determining whether or not they converge will require new tricks. Despite the fact that many sequences are not be monotonic, it is true that every sequence has a subsequence which is monotonic. We now embark on a proof of this surprising fact, and we begin by giving a careful definition of a subsequence. Intuitively, given a sequence $\langle a_n \rangle$, a subsequence of $\langle a_n \rangle$ is just a sequence that is obtained by dropping some terms from $\langle a_n \rangle$ in such a way that infinitely many remain, and then viewing the remaining terms as a sequence (by re-indexing). To illustrate, if we view our sequence $\langle a_n \rangle$ as

$$a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad \cdots$$

then a subsequence might look like

$$a_3 \quad a_5 \quad a_{10} \quad a_{11} \quad a_{18} \quad a_{42} \quad \cdots$$

Definition 2.4.1. *Let $\langle a_n \rangle_{n=1}^{\infty}$ be a sequence. Given a strictly increasing sequence $\langle n_k \rangle_{k=1}^{\infty}$ of positive natural numbers, we call $\langle a_{n_k} \rangle_{k=1}^{\infty}$ a* subsequence *of $\langle a_n \rangle_{n=1}^{\infty}$.*

For example, let $a_n = (-1)^n$ for every $n \in \mathbb{N}^+$, which we visualize as

$$-1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad \cdots$$

Letting $n_k = 2k$ for each $k \in \mathbb{N}^+$, we obtain following strictly increasing sequence of natural numbers:

$$2 \quad 4 \quad 6 \quad 8 \quad 10 \quad 12 \quad \cdots$$

The subsequence $\langle a_{n_k} \rangle$ is then $a_2, a_4, a_6, \ldots$, which is the sequence

$$1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad \cdots$$

One simple yet fundamental fact about strictly increasing sequences of natural numbers is the following.

Lemma 2.4.2. *Let $\langle n_k \rangle$ be a strictly increasing sequence of natural numbers. We then have $n_k \geq k$ for all $k \in \mathbb{N}^+$.*

Proof. We prove this by induction on k. For the base case, notice that since $n_1 \in \mathbb{N}^+$, we must have $n_1 \geq 1$. Now let $k \in \mathbb{N}^+$ be arbitrary with $n_k \geq k$. We then have

$$\begin{aligned} n_{k+1} &> n_k && \text{(since } \langle n_k \rangle \text{ is increasing)} \\ &\geq k && \text{(by the inductive hypothesis),} \end{aligned}$$

so $n_{k+1} > k$. Since $n_{k+1} \in \mathbb{N}^+$, it follows that $n_{k+1} \geq k + 1$. By induction, we conclude that $n_k \geq k$ for all $k \in \mathbb{N}^+$. □

Proposition 2.4.3. *Every sequence has a monotonic subsequence.*

Proof. Let $\langle a_n \rangle$ be a sequence. Given $n \in \mathbb{N}^+$, we say that n is a *peak point* of the sequence $\langle a_n \rangle$ if $a_n > a_m$ for every $m > n$. We now have two cases:

- *Case 1:* Suppose that $\langle a_n \rangle$ has infinitely many peak points. Let $n_1 < n_2 < n_3 < \cdots$ be the peak points of $\langle a_n \rangle$. For any $k \in \mathbb{N}$, we then have $a_{n_k} > a_{n_{k+1}}$ since n_k is a peak point and $n_k < n_{k+1}$. Therefore, $\langle a_{n_k} \rangle$ is a (strictly) decreasing subsequence of $\langle a_n \rangle$.

- *Case 2:* Suppose that $\langle a_n \rangle$ has only finitely many peak points. Fix $n_1 \in \mathbb{N}$ larger than every peak point. Since n_1 is not a peak point, we can fix $n_2 > n_1$ such that $a_{n_1} \leq a_{n_2}$. Suppose now that we have defined $n_1 < n_2 < \cdots < n_k$ so that $a_{n_1} \leq a_{n_2} \leq \cdots \leq a_{n_k}$. Since $n_k > n_1$, we know that n_k is not a peak point, so we can fix $n_{k+1} > n_k$ such that $a_{n_k} \leq a_{n_{k+1}}$. Notice that $\langle a_{n_k} \rangle$ is then an increasing subsequence of $\langle a_n \rangle$.

Therefore, in either case, we have found a monotonic subsequence. □

Theorem 2.4.4 (Bolzano-Weierstrass)**.** *Every bounded sequence has a convergent subsequence.*

Proof. Suppose that $\langle a_n \rangle$ is a bounded sequence. By Proposition 2.4.3, $\langle a_n \rangle$ has a monotonic subsequence which converges by Theorem 2.2.14 (notice that a subsequence of a bounded sequence is bounded). □

Although the Bolzano-Weierstrass Theorem is an elegant result, we often want to show that a sequence $\langle a_n \rangle$ converges, not that we can extract a convergent subsequence from it. At first glance, in order to determine if a sequence $\langle a_n \rangle$ converges, we need to put forward an $\ell \in \mathbb{R}$ and verify the definition of $\lim_{n\to\infty} a_n = \ell$. It would be nice to have a way to determine whether or not $\langle a_n \rangle$ converges simply from properties of the sequence itself without needing to have a particular ℓ in mind. So far, the only result we have in this direction is the Monotone Convergence Theorem. Can we develop a criterion that applies to general sequences?

Intuitively, if $\lim_{n\to\infty} a_n = \ell$, then the elements of the sequence $\langle a_n \rangle$ are getting closer and closer to ℓ as we make n very large. If we wanted to dispense with the limit ℓ and focus on the sequence itself, we might think that a sequence converges if and only if its terms are getting closer and closer to each other. The first step is to turn this idea into a precise definition.

Definition 2.4.5. *A sequence $\langle a_n \rangle$ is a* Cauchy sequence *if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$ such that $|a_n - a_m| < \varepsilon$ for all $n, m \geq N$.*

One often sees the above definition abbreviated as $\lim_{n,m\to\infty} |a_n - a_m| = 0$. It's not bad notation, but it can be a little confusing at first and takes some getting used to. It's meant to convey the idea that we can make $|a_n - a_m|$ very small provided that *both* n and m are very large. If we keep n small, say $n = 2$, and let m get very large, it need not be the case that we can make $|a_2 - a_m|$ very small.

We now embark on a proof of the fact that a sequence converges if and only if it is Cauchy. We start with the easier direction.

Proposition 2.4.6. *Every convergent sequence is a Cauchy sequence.*

Proof. Suppose that $\langle a_n \rangle$ converges and fix $\ell \in \mathbb{R}$ with $\lim_{n\to\infty} a_n = \ell$. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} a_n = \ell$, we can

fix $N \in \mathbb{N}^+$ such that $|a_n - \ell| < \frac{\varepsilon}{2}$ for all $n \geq N$. For any $n, m \geq N$, we then have

$$\begin{aligned} |a_n - a_m| &= |a_n - \ell + \ell - a_m| \\ &\leq |a_n - \ell| + |\ell - a_m| && \text{(by the Triangle Inequality)} \\ &= |a_n - \ell| + |a_m - \ell| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } n, m \geq N\text{)} \\ &= \varepsilon. \end{aligned}$$

Therefore, $\langle a_n \rangle$ is a Cauchy sequence. □

That wasn't so hard, but the converse is more interesting, and takes more work. If we know that the terms of a sequence are getting closer and closer to each other, why are they necessarily converging to something? The intuitive idea is that the real numbers have no "holes" in them (this is the main point of our completeness axiom), so the sequence is not able to squirm into some nether region of the real number line. However, a careful proof that all Cauchy sequences converge takes some ingenuity. We start by proving an analogue of Proposition 2.2.7 for Cauchy sequences.

Proposition 2.4.7. *Every Cauchy sequence is bounded.*

Proof. Suppose that $\langle a_n \rangle$ is a Cauchy sequence. Using $\varepsilon = 1$, we can fix $N \in \mathbb{N}^+$ such that $|a_n - a_m| < 1$ for all $n, m \geq N$. Let $d = \max\{|a_1|, |a_2|, \ldots, |a_{N-1}|, |a_N| + 1\}$. We claim that $|a_n| \leq d$ for all $n \in \mathbb{N}$. Let $n \in \mathbb{N}^+$ be arbitrary. If $n < N$, then we clearly have $|a_n| \leq d$. On the other hand, if $n \leq N$, then

$$\begin{aligned} |a_n| &= |a_N + a_n - a_N| \\ &\leq |a_N| + |a_n - a_N| && \text{(by the Triangle Inequality)} \\ &< |a_N| + 1 && \text{(since both } n \geq N \text{ and } N \geq N\text{)} \\ &\leq d. \end{aligned}$$

Therefore, $|a_n| \leq d$ for all $n \in \mathbb{N}^+$. It follows that $\langle a_n \rangle$ is bounded. □

Since every Cauchy sequence is bounded, we know from the Bolzano-Weierstrass Theorem that every Cauchy sequence has a convergent subsequence. At first, this might not seem particularly useful, because we want to argue that the sequence itself converges. However, extracting a convergent subsequence is incredibly helpful because that convergent subsequence has a limit ℓ, which gives us a candidate for the limit of our original sequence. Intuitively, the terms of the subsequence are getting closer and closer to ℓ, and the terms of the whole sequence are getting closer and closer to each other. Since elements of the subsequence appear arbitrarily far out in the sequence, we might hope to piece these ideas together to argue that the whole sequence converges to ℓ.

Lemma 2.4.8. *Suppose that $\langle a_n \rangle$ is a Cauchy sequence and that $\langle a_{n_k} \rangle$ is a convergent subsequence with $\lim\limits_{k \to \infty} a_{n_k} = \ell$. We then have $\lim\limits_{n \to \infty} a_n = \ell$.*

Proof. Let $\varepsilon > 0$. Since $\langle a_n \rangle$ is a Cauchy sequence, we can fix $N \in \mathbb{N}^+$ such that $|a_n - a_m| < \frac{\varepsilon}{2}$ for all $m, n \geq N$. Since $\lim\limits_{k \to \infty} a_{n_k} = \ell$, we can fix $K \in \mathbb{N}^+$ such that $|a_{n_k} - \ell| < \frac{\varepsilon}{2}$ for all $k \geq K$. Let $k = \max\{K, N\}$.

We then have $k \geq N$ and hence $n_k \geq n_N \geq N$ by Lemma 2.4.2. For any $n \geq N$, we therefore have

$$\begin{aligned}
|a_n - \ell| &= |a_n - a_{n_k} + a_{n_k} - \ell| \\
&\leq |a_n - a_{n_k}| + |a_{n_k} - \ell| && \text{(by the Triangle Inequality)} \\
&< \frac{\varepsilon}{2} + |a_{n_k} - \ell| && \text{(since both } n \geq N \text{ and } n_k \geq N\text{)} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } k \geq K\text{)} \\
&= \varepsilon.
\end{aligned}$$

Therefore, $\lim_{n\to\infty} a_n = \ell$. □

We have established all of the key ingredients to prove the converse of Proposition 2.4.6.

Theorem 2.4.9. *Every Cauchy sequence converges.*

Proof. Suppose that $\langle a_n \rangle$ is a Cauchy sequence. By Proposition 2.4.7, we know that $\langle a_n \rangle$ is bounded. Using the Bolzano-Weierstrass Theorem, we may conclude that $\langle a_n \rangle$ has a convergent subsequence. Therefore, $\langle a_n \rangle$ converges by Lemma 2.4.8. □

Chapter 3

Series of Real Numbers

Suppose I want to move (in a straight line) one meter forward from the position at which I'm currently standing. In order to do, I must first move $\frac{1}{2}$ the total distance, which is $\frac{1}{2}$ of a meter. Next, I need to move $\frac{1}{2}$ of the remaining distance of $\frac{1}{2}$ of a meter, which is another $\frac{1}{4}$ of a meter. After that, I need to move $\frac{1}{2}$ of the remaining distance of $\frac{1}{4}$ of a meter, which is another $\frac{1}{8}$ of a meter. And so on. Zeno applied to this line reasoning to argue that motion was impossible because it was necessary composed of these infinitely many actions, and he put it forward as one of his paradoxes. However, from a mathematical perspective, this line of argument suggests that

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1$$

in some strange sort of way. Thus, it should be possible to somehow make sense out the idea of adding infinitely many numbers together. We need to tread very carefully and make our definitions and arguments precise or else we will end up writing things such as:

$$\begin{aligned} 0 &= 0 + 0 + 0 + \dots \\ &= (1 + (-1)) + (1 + (-1)) + (1 + (-1)) + \dots \\ &= 1 + ((-1) + 1) + ((-1) + 1) + ((-1) + 1) + \dots \\ &= 1 + 0 + 0 + 0 + \dots \\ &= 1. \end{aligned}$$

Without a careful treatment, who's to say what's wrong with this derivation?

3.1 Definitions and Basic Results

Let's think about

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1$$

a little more deeply. Instead of writing all of those pluses, which right now have no coherent meaning, let's instead simply write down the sequence of summands:

$$\frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{16} \quad \cdots$$

Ok, so we are now in possession of a sequence, which we can write more formally as $\langle \frac{1}{2^n} \rangle$. We are trying to make sense of adding together all of these terms, and no formal definition seems apparent, so we make

do with what we know. Instead of trying to add all of the terms together, we simply cut off the sum at some finite stage. Taking just the first term gives the value $\frac{1}{2}$. Taking the first two terms and adding them together gives

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

Next we notice that the sum of the first three terms is

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}.$$

Each of these finite sums which arise by cutting off at some point is called a *partial sum* of the sequence $\langle \frac{1}{2^n} \rangle$. Of course, any particular cut off ignores practically all of terms, so there is no reason to think that we should use it as the value of the infinite sum. However, if the infinite sum is to mean anything, then later and later cut-offs should provide better and better approximations to it. We thus look at the sequence of partial sums:

$$\frac{1}{2} \qquad \frac{1}{2} + \frac{1}{4} \qquad \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \qquad \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} \qquad \cdots$$

which we can write as

$$\left\langle \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n} \right\rangle.$$

In order to get a cleaner expression for $\frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n}$, recall from Homework 1 that for any $r \neq 1$ and any $n \in \mathbb{N}$, we have

$$1 + r + r^2 + \cdots + r^n = \frac{r^{n+1} - 1}{r - 1}.$$

Since $\frac{1}{2} \neq 1$, we can conclude that for any $n \in \mathbb{N}^+$, we have

$$\begin{aligned}
\frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n} &= \frac{1}{2} \cdot \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{n-1}} \right) \\
&= \frac{1}{2} \cdot \left(1 + \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^{n-1} \right) \\
&= \frac{1}{2} \cdot \frac{1 - (\frac{1}{2})^n}{1 - \frac{1}{2}} \\
&= 1 - \left(\frac{1}{2}\right)^n \\
&= \frac{2^n - 1}{2^n}.
\end{aligned}$$

Thus, the sequence of partial sums is

$$\left\langle 1 - \left(\frac{1}{2}\right)^n \right\rangle = \left\langle \frac{2^n - 1}{2^n} \right\rangle,$$

which starts out as

$$\frac{1}{2} \qquad \frac{3}{4} \qquad \frac{7}{8} \qquad \frac{15}{16} \qquad \cdots$$

Now we're in business. If later and later cut-offs should provide better and better approximations to the infinite sum, then perhaps we should simply *define* the infinite sum to be the limit of the sequence of partial

sums, if it exists. If we define the infinite sum in this way, we then get the result that we want. To see this, we want to calculate

$$\lim_{n\to\infty}\left(\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^n}\right)=\lim_{n\to\infty}\left(1-\left(\frac{1}{2}\right)^n\right)$$

Using Proposition 2.3.11, we know that

$$\lim_{n\to\infty}\left(\frac{1}{2}\right)^n=0,$$

so it follows that

$$\lim_{n\to\infty}\left(\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^n}\right)=1-0=1.$$

With such success, we are now ready to make the general precise definition.

Definition 3.1.1. *Let $\langle a_n\rangle$ be a sequence. Define a new sequence $\langle s_n\rangle$ by letting*

$$s_n=\sum_{i=1}^{n}a_i=a_1+a_2+\cdots+a_n$$

for every $n\in\mathbb{N}^+$. We call $\langle s_n\rangle$ the sequence of partial sums *of $\langle a_n\rangle$.*

1. *If the sequence $\langle s_n\rangle$ converges, we say that the series $\sum_{n=1}^{\infty}a_n$* converges*, and we write $\sum_{n=1}^{\infty}a_n$ to denote $\lim_{n\to\infty}s_n$.*
2. *If the sequence $\langle s_n\rangle$ diverges, we say that the series $\sum_{n=1}^{\infty}a_n$* diverges*.*
3. *If the sequence $\langle s_n\rangle$ diverges to ∞, we say that the series $\sum_{n=1}^{\infty}a_n$* diverges to ∞*, and write $\sum_{n=1}^{\infty}a_n=\infty$.*
4. *If the sequence $\langle s_n\rangle$ diverges to $-\infty$, we say that the series $\sum_{n=1}^{\infty}a_n$* diverges to $-\infty$*, and write $\sum_{n=1}^{\infty}a_n=-\infty$.*

Let's return to the cautionary example of the "infinite sum"

$$1+(-1)+1+(-1)+1+(-1)+\ldots$$

mentioned above. To formalize this, we start with the sequence $\langle a_n\rangle$ defined by letting $a_n=(-1)^{n+1}$ for all $n\in\mathbb{N}^+$. We are then asking whether $\sum_{n=1}^{\infty}a_n$ converges. If we form the sequence $\langle s_n\rangle$ of partial sums, then we have

$$s_n=\begin{cases}1 & \text{if } n \text{ is odd}\\ 0 & \text{otherwise.}\end{cases}$$

Notice that $\langle s_n\rangle$ diverges, so by definition the series $\sum_{n=1}^{\infty}a_n=\sum_{n=1}^{\infty}(-1)^{n+1}$ diverges.

For another example, let's show that $\sum_{n=1}^{\infty}\frac{1}{n(n+1)}$ converges and that

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

Let $\langle s_n\rangle$ be the sequence of partial sums of $\langle\frac{1}{n(n+1)}\rangle$, i.e.

$$s_n=\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\cdots+\frac{1}{n(n+1)}.$$

for all $n \in \mathbb{N}^+$. It would be a lot easier to understand the sequence $\langle s_n \rangle$ if we could develop a simple formula for s_n that does not involve a sum. If we play around with small values, we notice the following:

- $s_1 = \frac{1}{1 \cdot 2} = \frac{1}{2}$.
- $s_2 = s_1 + \frac{1}{2 \cdot 3} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$.
- $s_3 = s_2 + \frac{1}{3 \cdot 4} = \frac{2}{3} + \frac{1}{12} = \frac{3}{4}$.
- $s_4 = s_3 + \frac{1}{4 \cdot 5} = \frac{3}{4} + \frac{1}{20} = \frac{4}{5}$.

From this simple experimentation, a pattern seems to emerge. We now prove that $s_n = \frac{n}{n+1}$ in 2 different ways:

- First, we argue that $s_n = \frac{n}{n+1}$ for all $n \in \mathbb{N}^+$ by induction. Notice that $s_1 = \frac{1}{1 \cdot 2}$, so we have our base case. Let $n \in \mathbb{N}^+$ be arbitrary such that $s_n = \frac{n}{n+1}$ is true. We then have

$$\begin{aligned} s_{n+1} &= \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} + \frac{1}{(n+1)(n+2)} \\ &= s_n + \frac{1}{(n+1)(n+2)} \\ &= \frac{n}{n+1} + \frac{1}{(n+1)(n+2)} && \text{(by induction)} \\ &= \frac{n^2 + 2n + 1}{(n+1)(n+2)} \\ &= \frac{(n+1)^2}{(n+1)(n+2)} \\ &= \frac{n+1}{n+2}, \end{aligned}$$

  completing the induction.

- Here's another argument. We want to use partial fractions and determine A and B so that $\frac{1}{n(n+1)} = \frac{A}{n} + \frac{B}{n+1}$ for every $n \in \mathbb{N}$. Let's solve for A and B. We have

$$\frac{A}{n} + \frac{B}{n+1} = \frac{A(n+1) + Bn}{n(n+1)} = \frac{(A+B)n + A}{n(n+1)}.$$

  Thus, if we want to choose A and B so that $\frac{1}{n(n+1)} = \frac{A}{n} + \frac{B}{n+1}$ for all $n \in \mathbb{N}^+$, then we should ensure that $A + B = 0$ and that $A = 1$. Solving these equations gives $A = 1$ and $B = -1$, and then we can check that we do indeed have $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$ for every $n \in \mathbb{N}^+$. Therefore,

$$\begin{aligned} s_n &= \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} \\ &= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right) \\ &= 1 + \left(-\frac{1}{2} + \frac{1}{2}\right) + \left(-\frac{1}{3} + \frac{1}{3}\right) + \cdots + \left(-\frac{1}{n} + \frac{1}{n}\right) - \frac{1}{n+1} \\ &= 1 - \frac{1}{n+1} \\ &= \frac{n}{n+1} \end{aligned}$$

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

We now know that $s_n = \frac{n}{n+1}$ for every $n \in \mathbb{N}$. Therefore,

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

(where the last equality follows from the fall quarter or simply by dividing top and bottom by n). Therefore, $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$.

We'll often have an interest in series that start at numbers other than 1. For instance, we'd like to consider a series like $\sum_{n=0}^{\infty} (\frac{1}{2})^n$, which should be the "infinite sum" $1 + \frac{1}{2} + \frac{1}{4} + \ldots$, or $\sum_{n=7}^{\infty} (\frac{1}{2})^n$, which should be the "infinite sum" $\frac{1}{2^7} + \frac{1}{2^8} + \frac{1}{2^9} + \ldots$. These aren't really problems, but we need to decide what we mean by the n^{th} partial sum of such a sequence.

One way to handle these sums is to simply re-index them so that they start at 1. In other words,

$$\sum_{n=7}^{\infty} \left(\frac{1}{2}\right)^n \text{ can be interpreted as shorthand for } \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n+6}$$

while

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n \text{ can be interpreted as shorthand for } \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1}.$$

Thus, when we consider $\sum_{n=7}^{\infty}(\frac{1}{2})^n$, our sequence of partial sums $\langle s_n \rangle$ would be

$$s_n = \left(\frac{1}{2}\right)^7 + \left(\frac{1}{2}\right)^8 + \cdots + \left(\frac{1}{2}\right)^{n+6}.$$

Similarly, when we consider $\sum_{n=0}^{\infty}(\frac{1}{2})^n$, our sequence of partial sums $\langle s_n \rangle$ would be

$$s_n = \left(\frac{1}{2}\right)^0 + \left(\frac{1}{2}\right)^1 + \cdots + \left(\frac{1}{2}\right)^{n-1}.$$

Another equally compelling way to handle the situation is to avoid re-indexing, and simply think of placing 0's in the positions that are not indexed. If we follow this approach, then when we consider $\sum_{n=7}^{\infty}(\frac{1}{2})^n$, our sequence of partial sums $\langle s_n \rangle$ would be $s_n = 0$ for all $n \leq 6$, $s_7 = (\frac{1}{2})^7$, $s_8 = (\frac{1}{2})^8$, and in general $s_n = (\frac{1}{2})^7 + (\frac{1}{2})^8 + \cdots + (\frac{1}{2})^n$ whenever $n \geq 7$. For a series like $\sum_{n=0}^{\infty}(\frac{1}{2})^n$ which starts at 0, our sequence of partial sums $\langle s_n \rangle$ would be $s_1 = (\frac{1}{2})^0 + (\frac{1}{2})^1$, $s_2 = (\frac{1}{2})^0 + (\frac{1}{2})^1 + (\frac{1}{2})^2$, and in general $s_n = (\frac{1}{2})^0 + (\frac{1}{2})^1 + (\frac{1}{2})^2 + \cdots + (\frac{1}{2})^n$.

The crucial thing to notice is that the two methods are equivalent in terms of whether the sequence of partial sums converges or diverges, and the value of the limit in the case of convergence, because the two sequences are shifts of each other (see Proposition 2.2.2). Hence, we can use either one, and never worry about these pesky questions again.

We now generalize the $\sum_{n=1}^{\infty}(\frac{1}{2})^n$ example to a class of series that will play a fundamental role in our study. Notice that if we take $a = \frac{1}{2}$ and $r = \frac{1}{2}$, then we do indeed obtain $\sum_{n=1}^{\infty}(\frac{1}{2})^n$.

Proposition 3.1.2 (Geometric Series)**.** *Let $a, r \in \mathbb{R}$ with $|r| < 1$. The series $\sum_{n=0}^{\infty} ar^n$ converges and*

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}.$$

Proof. Let $\langle s_n \rangle$ be the sequence of partial sums, i.e. $s_n = a + ar + ar^2 + \cdots + ar^n$ for all $n \in \mathbb{N}^+$. Using

Problem 1 on Homework 1, for any $n \in \mathbb{N}$, we have

$$\begin{aligned} s_n &= a + ar + ar^2 + \cdots + ar^n \\ &= a \cdot (1 + r + r^2 + \cdots + r^n) \\ &= a \cdot \frac{r^{n+1} - 1}{r - 1} \\ &= a \cdot \frac{1 - r^{n+1}}{1 - r}. \end{aligned}$$

Since $\lim_{n\to\infty} r^n = 0$ by Proposition 2.3.11, it follows (using either Proposition 2.2.2 or Theorem 2.2.8) that $\lim_{n\to\infty} r^{n+1} = 0$, and so

$$\lim_{n\to\infty} \frac{1 - r^{n+1}}{1 - r} = \frac{1}{1 - r}.$$

Therefore, we have

$$\begin{aligned} \lim_{n\to\infty} s_n &= \lim_{n\to\infty} a \cdot \frac{1 - r^{n+1}}{1 - r} \\ &= \frac{a}{1 - r}. \end{aligned}$$

By definition, we conclude that $\sum_{n=0}^{\infty} ar^n$ converges to $\frac{a}{1-r}$. □

Our next result establishes a fundamental implication that we can draw about the sequence $\langle a_n \rangle$ using knowledge that $\sum_{n=1}^{\infty} a_n$ converges.

Proposition 3.1.3. *If* $\sum_{n=1}^{\infty} a_n$ *converges, then* $\lim_{n\to\infty} a_n = 0$.

Proof. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$. Since $\sum_{n=1}^{\infty} a_n$ converges, we can fix $\ell \in \mathbb{R}$ such that $\lim_{n\to\infty} s_n = \ell$. Using Proposition 2.2.2, we also that $\lim_{n\to\infty} s_{n-1} = \ell$. Since $a_n = s_n - s_{n-1}$ for all $n \geq 2$, it follows that $\langle a_n \rangle$ converges and that

$$\begin{aligned} \lim_{n\to\infty} a_n &= \lim_{n\to\infty} (s_n - s_{n-1}) \\ &= \ell - \ell \\ &= 0. \end{aligned}$$

□

Here is an example of how to use Proposition 3.1.3 to argue that a series diverges. Consider the series $\sum_{n=1}^{\infty} \frac{(n+1)^2}{4n(n+2)}$. Notice that

$$\begin{aligned} \lim_{n\to\infty} \frac{(n+1)^2}{4n(n+2)} &= \lim_{n\to\infty} \frac{n^2 + 2n + 1}{4n^2 + 8n} \\ &= \lim_{n\to\infty} \frac{1 + (2/n) + (1/n^2)}{4 + (8/n)} \\ &= \frac{1 + 0 + 0}{4 + 0} \\ &= \frac{1}{4}. \end{aligned}$$

Since $\langle \frac{(n+1)^2}{4n(n+2)} \rangle$ does not converge to 0, we can use Proposition 3.1.3 to conclude that $\sum_{n=1}^{\infty} \frac{(n+1)^2}{4n(n+2)}$ diverges.

One fact that can not be emphasized enough is that the converse of Proposition 3.1.3 does not hold. That is, we can not conclude that $\sum_{n=1}^{\infty} a_n$ converges simply from the fact that $\lim_{n\to\infty} a_n = 0$. The next extremely important example shows how this can happen.

Proposition 3.1.4 (Harmonic Series)**.** *The series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges to ∞, i.e. we have $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$.*

Before jumping into the proof, let's explore the core idea. Notice that since $\frac{1}{n} > 0$ for all $n \in \mathbb{N}^+$, the sequence $\langle s_n \rangle$ of partial sums is strictly increasing. We will show that $\langle s_n \rangle$ is not bounded above. Notice first that

$$s_2 = 1 + \frac{1}{2} = \frac{3}{2}.$$

Can we find an n with $s_n \geq 2$? Notice that

$$\begin{aligned} s_4 &= s_2 + \frac{1}{3} + \frac{1}{4} \\ &> s_2 + \frac{1}{4} + \frac{1}{4} \\ &= s_2 + \frac{1}{2} \\ &= 2. \end{aligned}$$

Can we find another $\frac{1}{2}$ in the remaining terms? Notice that

$$\begin{aligned} s_8 &= s_4 + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \\ &> s_4 + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &> s_2 + \frac{1}{2} \\ &= \frac{5}{2}. \end{aligned}$$

A similar argument show that $s_{16} \geq s_8 + \frac{1}{2}$. Continuing this logic, we arrive at the following proof.

Proof. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle \frac{1}{n} \rangle$, so that $s_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ for every $n \in \mathbb{N}^+$. We show that $s_{2^n} \geq 1 + \frac{n}{2}$ for all $n \in \mathbb{N}$ by induction. Notice that $s_2 = 1 + \frac{1}{2}$, so the statement is true when $n = 1$. Suppose that $n \in \mathbb{N}^+$ and $s_{2^n} \geq 1 + \frac{n}{2}$. We then have

$$\begin{aligned} s_{2^{n+1}} &= 1 + \frac{1}{2} + \cdots + \frac{1}{2^{n+1}} \\ &= \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^n}\right) + \left(\frac{1}{2^n+1} + \frac{1}{2^n+2} + \cdots + \frac{1}{2^{n+1}}\right) \\ &= s_{2^n} + \left(\frac{1}{2^n+1} + \frac{1}{2^n+2} + \cdots + \frac{1}{2^{n+1}}\right) \\ &\geq s_{2^n} + \left(\frac{1}{2^{n+1}} + \frac{1}{2^{n+1}} + \cdots + \frac{1}{2^{n+1}}\right) \\ &\geq \left(1 + \frac{n}{2}\right) + 2^n \cdot \frac{1}{2^{n+1}} && \text{(using induction)} \\ &= 1 + \frac{n}{2} + \frac{1}{2} \\ &= 1 + \frac{n+1}{2}. \end{aligned}$$

Therefore, we have $s_{2^n} \geq 1 + \frac{n}{2}$ for all $n \in \mathbb{N}^+$.

We now show that $\lim_{n\to\infty} s_n = \infty$. Let $z \in \mathbb{R}$ be arbitrary. By Proposition 1.4.5, we can fix $N \in \mathbb{N}$ with $N > z$. For any $n \geq 2^{2N}$, we have

$$\begin{aligned} s_n &\geq s_{2^{2N}} \\ &\geq 1 + \frac{2N}{2} \\ &= 1 + N \\ &> z. \end{aligned}$$

Therefore, we have $\lim_{n\to\infty} s_n = \infty$, and hence $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. □

Our next result shows that infinite series behave well with respect to constant multiples and addition. Fortunately, we can use the results from Chapter 2 to do all of the heavy lifting.

Proposition 3.1.5.

1. *Suppose that $\sum_{n=1}^{\infty} a_n$ converges. For any $c \in \mathbb{R}$, the series $\sum_{n=1}^{\infty} c \cdot a_n$ converges and*

$$\sum_{n=1}^{\infty} c \cdot a_n = c \cdot \sum_{n=1}^{\infty} a_n.$$

2. *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge, then $\sum_{n=1}^{\infty} (a_n + b_n)$ converges and*

$$\sum_{n=1}^{\infty} (a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n.$$

Proof.

1. Let $c \in \mathbb{R}$ be arbitrary. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$, so $s_n = a_1 + a_2 + \cdots + a_n$ for every $n \in \mathbb{N}^+$. Since $\sum_{n=1}^{\infty} a_n$ converges, we know by definition that $\langle s_n \rangle$ converges and that $\lim_{n\to\infty} s_n = \sum_{n=1}^{\infty} a_n$. Notice that for every $n \in \mathbb{N}^+$, we have

$$\begin{aligned} c \cdot a_1 + c \cdot a_2 + \cdots + c \cdot a_n &= c \cdot (a_1 + a_2 + \cdots + a_n) \\ &= c \cdot s_n. \end{aligned}$$

   Hence, $\langle c \cdot s_n \rangle$ is the sequence of partial sums of $\langle c \cdot a_n \rangle$. Since $\langle s_n \rangle$ converges, we can use Proposition 2.2.3 to conclude that $\langle c \cdot s_n \rangle$ converges and that $\lim_{n\to\infty} c \cdot s_n = c \cdot \lim_{n\to\infty} s_n$. Therefore, $\sum_{n=1}^{\infty} c \cdot a_n$ converges and

$$\begin{aligned} \sum_{n=1}^{\infty} c \cdot a_n &= \lim_{n\to\infty} c \cdot s_n \\ &= c \cdot \lim_{n\to\infty} s_n \\ &= c \cdot \sum_{n=1}^{\infty} a_n. \end{aligned}$$

2. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$ and let $\langle t_n \rangle$ be the sequence of partial sums of $\langle b_n \rangle$. Since $\sum_{n=1}^{\infty} a_n$ converges, we know by definition that $\langle s_n \rangle$ converges and that $\lim_{n\to\infty} s_n = \sum_{n=1}^{\infty} a_n$. Similarly, since $\sum_{n=1}^{\infty} b_n$ converges, we know by definition that $\langle t_n \rangle$ converges and that $\lim_{n\to\infty} t_n = \sum_{n=1}^{\infty} t_n$. Notice that for every $n \in \mathbb{N}^+$, we have

$$\begin{aligned}(a_1 + b_1) + (a_2 + b_2) + \cdots + (a_n + b_n) &= (a_1 + a_2 + \cdots + a_n) + (b_1 + b_2 + \cdots + b_n) \\ &= s_n + t_n.\end{aligned}$$

Therefore, $\langle s_n + t_n \rangle$ is the sequence of partial sums of $\langle a_n + b_n \rangle$. Since $\langle s_n \rangle$ and $\langle t_n \rangle$ both converge, we can use Theorem 2.2.8 to conclude that $\langle s_n + t_n \rangle$ converges and that

$$\lim_{n\to\infty} (s_n + t_n) = \lim_{n\to\infty} s_n + \lim_{n\to\infty} t_n.$$

Therefore, $\sum_{n=1}^{\infty}(a_n + b_n)$ converges, and we have

$$\begin{aligned}\sum_{n=1}^{\infty}(a_n + b_n) &= \lim_{n\to\infty} (s_n + t_n) \\ &= \lim_{n\to\infty} s_n + \lim_{n\to\infty} t_n \\ &= \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n.\end{aligned}$$

□

Here's another result that basically says that finitely many terms don't matter.

Proposition 3.1.6. *Let $\langle a_n \rangle$ be a sequence and let $N, M \in \mathbb{N}^+$ with $N > M$. We then have that $\sum_{n=N}^{\infty} a_n$ converges if and only if $\sum_{n=M}^{\infty} a_n$ converges, and moreover in this case we have*

$$\sum_{n=M}^{\infty} a_n = (a_M + a_{M+1} + \cdots + a_{N-1}) + \sum_{n=N}^{\infty} a_n.$$

Proof. Let $N, M \in \mathbb{N}^+$ with $N > M$. Let $\langle s_n \rangle$ be the sequence of partial sums corresponding to the series $\sum_{n=M}^{\infty} a_n$, where we let $s_n = 0$ for $n < M$, and let $s_n = a_M + \cdots + a_n$ for $n \geq M$. Similarly, let $\langle t_n \rangle$ be the sequence of partial sums corresponding to the series $\sum_{n=N}^{\infty} a_n$, where we let $t_n = 0$ for $n < N$, and let $t_n = b_M + \cdots + b_n$ for $n \geq N$.

Notice that for any $n \geq N$, we have $s_n = a_M + a_{M+1} + \cdots + a_n$ and $t_n = a_N + a_{N+1} + \cdots + a_n$. Therefore, for any $n \geq N$, we have

$$t_n = (a_M + a_{M+1} + \cdots + a_{N-1}) + s_n.$$

In other words, with the exception of finitely many terms, the sequence $\langle s_n \rangle$ is the sequence $\langle t_n \rangle$ plus the constant sequence $\langle a_M + a_{M+1} + \cdots + a_{N-1} \rangle$. Using Proposition 2.2.1 and Proposition 2.2.8, we conclude that $\langle s_n \rangle$ converges if and only if $\langle t_n \rangle$ converges, and in this case we have

$$\begin{aligned}\sum_{n=M}^{\infty} a_n &= \lim_{n\to\infty} s_n \\ &= (a_M + a_{M+1} + \cdots + a_{N-1}) + \lim_{n\to\infty} t_n \\ &= (a_M + a_{M+1} + \cdots + a_{N-1}) + \sum_{n=N}^{\infty} a_n.\end{aligned}$$

□

Let $\langle a_n \rangle$ be a sequence, and let $\langle s_n \rangle$ be the sequence of partial sums. If we want to know whether or not $\sum_{n=1}^{\infty} a_n$ converges, we are asking whether the sequence $\langle s_n \rangle$ converges. If we use the fact that a sequence converges if and only if it is a Cauchy sequence, we arrive at the following.

Proposition 3.1.7. *Let $\langle a_n \rangle$ be a sequence. The following are equivalent:*

1. *The series $\sum_{n=1}^{\infty} a_n$ converges.*
2. *For all $\varepsilon > 0$ there exists $N \in \mathbb{N}^+$ such that $|a_{m+1} + a_{m+2} + \cdots + a_n| < \varepsilon$ whenever $n > m \geq N$.*

Proof. Throughout the argument, let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$.

- (1) $\Rightarrow$ (2): Suppose first that $\sum_{n=1}^{\infty} a_n$ converges. By definition, this means that the sequence $\langle s_n \rangle$ converges. Using Proposition 2.4.6, it follows that $\langle s_n \rangle$ is a Cauchy sequence. Now let $\varepsilon > 0$. Since $\langle s_n \rangle$ is a Cauchy sequence, we can fix $N \in \mathbb{N}^+$ such that $|s_n - s_m| < \varepsilon$ whenever $n, m \geq N$. Now given any $n, m \in \mathbb{N}^+$ with $n > m \geq N$, we have

$$\begin{aligned} |a_{m+1} + a_{m+2} + \cdots + a_n| &= |(a_1 + a_2 + \cdots + a_n) - (a_1 + a_2 + \cdots + a_m)| \\ &= |s_n - s_m| \\ &< \varepsilon. && \text{(since } n, m \geq N\text{)} \end{aligned}$$

- (2) $\Rightarrow$ (1): Assume (2). We will show that $\langle s_n \rangle$ is a Cauchy sequence. Let $\varepsilon > 0$. By our assumption, we can fix $N \in \mathbb{N}^+$ such that $|a_{m+1} + a_{m+2} + \cdots + a_n| < \varepsilon$ whenever $n > m \geq N$. Let $n, m \geq N$ be arbitrary. We have three cases:
  - *Case 1:* Suppose that $m = n$. We then have $|s_n - s_m| = 0 < \varepsilon$.
  - *Case 2:* Suppose that $n > m$. We then have

$$\begin{aligned} |s_n - s_m| &= |(a_1 + a_2 + \cdots + a_n) - (a_1 + a_2 + \cdots + a_m)| \\ &= |a_{m+1} + a_{m+2} + \cdots + a_n| \\ &< \varepsilon. && \text{(since } n > m \geq N\text{).} \end{aligned}$$

  - *Case 3:* Suppose that $m > n$. We then have

$$\begin{aligned} |s_n - s_m| &= |(a_1 + a_2 + \cdots + a_n) - (a_1 + a_2 + \cdots + a_m)| \\ &= |-a_{n+1} - a_{n+2} - \cdots - a_m| \\ &= |a_{n+1} + a_{n+2} + \cdots + a_m| \\ &< \varepsilon. && \text{(since } m > n \geq N\text{).} \end{aligned}$$

  Thus, in all cases, we have $|s_n - s_m| < \varepsilon$. It follows that $\langle s_n \rangle$ is a Cauchy sequence. Therefore, $\langle s_n \rangle$ converges by Theorem 2.4.9, so $\sum_{n=1}^{\infty} a_n$ converges.

□

3.2 Series of Nonnegative Terms

In the special case when a series consists of nonnegative terms, the sequence of partial sums is increasing, so in order to check for convergence, we need only determine whether the sequence of partial sums is bounded above. In other words, we have the following.

Proposition 3.2.1. *Suppose that $a_n \geq 0$ for all $n \in \mathbb{N}$. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$.*

1. *If $\langle s_n \rangle$ is bounded above, then $\sum_{n=1}^{\infty} a_n$ converges and $\sum_{n=1}^{\infty} a_n = \sup\{s_n : n \in \mathbb{N}\}$.*
2. *If $\langle s_n \rangle$ is not bounded above, then $\sum_{n=1}^{\infty} a_n = \infty$.*

In particular, $\sum_{n=1}^{\infty} a_n$ converges if and only if $\langle s_n \rangle$ is bounded above.

Proof. Notice that for any $n \in \mathbb{N}^+$, we have $s_{n+1} = s_n + a_n \geq s_n$ because $a_n \geq 0$. Therefore, $\langle s_n \rangle$ is an increasing sequence.

1. Suppose that $\langle s_n \rangle$ is bounded above. Using the Monotone Convergence Theorem, we then know that $\langle s_n \rangle$ converges and that $\lim\limits_{n\to\infty} s_n = \sup\{s_n : n \in \mathbb{N}^+\}$. Therefore, $\sum_{n=1}^{\infty} a_n$ converges and $\sum_{n=1}^{\infty} a_n = \sup\{s_n : n \in \mathbb{N}\}$.

2. Suppose that $\langle s_n \rangle$ is not bounded above. Let $z \in \mathbb{R}$ be arbitrary. Since $\langle s_n \rangle$ is not bounded above, we know that z is not an upper bound for $\{s_n : n \in \mathbb{N}^+\}$, so we can fix $N \in \mathbb{N}^+$ with $s_N > z$. Since $\langle s_n \rangle$ is increasing, we know that for any $n \geq N$, we have $s_n \geq s_N > z$. It follows that $\lim\limits_{n\to\infty} s_n = \infty$, so $\sum_{n=1}^{\infty} a_n = \infty$.

□

It may seem like we have solved the problem of determining whether a series $\sum_{n=1}^{\infty} a_n$ with nonnegative terms converges, but figuring out whether or not the sequence of partial sums is bounded above is often extremely difficult. At this point, the only tool that we currently have is to compare a given series to one we already know converges.

Proposition 3.2.2 (Comparison Test)**.** *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences. Suppose that there exists $N \in \mathbb{N}^+$ such that $0 \leq a_n \leq b_n$ for all $n \geq N$. If $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges and $\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n$.*

Proof. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$ and let $\langle t_n \rangle$ be the sequence of partial sums of $\langle b_n \rangle$. Since $a_n \leq b_n$ for every $n \in \mathbb{N}^+$, it follows by induction that $s_n \leq t_n$ for every $n \in \mathbb{N}^+$. By Proposition 3.2.1, we know that $\{t_n : n \in \mathbb{N}\}$ is bounded above. Now $s_k \leq t_k \leq \sup\{t_n : n \in \mathbb{N}\}$ for every $k \in \mathbb{N}$, hence $\sup\{t_n : n \in \mathbb{N}\}$ is an upper bound for $\{s_n : n \in \mathbb{N}\}$, and it follows that $\sup\{s_n : n \in \mathbb{N}\} \leq \sup\{t_n : n \in \mathbb{N}\}$. Using Proposition 3.2.1, we conclude that $\sum_{n=1}^{\infty} a_n$ converges and that

$$\begin{aligned} \sum_{n=1}^{\infty} a_n &= \sup\{s_n : n \in \mathbb{N}^+\} \\ &\leq \sup\{t_n : n \in \mathbb{N}^+\} \\ &= \sum_{n=1}^{\infty} b_n. \end{aligned}$$

□

For example, suppose we want to show that

$$\sum_{n=1}^{\infty} \frac{4 + \sin(n^2)}{3^n + 182n + 1}$$

converges. In this case, we do not want to handle the partial sums directly, because they are not pretty. Instead, we compare the given series to a known geometric series. Since $-1 \leq \sin x \leq 1$ for all $x \in \mathbb{R}$, and $3^n + 182n + 1 \geq 3^n$ for all $n \in \mathbb{N}^+$, it follows that

$$0 \leq \frac{4 + \sin(n^2)}{3^n + 182n + 1} \leq \frac{5}{3^n}$$

for all $n \in \mathbb{N}^+$. Now we know that $\sum_{n=1}^{\infty} \frac{5}{3^n}$ converges by Proposition 3.1.2, and in fact we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{5}{3^n} &= \sum_{n=0}^{\infty} \frac{5}{3^{n+1}} \\
&= \sum_{n=0}^{\infty} \frac{5/3}{3^n} \\
&= \sum_{n=0}^{\infty} \frac{5}{3} \cdot \left(\frac{1}{3}\right)^n \\
&= \frac{5}{3} \cdot \frac{1}{1 - (1/3)} \\
&= \frac{5}{3} \cdot \frac{3}{2} \\
&= \frac{5}{2}.
\end{aligned}$$

Therefore, our series converges by the Comparison Test, and we know that

$$\sum_{n=1}^{\infty} \frac{3 + \sin(n^2)}{2^n + 182n + 1} \leq \frac{5}{2}.$$

If we weaken the assumption that $0 \leq a_n \leq b_n$ to the assumption that *eventually* we have $0 \leq a_n \leq b_n$, then we can still conclude that $\sum_{n=1}^{\infty} a_n$ converges whenever $\sum_{n=1}^{\infty} b_n$ does, but we lose out on the resulting inequality.

Corollary 3.2.3. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences. Suppose that there exists $N \in \mathbb{N}^+$ such that $0 \leq a_n \leq b_n$ for all $n \geq N$. If $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.*

Proof. Assume that $\sum_{n=1}^{\infty} b_n$ converges. Fix $N \in \mathbb{N}^+$ such that $0 \leq a_n \leq b_n$ for all $n \geq N$. Since $\sum_{n=1}^{\infty} b_n$ converges, we can use Proposition 3.1.6, to conclude that $\sum_{n=N}^{\infty} b_n$ converges. Now we know that $0 \leq a_n \leq b_n$ for all $n \geq N$, so we can use the Comparison Test to conclude that $\sum_{n=N}^{\infty} a_n$ converges. Applying Proposition 3.1.6 again, it follows that $\sum_{n=1}^{\infty} a_n$ converges. □

What if we consider the series

$$\sum_{n=1}^{\infty} \frac{1}{2^n - 1}?$$

In this case, it feels natural to say that the terms grow at the same rate as the terms of $\sum_{n=1}^{\infty} \frac{1}{2^n}$, so we might guess that our series converges. Unfortunately, the inequality goes the wrong way! Since $2^n - 1 < 2^n$ for all $n \in \mathbb{N}^+$, we have

$$\frac{1}{2^n - 1} > \frac{1}{2^n}$$

for all $n \in \mathbb{N}^+$. As a result, we can not directly use the Comparison Test. To get around this obstacle, you might think to compare our series $\sum_{n=1}^{\infty} \frac{1}{2^n-1}$ with a slightly bigger geometric series, say $\sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^n$. You could then do some inequality work to show that

$$\frac{1}{2^n - 1} \leq \left(\frac{2}{3}\right)^n$$

for all $n \geq 2$, and then apply Corollary 3.2.3 together with Proposition 3.1.2. This works, but it's tiresome to work out the inequalities by hand. Instead, we can apply the following test, which intuitively says that if $\langle a_n \rangle$ and $\langle b_n \rangle$ are nonnegative sequences that grow at roughly the same rate, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.

Proposition 3.2.4 (Limit Comparison Test)**.** *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences such that $a_n \geq 0$ and $b_n \geq 0$ for all $n \in \mathbb{N}^+$.*

1. *If $\lim_{n\to\infty} \frac{a_n}{b_n} = c$ where $c \in \mathbb{R}$ and $c > 0$, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.*
2. *If $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.*
3. *If $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$ and $\sum_{n=1}^{\infty} a_n$ converges, then $\sum_{n=1}^{\infty} b_n$ converges.*

Proof.

1. Assume that $\lim_{n\to\infty} \frac{a_n}{b_n} = c$ where $c \in \mathbb{R}$ and $c > 0$.
   - Suppose first that $\sum_{n=1}^{\infty} b_n$ converges. Since $\lim_{n\to\infty} \frac{a_n}{b_n} = c$ and $c > 0$, we can fix $N \in \mathbb{N}^+$ such that $|\frac{a_n}{b_n} - c| < c$ for all $n \geq N$. For any $n \geq N$, we then have $\frac{a_n}{b_n} < 2c$, and hence $a_n < 2cb_n$. Since $\sum_{n=1}^{\infty} b_n$ converges, Proposition 3.1.5 tells us that $\sum_{n=1}^{\infty} 2cb_n$ also converges. Since $0 \leq a_n < 2cb_n$ for all $n \geq N$, we can use Corollary 3.2.3 to conclude that $\sum_{n=1}^{\infty} a_n$ converges.
   - Conversely, suppose that $\sum_{n=1}^{\infty} a_n$ converges. Using Theorem 2.2.8, it follows that $\lim_{n\to\infty} \frac{b_n}{a_n} = \frac{1}{c}$. Since $\frac{1}{c} > 0$, we can use what we just proved to conclude that $\sum_{n=1}^{\infty} b_n$ converges.
2. Assume that $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$ and that $\sum_{n=1}^{\infty} b_n$ converges. By definition, we can fix $N \in \mathbb{N}$ such that $|\frac{a_n}{b_n} - 0| < 1$ for all $n \geq N$. We then have that $0 \leq a_n < b_n$ for all $n \geq N$. Since $\langle b_n \rangle$ converges, we can use Corollary 3.2.3 to conclude that $\sum_{n=1}^{\infty} a_n$ converges.
3. Assume that $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$ and $\sum_{n=1}^{\infty} a_n$ converges. By Proposition 2.3.7, it follows that $\lim_{n\to\infty} \frac{b_n}{a_n} = 0$. Using part (2), we immediately conclude that $\sum_{n=1}^{\infty} b_n$ converges.

□

For example, we can use the Limit Comparison Test to show that

$$\sum_{n=1}^{\infty} \frac{1}{2^n - 1}$$

converges. As mentioned above, this series looks a lot like $\sum_{n=1}^{\infty} \frac{1}{2^n}$, so we compare the two. Notice that $\frac{1}{2^n} \geq 0$ and $\frac{1}{2^n-1} \geq 0$ for all $n \in \mathbb{N}$. We also have

$$\begin{aligned} \frac{1/(2^n - 1)}{1/2^n} &= \frac{2^n}{2^n - 1} \\ &= \frac{1}{1 - (1/2)^n} \end{aligned}$$

for all $n \in \mathbb{N}^+$. Now we know that $\lim_{n\to\infty}(\frac{1}{2})^n = 0$ by Proposition 2.3.11, so we can use Theorem 2.2.8 to conclude that $\langle \frac{1/(2^n-1)}{1/2^n} \rangle$ converges and

$$\lim_{n\to\infty} \frac{1/(2^n-1)}{1/2^n} = \frac{1}{1-0} = 1.$$

Since $\sum_{n=1}^{\infty} \frac{1}{2^n}$ converges, it follows from the Limit Comparison Test that $\sum_{n=1}^{\infty} \frac{1}{2^n-1}$ converges.

The most important class of series that we understand well is the class of geometric series. Given a geometric series $\sum_{n=0}^{\infty} ar^n$, the defining characteristic is that the quotient of two consecutive terms in constant. That is, for any $n \in \mathbb{N}$, we have

$$\frac{ar^{n+1}}{ar^n} = r.$$

Suppose now that we have a series $\sum_{n=1}^{\infty} a_n$, where $a_n > 0$ for all $n \in \mathbb{N}^+$, and we want to determine whether it converges. We might try to compare our series to a geometric series, but we are not sure which r to pick. In this case, The natural instinct is to look at the quotient of consecutive terms $\frac{a_{n+1}}{a_n}$ and see we want we observe. If these quotients approach a limit like $\frac{2}{3}$, then we conclude that a_{n+1} is approximately $\frac{2}{3}a_n$ for large values of n, and so we might think to compare it to the series $\sum_{n=0}^{\infty}(\frac{2}{3})^n$. There is a slight problem here in that if $\langle \frac{a_{n+1}}{a_n} \rangle$ converges to $\frac{2}{3}$, then we can not directly use the Comparison Test because the quotients might be slightly larger than $\frac{2}{3}$ for large values of n. However, since $\frac{2}{3} < 1$, we can pick a slightly larger value of r that is still less than $\frac{2}{3}$, such as $\frac{3}{4}$. Now if $\langle \frac{a_{n+1}}{a_n} \rangle$ converges to $\frac{2}{3}$, then for large values of n we will have $\frac{a_{n+1}}{a_n} < \frac{3}{4}$, and hence $a_{n+1} < \frac{3}{4} \cdot a_n$ for large values of n. We will then be able to compare the tail of our series with the tail of the series $\sum_{n=0}^{\infty}(\frac{2}{3})^n$, which converges. Putting these ideas together leads to the following test.

Proposition 3.2.5 (Ratio Test for Series with Nonnegative Terms)**.** *Suppose that $a_n \geq 0$ for all $n \in \mathbb{N}^+$ and that $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = r$ for some $r \in \mathbb{R}$.*

1. *If $r < 1$ then $\sum_{n=1}^{\infty} a_n$ converges.*
2. *If $r > 1$ then $\lim_{n\to\infty} a_n = \infty$, and hence $\sum_{n=1}^{\infty} a_n = \infty$.*

Proof.

1. Suppose that $r < 1$. Let $s = \frac{r+1}{2}$, and notice that $r < s < 1$. Since $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = r$, we can fix $N \in \mathbb{N}^+$ such that

$$\left| \frac{a_{n+1}}{a_n} - r \right| < s - r$$

for all $n \geq N$. Working through the inequalities, it follows that for all $n \geq N$, we have $\frac{a_{n+1}}{a_n} < s$, and hence $a_{n+1} < sa_n$. In particular, we have $a_{N+1} < sa_N$, hence

$$\begin{aligned} a_{N+2} &< sa_{N+1} \\ &< s(sa_N) \\ &= s^2 a_N, \end{aligned}$$

and so

$$\begin{aligned} a_{N+3} &< sa_{N+2} \\ &< s(s^2 a_N) \\ &= s^3 a_N, \end{aligned}$$

Using this argument, a simple induction shows that $a_{N+k} \leq s^k a_N$ for all $k \in \mathbb{N}^+$. Since $0 < s < 1$, we know that $\sum_{k=0}^{\infty} a_N s^k$ converges by Proposition 3.1.2. Using the Comparison Test, we conclude that $\sum_{k=0}^{\infty} a_{N+k}$ converges, and hence $\sum_{n=N}^{\infty} a_n$ converges (since the partial sums are the latter series are just a shift of the partial sums of the former series). Finally, Proposition 3.1.6 tells us that $\sum_{n=1}^{\infty} a_n$ converges.

2. Suppose that $r > 1$. Let $s = \frac{1+r}{2}$, and notice that $1 < s < r$. Since $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = r$, we can fix $N \in \mathbb{N}^+$ such that
$$\left| \frac{a_{n+1}}{a_n} - r \right| < r - s$$
for all $n \geq N$. We may assume, by making N larger if necessary, that $a_N > 0$ (there can be at most finitely many terms of the sequence that equal 0, since the limit of $\langle a_{n+1} a_n \rangle$ exists). Working through the inequalities, it follows that for all $n \geq N$, we have $\frac{a_{n+1}}{a_n} > s$. Arguing as above, this implies that $a_{N+k} > s^k a_N$ for all $k \in \mathbb{N}$. Since $s > 1$, we may use Proposition 2.3.11 to conclude that $\lim_{k\to\infty} s^k = \infty$. Since $a_N > 0$, Proposition 2.3.6 implies that $\lim_{k\to\infty} a_N s^k = \infty$ as well. Finally, we can use Proposition 2.3.9 to conclude that $\lim_{k\to\infty} a_{N+k} = \infty$. Therefore, $\lim_{n\to\infty} a_n = \infty$. In particular, it is not the case that $\lim_{n\to\infty} a_n = 0$, and hence $\sum_{n=1}^{\infty} a_n$ diverges by Proposition 3.1.3.

□

Here is a simple but important example of using the Ratio Test to conclude that a series converges.

Proposition 3.2.6. *For every $x \in \mathbb{R}$ with $x > 0$, the series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges.*

Proof. Let $x \in \mathbb{R}$ with $x > 0$ be arbitrary. Notice that $\frac{x^n}{n!} > 0$ for all $n \in \mathbb{N}$ and
$$\begin{aligned} \frac{x^{n+1}/(n+1)!}{x^n/n!} &= \frac{n! \cdot x^{n+1}}{(n+1)! \cdot x^n} \\ &= \frac{x}{n+1} \end{aligned}$$
for all $n \in \mathbb{N}$. Now we know that $\lim_{n\to\infty} \frac{1}{n} = 0$, so we know that $\lim_{n\to\infty} \frac{1}{n+1} = 0$ (by Proposition 2.2.2, for example). Using Proposition 2.2.3, it follows that $\lim_{n\to\infty} \frac{x}{n+1} = 0$. Therefore, we have
$$\lim_{n\to\infty} \frac{x^{n+1}/(n+1)!}{x^n/n!} = 0.$$
Using the Ratio Test above, we conclude that $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges. □

3.3 Absolute Convergence

In the previous section, we discussed a few ways to analyze the convergence of a series with nonnegative terms. What do we do with general series? One idea is to try to reduce the problem to the previous section by taking the absolute value of the terms.

Definition 3.3.1. *We say that the series $\sum_{n=1}^{\infty} a_n$ is* absolutely convergent *if $\sum_{n=1}^{\infty} |a_n|$ is convergent.*

At first, this might seem like a strange definition. Why would take the absolute value of the terms and investigate that new series? Our first result here will be that that if $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then it converges. Although this result helps to explain our interest in our new concept, we will see in this section that absolutely convergent series are the series that converge "well". In fact, we will eventually grow to appreciate absolutely convergent series as the ones that are most worthy of our attention.

Theorem 3.3.2. *Every absolutely convergent series converges.*

Proof. Suppose that $\sum_{n=1}^{\infty} a_n$ is absolutely convergent. We show that (2) of Proposition 3.1.7 holds for the series $\sum_{n=1}^{\infty} a_n$. Let $\varepsilon > 0$. Since $\sum_{n=1}^{\infty} |a_n|$ converges, we know from Proposition 3.1.7 that we can fix $N \in \mathbb{N}$ such that $||a_{m+1}| + |a_{m+2}| + \cdots + |a_n|| < \varepsilon$ whenever $n > m \geq N$. Now let $n, m \in \mathbb{N}^+$ be arbitrary with $n > m \geq N$. We then have

$$\begin{aligned} |a_{m+1} + a_{m+2} + \cdots + a_n| &\leq |a_{m+1}| + |a_{m+2}| + \cdots + |a_n| && \text{(by the Triangle Inequality)} \\ &= ||a_{m+1}| + |a_{m+2}| + \cdots + |a_n|| \\ &< \varepsilon. \end{aligned}$$

Therefore, $\sum_{n=1}^{\infty} a_n$ converges by Proposition 3.1.7. □

This theorem is useful because we know a lot about series with nonnegative terms, so given a series possibly with negative terms, one approach to show convergence is to take the absolute value of each of terms and use our tools for series of nonnegative terms to show that this new series converges, hence by the theorem the original series converges. Here's an example.

Proposition 3.3.3. *For every $x \in \mathbb{R}$, the series $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!}$ is absolutely convergent, and hence converges.*

Proof. Proposition 3.2.6 handles the case where $x > 0$. If $x = 0$, the result is clear. Suppose that $x < 0$. Since $|x| > 0$, we know that $\sum_{n=0}^{\infty} \frac{|x|^n}{n!}$ converges by Proposition 3.2.6. Since $\left|\frac{x^n}{n!}\right| = \frac{|x|^n}{n!}$ for all $n \in \mathbb{N}$, if follows that $\sum_{n=0}^{\infty} \left|\frac{x^n}{n!}\right|$ converges, i.e. that $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges absolutely. Therefore, $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges by Theorem 3.3.2. □

Unfortunately, this procedure of simply slapping absolute values on all the terms and checking convergence of the new series does not always work. We'll see below that there are series which converge but are not absolutely convergent. Nonetheless, since the case $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} > 1$ in Proposition 3.2.5 tells us not only that the series $\sum_{n=1}^{\infty} a_n$ diverges, but that it diverges *badly*, we can extend the Ratio Test to the case when some terms may be negative.

Theorem 3.3.4 (Ratio Test)**.** *Suppose that $\langle a_n \rangle$ is a sequence and that $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = r$ where $r \in \mathbb{R}$.*

1. *If $r < 1$ then $\sum_{n=1}^{\infty} a_n$ converges absolutely, and thus converges.*
2. *If $r > 1$ then $\lim_{n\to\infty} |a_n| = \infty$, so in particular $\sum_{n=1}^{\infty} a_n$ diverges.*

Proof.

1. Since $\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = r < 1$, it follows from Proposition 3.2.5 that $\sum_{n=1}^{\infty} |a_n|$ converges. Thus, $\sum_{n=1}^{\infty} a_n$ converges absolutely, and hence $\sum_{n=1}^{\infty} a_n$ converges by Theorem 3.3.2.
2. Since $\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = r > 1$, it follows from Proposition 3.2.5 that $\lim_{n\to\infty} |a_n| = \infty$. Therefore, the set $\{|a_n| : n \in \mathbb{N}^+\}$ is not bounded. In particular, it is not the case that $\lim_{n\to\infty} a_n = 0$, and hence $\sum_{n=1}^{\infty} a_n$ diverges by Proposition 3.1.3.

□

Let's return to the question of whether every convergent series is absolutely convergent. Consider the series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots$$

Since $|\frac{(-1)^{n+1}}{n}| = \frac{1}{n}$ for all $n \in \mathbb{N}^+$, and the harmonic series diverges, we conclude that the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ is *not* absolutely convergent. Nonetheless, the series above does converge. To get an intuitive sense of what is going on, consider the sequence $\langle s_n \rangle$ of partial sums of $\langle \frac{(-1)^{n+1}}{n} \rangle$. We begin with $s_1 = 1$. Next notice that $s_2 = 1 - \frac{1}{2} = \frac{1}{2}$. Now when we form $s_3 = s_2 + \frac{1}{3}$, notice that we move back in the other direction towards $s_1 = 1$, but we do not get all the way there because $\frac{1}{3} < \frac{1}{2}$. Next, s_4 will be less than s_3 but greater than s_2 because we now subtract the smaller value of $\frac{1}{4}$. In general, it appears that we have

$$\frac{1}{2} = s_2 \leq s_4 \leq s_6 \leq \dots \quad \leq \quad \dots \leq s_5 \leq s_3 \leq s_1 = 1.$$

In other words, it looks like the partial sums are bouncing back and forth, and since $\lim_{n\to\infty} \frac{(-1)^{n+1}}{n} = 0$, the jumps back and forth will become smaller and smaller. As a result, it seems that the supremum of the even s_n's will align with the infimum of the odd s_n's, and that this common value will be the limit of the sequence $\langle s_n \rangle$. We can formalize this idea for sequences that look like $\langle \frac{(-1)^{n+1}}{n} \rangle$ in the following test.

Theorem 3.3.5 (Alternating Series Test)**.** *Suppose that $\langle a_n \rangle$ is decreasing and that $a_n \geq 0$ for all $n \in \mathbb{N}^+$, i.e. suppose that $a_1 \geq a_2 \geq a_3 \geq \dots \geq 0$. Suppose also that $\lim_{n\to\infty} a_n = 0$. We then have that the series $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ converges. Furthermore, if $\langle s_n \rangle$ is the sequence of partial sums of $\langle a_n \rangle$, then*

$$s_{2k} \leq \sum_{n=1}^{\infty} (-1)^{n+1} a_n \leq s_{2m-1}$$

for any $k, m \in \mathbb{N}$.

Proof. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle (-1)^{n+1} a_n \rangle$, i.e. let

$$s_n = a_1 - a_2 + a_3 - \dots + (-1)^{n+1} a_n$$

for all $n \in \mathbb{N}^+$. We first establish three facts:

1. The sequence $\langle s_{2n-1} \rangle$ is decreasing, i.e. $s_1 \geq s_3 \geq s_5 \geq \dots$: Let $n \in \mathbb{N}^+$ be arbitrary. We have

$$\begin{aligned} s_{2(n+1)-1} &= s_{2n+1} \\ &= s_{2n-1} - a_{2n} + a_{2n+1} \\ &= s_{2n-1} - (a_{2n} - a_{2n+1}) \\ &\leq s_{2n-1} && \text{(since } a_{2n} - a_{2n+1} \geq 0\text{).} \end{aligned}$$

   Hence, $\langle s_{2n-1} \rangle$ is decreasing.

2. The sequence $\langle s_{2n} \rangle$ is increasing, i.e. $s_2 \leq s_4 \leq s_6 \leq \dots$: Let $n \in \mathbb{N}^+$ be arbitrary. We have

$$\begin{aligned} s_{2(n+1)} &= s_{2n+2} \\ &= s_{2n} + (a_{2n+1} - a_{2n+2}) \\ &\geq s_{2n} && \text{(since } a_{2n+1} - a_{2n+2} \geq 0\text{).} \end{aligned}$$

   Hence, $\langle s_{2n} \rangle$ is increasing.

3. For any $k, m \in \mathbb{N}^+$, we have $s_{2k} \leq s_{2m-1}$: Let $k, m \in \mathbb{N}^+$ be arbitrary. If $k \leq m-1$, then

$$\begin{aligned} s_{2k} &\leq s_{2(m-1)} && \text{(since } \langle s_{2n} \rangle \text{ is increasing and } k \leq m\text{)} \\ &= s_{2m-2} \\ &\leq s_{2m-2} + a_{2m-1} && \text{(since } a_{2m+1} \geq 0\text{)} \\ &= s_{2m-1} && \text{(since } (-1)^{2m} = 1\text{).} \end{aligned}$$

On the other hand, if $k \geq m$, then $m \leq k$, hence

$$\begin{aligned} s_{2m-1} &\geq s_{2k-1} && \text{(since } \langle s_{2n+1} \rangle \text{ is decreasing and } m \leq k\text{)} \\ &\geq s_{2k-1} - a_{2k} && \text{(since } a_{2k} \geq 0\text{)} \\ &= s_{2k} && \text{(since } (-1)^{2k+1} = -1\text{).} \end{aligned}$$

Since $\langle s_{2n-1} \rangle$ is decreasing and bounded below by s_2 (using (3)), we know from the Monotone Convergence Theorem that $\langle s_{2n-1} \rangle$ converges to $\ell_1 = \inf\{s_{2n-1} : n \in \mathbb{N}^+\}$. Similarly, since $\langle s_{2n} \rangle$ is increasing and bounded above by s_1 (using (3)), it converges to $\ell_2 = \sup\{s_{2n} : n \in \mathbb{N}^+\}$. Now we are assuming that $\lim_{n\to\infty} a_n = 0$, so it follows that $\lim_{n\to\infty} a_{2n} = 0$. Therefore, we have

$$0 = \lim_{n\to\infty} a_{2n} = \lim_{n\to\infty} (s_{2n} - s_{2n-1}) = \ell_2 - \ell_1,$$

and hence $\ell_1 = \ell_2$. Let ℓ be this common value. It is now a nice exercise to check that $\lim_{n\to\infty} s_n = \ell$. We conclude that $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ converges to ℓ.

Finally, since $\ell = \ell_1 = \inf\{s_{2n-1} : n \in \mathbb{N}^+\}$, we have $\ell \leq s_{2m-1}$ for every $m \in \mathbb{N}^+$, and since $\ell = \ell_2 = \sup\{s_{2n} : n \in \mathbb{N}^+\}$, we have $s_{2k} \leq \ell$ for every $k \in \mathbb{N}^+$. □

Since the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ satisfies the hypotheses of the Alternating Series Test, it is a convergent series. We saw above that $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ does not converge absolutely, so it provides a counterexample to the converse of Theorem 3.3.2. Series of this type are given a special name.

Definition 3.3.6. *We say that the series $\sum_{n=1}^{\infty} a_n$ is* conditionally convergent *if it is convergent but not absolutely convergent.*

You might think that we can rearrange the terms in a series without affecting convergence or divergence, and without affecting the limit (in the case of convergence). After all, addition is commutative, right? Of course, addition is commutative, but the definition of a series involves a limit over finite cut-offs in a particular order, so there is some wiggle room. To see some of the potential issues, consider the series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \dots$$

We just showed that this series converges. What does it converge to? By the Alternating Series Test, we know that

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

for all $k, m \in \mathbb{N}^+$. In particular, since $s_3 = \frac{5}{6}$, we know that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \leq \frac{5}{6} < 1.$$

In fact, the actual value of $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ is $\ln 2 \approx .693$, but we will need more tools to argue this fact.

Suppose that we rearrange the series so that we put two positive terms, then a negative term, then two positive terms, then a negative term, etc. In other words, we look at the series

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \frac{1}{13} + \dots$$

It is possible to show that this infinite series also converges by following logic that is similar to the proof of the Alternating Series Test. However, let's look at the partial sums $\langle s_n \rangle$ of this new series. We see something interesting if we look at every third term:

- $s_1 = 1$.
- $s_4 = 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} > s_1 = 1$ because $\frac{1}{3} - \frac{1}{2} + \frac{1}{5} > 0$.
- $s_7 = s_4 + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} > s_4 > 1$ because $\frac{1}{7} - \frac{1}{4} + \frac{1}{9} > 0$.

In general, since

$$\begin{aligned} \frac{1}{2n-1} - \frac{1}{n} + \frac{1}{2n+1} &= \frac{n(2n+1) - (2n-1)(2n+1) + n(2n+1)}{n(2n-1)(2n+1)} \\ &= \frac{1}{n(2n-1)(2n+1)} \\ &> 0 \end{aligned}$$

for every $n \in \mathbb{N}^+$, we have

$$1 = s_1 < s_4 < s_7 < s_{10} < s_{13} \leq \dots$$

In particular, the sequence of partial sums of this sequence does *not* converge to value less than 1, like the sequence of partial sums of $\langle \frac{(-1)^{n+1}}{n} \rangle$. So if you accept that this series converges (or work through the details), then it follows that we can rearrange to terms of a convergent series to obtain a new convergent series that converges to a different value. Before diving into what is going on here, let's first formally define this concept.

Definition 3.3.7. *Let $\langle a_n \rangle$ be a sequence. We say that $\langle b_n \rangle$ is a* rearrangement *of $\langle a_n \rangle$ if there exists a bijection $f \colon \mathbb{N}^+ \to \mathbb{N}^+$ such that $b_n = a_{f(n)}$ for every $n \in \mathbb{N}^+$.*

Reflecting on the above example, we can see that we are "front-loading" more positive terms in the rearrangement, which seems to be pushing up the value of the limit. In order to explore the underlying issues behind this phenomenon, and to help explain this very counterintuitive idea, we introduce the following definition and prove a relatively straightforward result.

Definition 3.3.8. *Let $\langle a_n \rangle$ be a sequence. We define two new sequences $\langle a_n^+ \rangle$ and $\langle a_n^- \rangle$ as follows. For each $n \in \mathbb{N}^+$, we let*

$$a_n^+ = \begin{cases} a_n & \text{if } a_n \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad a_n^- = \begin{cases} -a_n & \text{if } a_n \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Notice that $a_n^+ \geq 0$ and $a_n^- \geq 0$ for all $n \in \mathbb{N}^+$.

Proposition 3.3.9.

1. *The series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent if and only if both $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ converge. Furthermore, in this case we have that $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-$.*

2. *If* $\sum_{n=1}^{\infty} a_n^+ = \infty$ *and* $\sum_{n=1}^{\infty} a_n^-$ *converges, then* $\sum_{n=1}^{\infty} a_n = \infty$.

3. *If* $\sum_{n=1}^{\infty} a_n^+$ *converges and* $\sum_{n=1}^{\infty} a_n^- = \infty$, *then* $\sum_{n=1}^{\infty} a_n = -\infty$.

Proof.

1. Suppose first that $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, so that $\sum_{n=1}^{\infty} |a_n|$ converges. Since $0 \leq a_n^+ \leq |a_n|$ and $0 \leq a_n^- \leq |a_n|$, it follows from the Comparison Test that both $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ converge.

   Suppose conversely that both $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ converge. Since $|a_n| = a_n^+ + a_n^-$ for every $n \in \mathbb{N}^+$, we can use Proposition 3.1.5 to conclude that $\sum_{n=1}^{\infty} |a_n|$ converges, hence $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

   Finally, suppose that we are in this situation. Since $a_n = a_n^+ - a_n^-$ for every $n \in \mathbb{N}^+$ and both $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ converge, Proposition 3.1.5 tells us that $\sum_{n=1}^{\infty} a_n$ converges and that

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-.$$

2. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n^+ \rangle$ and let $\langle t_n \rangle$ be the sequence of partial sums of $\langle a_n^- \rangle$. Since $a_n = a_n^+ - a_n^-$ for all $n \in \mathbb{N}^+$, it follows that $\langle s_n - t_n \rangle$ is the sequence of partial sums of $\langle a_n \rangle$. Since $\lim_{n\to\infty} s_n = \infty$ and $\langle t_n \rangle$ converges, we can use Proposition 2.3.5 to conclude that $\lim_{n\to\infty} (s_n - t_n) = \infty$. Therefore, we have $\sum_{n=1}^{\infty} a_n = \infty$.

3. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n^+ \rangle$ and let $\langle t_n \rangle$ be the sequence of partial sums of $\langle a_n^- \rangle$. As above, $\langle s_n - t_n \rangle$ is the sequence of partial sums of $\langle a_n \rangle$. Since $\langle s_n \rangle$ converges and $\lim_{n\to\infty} t_n = \infty$,we can use Proposition 2.3.5 to conclude that $\lim_{n\to\infty} (s_n - t_n) = -\infty$. Therefore, $\sum_{n=1}^{\infty} a_n = -\infty$.

□

Corollary 3.3.10. *Suppose that* $\sum_{n=1}^{\infty} a_n$ *is conditionally convergent. We then have that both* $\sum_{n=1}^{\infty} a_n^+ = \infty$ *and* $\sum_{n=1}^{\infty} a_n^- = \infty$.

Proof. Immediate from Theorem 3.3.9. □

We can use this result to help explain what is going with our rearrangement of the series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \dots$$

If we let $a_n = \frac{(-1)^{n+1}}{n}$, then we know from Corollary 3.3.10 that both $\sum_{n=1}^{\infty} a_n^+ = \infty$ and $\sum_{n=1}^{\infty} a_n^- = \infty$. Thus, the only reason that $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ converges is because the positive and negative terms cancel out these huge excesses in either direction. But using the fact that $\sum_{n=1}^{\infty} a_n^+ = \infty$, we can"front-load" positive terms of the series in order to make the partial sums large, and then throw in the negative terms here and there to disturb and skew this perfect balance. In fact, by following through with this idea, we can show that it is possible to rearrange the terms of any conditionally convergent series to achieve any limit we would like.

Theorem 3.3.11. *Suppose that* $\sum_{n=1}^{\infty} a_n$ *is conditionally convergent. For every* $x \in \mathbb{R}$, *there exists a rearrangement* $\langle b_n \rangle$ *of* $\langle a_n \rangle$ *such that* $\sum_{n=1}^{\infty} b_n = x$.

Proof. Let $x \in \mathbb{R}$ be arbitrary. By Corollary 3.3.10, we have that both $\sum_{n=1}^{\infty} a_n^+ = \infty$ and $\sum_{n=1}^{\infty} a_n^- = \infty$.

We begin by defining two strictly increasing sequences of positive natural numbers $c_1 < c_2 < c_3 < \ldots$ and $d_1 < d_2 < d_3 < \ldots$ recursively as follows. Let c_1 be the least natural number such that $\sum_{n=1}^{c_1} a_n^+ > x$ (such a c_1 exists because $\sum_{n=1}^{\infty} a_n^+ = \infty$). Given c_1, let d_1 be the least natural number such that $\sum_{n=1}^{c_1} a_n^+ - \sum_{n=1}^{d_1} a_n^- < x$ (such a d_1 exists because $\sum_{n=1}^{\infty} a_n^- = \infty$). Suppose that $k \in \mathbb{N}^+$ and we've already defined c_k and d_k. Let c_{k+1} be the least natural number greater that c_k such that $\sum_{n=1}^{c_{k+1}} a_n^+ - \sum_{n=1}^{d_k} a_n^- > x$. Next, let d_{k+1} be the least natural number greater that d_k such that $\sum_{n=1}^{c_{k+1}} a_n^+ - \sum_{n=1}^{d_{k+1}} a_n^- < x$. This completes the description of the sequences $\langle c_k \rangle$ and $\langle d_k \rangle$.

We now define our rearrangement be $\langle b_n \rangle$ of $\langle a_n \rangle$. Begin the sequence $\langle b_n \rangle$ by taking the nonnegative elements of $\langle a_n \rangle$ from $n = 1$ to $n = c_1$ in order, and then follow them by taking the negative elements of $\langle a_n \rangle$ from $n = 1$ to $n = d_1$ in order. Once we have included the nonnegative terms through c_k and the negative terms through d_k, continue our rearrangement by first taking the nonnegative elements of $\langle a_n \rangle$ from $n = c_k + 1$ to $n = c_{k+1}$ in order, and then follow them by taking the negative elements of $\langle a_n \rangle$ from $n = d_k + 1$ to $n = d_{k+1}$ in order. In this way, we define a rearrangement $\langle b_n \rangle$ of $\langle a_n \rangle$.

We need only verify that $\sum_{n=1}^{\infty} b_n = x$. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle b_n \rangle$. Since $\sum_{n=1}^{\infty} a_n$ converges, we know that $\lim_{n\to\infty} a_n = 0$. The key idea is that since we always took the *least* choice of c_k (resp. d_k) that put us above (resp. below) x, and the sequence $\langle a_n \rangle$ converges to 0, we eventually only overshoot and undershoot x by very small amounts. The details are left as exercise. □

The central idea in the above argument was to exploit the fact that both $\sum_{n=1}^{\infty} a_n^+ = \infty$ and $\sum_{n=1}^{\infty} a_n^- = \infty$ in order to offset the balance and achieve any sum we wanted. A similar argument show that if $\sum_{n=1}^{\infty} a_n$ is conditionally convergent, then we can form a rearrangement that diverges. In fact, we can make it diverge to ∞, diverge to $-\infty$, or oscillate however we would like.

At this point, you might think that this whole series thing is not worth studying. Before throwing in the towel, notice that these ideas do not seem to work for absolutely convergent series, since Proposition 3.3.9 tells us that both $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ converge in this case. In fact, we now argue that we can rearrange the terms of an absolutely convergent series in any manner whatsoever without disturbing convergence and without affecting the limit.

Proposition 3.3.12. *Suppose that $\sum_{n=1}^{\infty} a_n$ is absolutely convergent. For every rearrangement $\langle b_n \rangle$ of $\langle a_n \rangle$, we have that $\sum_{n=1}^{\infty} b_n$ is absolutely convergent and $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} a_n$.*

Proof. Let $\langle b_n \rangle$ be a rearrangement of $\langle a_n \rangle$, and fix a bijection $f : \mathbb{N}^+ \to \mathbb{N}^+$ with $b_n = a_{f(n)}$ for all $n \in \mathbb{N}^+$. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$ and let $\langle t_n \rangle$ be the sequence of partial sums of $\langle b_n \rangle$. Thus, we have

$$\begin{aligned} t_n &= b_1 + b_2 + \cdots + b_n \\ &= a_{f(1)} + a_{f(2)} + \cdots + a_{f(n)} \end{aligned}$$

for every $n \in \mathbb{N}^+$.

We first show that $\sum_{n=1}^{\infty} b_n$ is absolutely convergent. Since $\sum_{n=1}^{\infty} |a_n|$ converges, we know from Proposition 3.2.1 that $\{|a_1| + |a_2| + \cdots + |a_n| : n \in \mathbb{N}^+\}$ is bounded above. Now given any $m \in \mathbb{N}^+$, if we let $k = \max\{f(1), f(2), \ldots, f(m)\}$, then

$$\begin{aligned} |b_1| + |b_2| + \cdots + |b_m| &= |a_{f(1)}| + |a_{f(2)}| + \cdots + |a_{f(m)}| \\ &\leq |a_1| + |a_2| + \cdots + |a_k| \\ &\leq \sup\{|a_1| + |a_2| + \cdots + |a_n| : n \in \mathbb{N}^+\}. \end{aligned}$$

Therefore, $\{|b_1| + |b_2| + \cdots + |b_m| : m \in \mathbb{N}^+\}$ is also bounded above. Using Proposition 3.2.1 again, it follows that $\sum_{n=1}^{\infty} |b_n|$ converges, which means that $\sum_{n=1}^{\infty} b_n$ is absolutely convergent.

Let $\ell = \sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} s_n$. We need to show that $\lim_{n\to\infty} t_n = \ell$. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = \ell$, we can fix $N_1 \in \mathbb{N}^+$ such that $|s_n - \ell| < \frac{\varepsilon}{2}$ for all $n \geq N_1$. Since $\sum_{n=1}^{\infty} |a_n|$ converges, we can use Proposition 3.1.7 to fix $N_2 \in \mathbb{N}^+$ such that whenever $k > m \geq N_2$, we have

$$|a_{m+1}| + |a_{m+2}| + \cdots + |a_k| < \frac{\varepsilon}{2}.$$

Let $M = \max\{N_1, N_2\}$, and let

$$N = \max\{f^{-1}(1), f^{-1}(2), \ldots, f^{-1}(M)\}.$$

Now let $n \geq N$ be arbitrary. Notice that

$$\begin{aligned} |t_n - s_M| &= |(b_1 + b_2 + \cdots + b_n) - (a_1 + a_2 + \cdots + a_M)| \\ &= |a_{f(1)} + a_{f(2)} + \cdots + a_{f(n)} - a_1 - a_2 - \cdots - a_M)|. \end{aligned}$$

Since $n \geq f^{-1}(i)$ for each $i \in \{1, 2, \ldots, M\}$, each of the negated terms above is canceled by a corresponding positive term, so the sum inside the absolute value can be viewed as a finite sum of terms of the form a_i, where each of i's is at least $M + 1 \geq N_2 + 1$. Using the triangle inequality and the fact that

$$|a_{m+1}| + |a_{m+2}| + \cdots + |a_k| < \frac{\varepsilon}{2}.$$

whenever $k > m \geq N_2$, we conclude that $|t_n - s_M| < \frac{\varepsilon}{2}$. Therefore, we have

$$\begin{aligned} |t_n - \ell| &= |t_n - s_M + s_M - \ell| \\ &\leq |t_n - s_M| + |s_M - \ell| \\ &< \frac{\varepsilon}{2} + |s_M - \ell| && \text{(from above)} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } M \geq N_1\text{)} \\ &= \varepsilon. \end{aligned}$$

We conclude that $\lim_{n\to\infty} t_n = \ell$, so $\sum_{n=1}^{\infty} b_n = \ell = \sum_{n=1}^{\infty} a_n$. □

The next question to tackle is how to multiply series. You might think "Why not just multiply two series term-by-term?" so that the product of $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ should be $\sum_{n=1}^{\infty} a_n \cdot b_n$. However, if a series is supposed to be a formalization of an infinite sum, then intuitively we should be able to do a massive distributive law and arrive at the following:

$$\begin{aligned} \left(\sum_{n=1}^{\infty} a_n\right) \cdot \left(\sum_{n=1}^{\infty} b_n\right) &= (a_1 + a_2 + a_3 + \dots) \cdot (b_1 + b_2 + b_3 + \dots) \\ &= a_1(b_1 + b_2 + b_3 + \dots) + a_2(b_1 + b_2 + b_3 + \dots) + a_3(b_1 + b_2 + b_3 + \dots) + \dots \\ &= a_1b_1 + a_1b_2 + a_1b_3 + \cdots + a_2b_1 + a_2b_2 + a_2b_3 + \cdots + a_3b_1 + a_3b_2 + a_3b_3 + \cdots + \dots \end{aligned}$$

The final line is a bit mysterious because the terms are not arranged in one infinite list, as the definition of a series requires. How can we overcome this obstacle? Somehow, we want to add up all the terms in the following table:

$$\begin{array}{cccc} a_1b_1 & a_1b_2 & a_1b_3 & \dots \\ a_2b_1 & a_2b_2 & a_2b_3 & \dots \\ a_3b_1 & a_3b_2 & a_3b_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

Aha! We do know how to arrange a table like this into a list from our work on countability. We can simply walk along the finite diagonals to form the following infinite list:

$$a_1b_1 \quad a_1b_2 \quad a_2b_1 \quad a_1b_3 \quad a_2b_2 \quad a_3b_1 \quad \dots$$

Although we could take the above terms as the terms of a new series, it is a bit cumbersome to work with, since writing a formula for the n^{th} term is tricky. Moreover, when we work with power series later, it will be convenient and more natural to arrange the terms differently. The idea is to add the terms along the n^{th} diagonal and put this as the n^{th} term of the new sequence. In other words, we can form the the following sequence"

$$a_1b_1 \quad (a_1b_2 + a_2b_1) \quad (a_1b_3 + a_2b_2 + a_3b_1) \quad \cdots$$

Let's turn this idea into a formal definition.

Definition 3.3.13. *Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences. We define a new sequence $\langle c_n \rangle$, called the* Cauchy product *of $\langle a_n \rangle$ and $\langle b_n \rangle$, by letting*

$$\begin{aligned} c_n &= \sum_{k=1}^{n} a_k b_{n+1-k} \\ &= a_1b_n + a_2b_{n-1} + \cdots + a_nb_1 \end{aligned}$$

for every $n \in \mathbb{N}^+$.

In a perfect world, whenever the two series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge, and we consider the Cauchy product $\langle c_n \rangle$ of $\langle a_n \rangle$ and $\langle b_n \rangle$, then $\sum_{n=1}^{\infty} c_n$ would converge and we would have

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} a_n \cdot \sum_{n=1}^{\infty} b_n.$$

Unfortunately, the world is far from perfect. To see a counterexample, consider the sequences $\langle a_n \rangle$ and $\langle b_n \rangle$ defined by letting

$$a_n = \frac{(-1)^{n+1}}{\sqrt{n}} = b_n$$

for every $n \in \mathbb{N}^+$. Using the Alternating Series Test, it is straightforward to check that $\sum_{n=1}^{\infty} a_n$ (and hence also $\sum_{n=1}^{\infty} b_n$) converges. Let $\langle c_n \rangle$ be the Cauchy product of $\langle a_n \rangle$ and $\langle b_n \rangle$. We claim that $\sum_{n=1}^{\infty} c_n$ diverges. To see this, notice that for any $n \in \mathbb{N}^+$, we have

$$\begin{aligned} |c_n| &= |a_1b_n + a_2b_{n-1} + \cdots + a_nb_1| \\ &= \left| \frac{(-1)^2}{\sqrt{1}} \cdot \frac{(-1)^{n+1}}{\sqrt{n}} + \frac{(-1)^3}{\sqrt{2}} \cdot \frac{(-1)^n}{\sqrt{n-1}} + \cdots + \frac{(-1)^{n+1}}{\sqrt{n}} \cdot \frac{(-1)^2}{\sqrt{1}} \right| \\ &= \left| (-1)^{n+3} \cdot \left(\frac{1}{\sqrt{1} \cdot \sqrt{n}} + \frac{1}{\sqrt{2} \cdot \sqrt{n-1}} + \cdots + \frac{1}{\sqrt{n} \cdot \sqrt{1}} \right) \right| \\ &= \frac{1}{\sqrt{1} \cdot \sqrt{n}} + \frac{1}{\sqrt{2} \cdot \sqrt{n-1}} + \cdots + \frac{1}{\sqrt{n} \cdot \sqrt{1}} \\ &\geq \frac{1}{\sqrt{n} \cdot \sqrt{n}} + \frac{1}{\sqrt{n} \cdot \sqrt{n}} + \cdots + \frac{1}{\sqrt{n} \cdot \sqrt{n}} \\ &= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} \\ &= 1. \end{aligned}$$

Since $|c_n| \geq 1$ for all $n \in \mathbb{N}^+$, it is certainly not the case that $\langle c_n \rangle$ converges to 0, so $\sum_{n=1}^{\infty} c_n$ diverges by Proposition 3.1.3.

Fortunately, if we assume absolute convergence, rather than just convergence, then all is right with the world. We will state and prove this result below in Proposition 3.3.15. The key step will be relating the product of the partial sums of $\sum_{n=1}^{\infty} a_n$ of $\sum_{n=1}^{\infty} b_n$ with the partial sums of the Cauchy product. We pull out this key connection in the following lemma.

Lemma 3.3.14. *Suppose that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are both absolutely convergent. Let $\langle c_n \rangle$ be the Cauchy product of $\langle a_n \rangle$ and $\langle b_n \rangle$. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$, let $\langle t_n \rangle$ be the sequence of partial sums of $\langle b_n \rangle$, and let $\langle u_n \rangle$ be the sequence of partial sums of $\langle c_n \rangle$. We then have that $\langle s_n \cdot t_n - u_n \rangle$ converges to 0.*

Proof. Let $\alpha = \sum_{n=1}^{\infty} |a_n|$ and let $\beta = \sum_{n=1}^{\infty} |b_n|$. Notice that if $\alpha = 0$, then $a_n = 0$ for all $n \in \mathbb{N}^+$, so both $s_n = 0$ and $u_n = 0$ for all $n \in \mathbb{N}^+$, and hence $\langle s_n \cdot t_n - u_n \rangle$ is the constant zero sequence. A similar argument applies if $\beta = 0$. Therefore, we can assume that $\alpha > 0$ and $\beta > 0$.

Notice that for every $n \in \mathbb{N}^+$, we have

$$\begin{aligned} u_n &= c_1 + c_2 + \cdots + c_n \\ &= (a_1 b_1) + (a_1 b_2 + a_2 b_1) + \cdots + (a_1 b_n + a_2 b_{n-1} + \cdots + a_n b_1) \\ &= a_1 \cdot (b_1 + \cdots + b_n) + a_2 \cdot (b_1 + \cdots + b_{n-1}) + \cdots + a_{n-1} \cdot (b_1 + b_2) + a_n \cdot b_1 \end{aligned}$$

and

$$\begin{aligned} s_n \cdot t_n &= (a_1 + a_2 + \cdots + a_n)(b_1 + b_2 + \cdots + b_n) \\ &= a_1 \cdot (b_1 + \cdots + b_n) + a_2 \cdot (b_1 + \cdots + b_n) + \cdots + a_{n-1} \cdot (b_1 + \cdots + b_n) + a_n \cdot (b_1 + \cdots + b_n), \end{aligned}$$

so

$$\begin{aligned} |s_n \cdot t_n - u_n| &= |a_2 \cdot b_n + a_3 \cdot (b_{n-1} + b_n) + \cdots + a_{n-1} \cdot (b_3 + \cdots + b_n) + a_n \cdot (b_2 + \cdots + b_n)| \\ &\leq |a_2| \cdot |b_n| + |a_3| \cdot |b_{n-1}| + |a_3| \cdot |b_n| + \cdots + |a_n| \cdot |b_2| + \cdots + |a_n| \cdot |b_n| \end{aligned}$$

The key idea now is to notice that for every term $|a_i| \cdot |b_j|$ in the above sum, we have $i + j \geq n + 2$, hence either $i > \frac{n}{2}$ or $j > \frac{n}{2}$. Therefore, if $n, m \in \mathbb{N}^+$ and $n \geq 2m$, we have

$$|s_n \cdot t_n - u_n| \leq \left(\sum_{k=1}^{n} |a_k| \right) \cdot \left(\sum_{k=m+1}^{n} |b_k| \right) + \left(\sum_{k=1}^{n} |b_k| \right) \cdot \left(\sum_{k=m+1}^{n} |a_k| \right)$$

Now let $\varepsilon > 0$ be arbitrary. Since $\sum_{n=1}^{\infty} |a_n|$ converges, we can use Proposition 3.1.7 to fix $N_1 \in \mathbb{N}^+$ such that whenever $n > m \geq N_1$, we have

$$\sum_{k=m+1}^{n} |a_k| < \frac{\varepsilon}{2\beta}.$$

Similarly, since $\sum_{n=1}^{\infty} |b_n|$ converges, we can use Proposition 3.1.7 to fix $N_2 \in \mathbb{N}^+$ such that whenever $n > m \geq N_2$, we have

$$\sum_{k=m+1}^{n} |b_k| < \frac{\varepsilon}{2\alpha}.$$

Let $M = \max\{N_1, N_2\}$ and let $N = 2M$. For any $n \geq N$, we have $n \geq 2M > M$, so

$$\begin{aligned}
|s_n \cdot t_n - u_n| &\leq \left(\sum_{k=1}^{n} |a_k|\right) \cdot \left(\sum_{k=M+1}^{n} |b_k|\right) + \left(\sum_{k=1}^{n} |b_k|\right) \cdot \left(\sum_{k=M+1}^{n} |a_k|\right) && \text{(from above)} \\
&\leq \alpha \cdot \left(\sum_{k=M+1}^{n} |b_k|\right) + \beta \cdot \left(\sum_{k=M+1}^{n} |a_k|\right) \\
&< \alpha \cdot \frac{\varepsilon}{2\alpha} + \beta \cdot \frac{\varepsilon}{2\beta} && \text{(since } n > M \geq \max\{N_1, N_2\}\text{)} \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

□

Proposition 3.3.15. *Suppose that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are both absolutely convergent. Let $\langle c_n \rangle$ be the Cauchy product of $\langle a_n \rangle$ and $\langle b_n \rangle$. We then have that $\sum_{n=1}^{\infty} c_n$ is absolutely convergent and*

$$\sum_{n=1}^{\infty} c_n = \left(\sum_{n=1}^{\infty} a_n\right) \cdot \left(\sum_{n=1}^{\infty} b_n\right).$$

Proof. Let $\langle s_n \rangle$ be the sequence of partial sums of $\langle a_n \rangle$, let $\langle t_n \rangle$ be the sequence of partial sums of $\langle b_n \rangle$, and let $\langle u_n \rangle$ be the sequence of partial sums of $\langle c_n \rangle$. Let $\ell = \lim\limits_{n\to\infty} s_n = \sum_{n=1}^{\infty} a_n$ and let $m = \lim\limits_{n\to\infty} t_n = \sum_{n=1}^{\infty} b_n$. Finally, let $\alpha = \sum_{n=1}^{\infty} |a_n|$ and let $\beta = \sum_{n=1}^{\infty} |b_n|$. As in the proof of the above lemma, we may assume that both $\alpha > 0$ and $\beta > 0$ because the result is trivial if either of these is equal to 0 (as this implies that either $a_n = 0$ for all $n \in \mathbb{N}^+$, or $b_n = 0$ for all $n \in \mathbb{N}^+$).

We first show that $\sum_{n=1}^{\infty} c_n$ is absolutely convergent. For every $n \in \mathbb{N}^+$, we have

$$\begin{aligned}
|c_1| + |c_2| + \cdots + |c_n| &= |a_1 b_1| + |a_1 b_2 + a_2 b_1| + \cdots + |a_1 b_n + a_2 b_{n-1} + \cdots + a_n b_1| \\
&\leq |a_1| \cdot |b_1| + |a_1| \cdot |b_2| + |a_2| \cdot |b_1| + \cdots + |a_1| \cdot |b_n| + |a_2| \cdot |b_{n-1}| + \cdots + |a_n| \cdot |b_1| \\
&= |a_1| \cdot (|b_1| + \cdots + |b_n|) + |a_2| \cdot (|b_1| + \cdots + |b_{n-1}|) + \cdots + |a_n| \cdot |b_1| \\
&\leq |a_1| \cdot (|b_1| + \cdots + |b_n|) + |a_2| \cdot (|b_1| + \cdots + |b_n|) + \cdots + |a_n| \cdot (|b_1| + \cdots + |b_n|) \\
&= (|a_1| + |a_2| + \cdots + |a_n|) \cdot (|b_1| + |b_2| + \cdots + |b_n|) \\
&\leq \alpha \cdot \beta.
\end{aligned}$$

It follows that $\{|c_1| + |c_2| + \cdots + |c_n| : n \in \mathbb{N}^+\}$ is bounded above. Therefore, $\sum_{n=1}^{\infty} |c_n|$ converges by Proposition 3.2.1, which means that $\sum_{n=1}^{\infty} c_n$ is absolutely convergent.

To complete the argument, we need to show that $\lim\limits_{n\to\infty} u_n = \ell \cdot m$. Let $\varepsilon > 0$. Since $\lim\limits_{n\to\infty} s_n = \ell$ and $\lim\limits_{n\to\infty} t_n = m$, we know from Theorem 2.2.8 that $\lim\limits_{n\to\infty} s_n \cdot t_n = \ell \cdot m$. Thus, we can fix $N_1 \in \mathbb{N}^+$ such that $|s_n \cdot t_n - \ell \cdot m| < \frac{\varepsilon}{2}$ for all $n \geq N_1$. Using Lemma 3.3.14, we can fix $N_2 \in \mathbb{N}^+$ such that $|s_n \cdot t_n - u_n| < \frac{\varepsilon}{2}$ for all $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. For any $n \geq N$, we have

$$\begin{aligned}
|u_n - \ell \cdot m| &= |u_n - s_n \cdot t_n + s_n \cdot t_n - \ell \cdot m| \\
&\leq |u_n - s_n \cdot t_n| + |s_n \cdot t_n - \ell \cdot m| \\
&< |u_n - s_n \cdot t_n| + \frac{\varepsilon}{2} && \text{(since } n \geq N \geq N_1\text{)} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } n \geq N \geq N_2\text{)} \\
&= \varepsilon.
\end{aligned}$$

Therefore, $\lim_{n\to\infty} u_n = \ell \cdot m$, and hence

$$\sum_{n=1}^{\infty} c_n = \ell \cdot m = \left(\sum_{n=1}^{\infty} a_n\right) \cdot \left(\sum_{n=1}^{\infty} b_n\right).$$

□

Chapter 4

Topology of the Real Line

Before moving on to discuss analytic properties of functions from $\mathbb{R}$ to $\mathbb{R}$, we first spend some time studying subsets of $\mathbb{R}$. The subsets that you might be most familiar with are the open intervals and the closed intervals. However, there are many other interesting and exotic subsets of $\mathbb{R}$. A relatively simple example of such a subset is $\mathbb{Q}$. Although this is a countable subset of $\mathbb{R}$, we know from the Density of $\mathbb{Q}$ in $\mathbb{R}$ that every real is "near" an element of this subset. A more complicated, and somewhat counterintuitive, subset of $\mathbb{R}$ is the Cantor set. We will wait to introduce it, but the Cantor set is an uncountable subset of $\mathbb{R}$ that feels "tiny". In order to get a handle on subsets, we begin by introducing some core topological concepts.

4.1 Interiors and Closures

Definition 4.1.1. *Given $a \in \mathbb{R}$ and $\varepsilon > 0$, we define*

$$V_\varepsilon(a) = \{x \in \mathbb{R} : |x - a| < \varepsilon\},$$

and we call $V_\varepsilon(a)$ the ε-neighborhood of a.

Notice that $V_\varepsilon(a)$ is just different notation for the open interval $(a - \varepsilon, a + \varepsilon)$. Thus, you should think about an ϵ-neighborhood of a as "rounding out" the point a to a very small open interval.

Definition 4.1.2. *Let $A \subseteq \mathbb{R}$ and let $b \in \mathbb{R}$.*

- *We say that b is an* interior point *of A if there exists $\varepsilon > 0$ with $V_\varepsilon(b) \subseteq A$.*
- *We say that b is a* closure point *of A if for all $\varepsilon > 0$, the set $V_\varepsilon(b) \cap A$ is nonempty.*
- *We say that b is a* limit point *of A if for all $\varepsilon > 0$, the set $V_\varepsilon(b) \cap A$ contains at least one point distinct from b.*

Intuitively, a point b is an interior point of A is it sits "comfortably" inside A because we can fatten up b to a whole (potentially very small) interval of points around b that completely sits inside A. A point b is a closure point of A if it is very friendly with points of A in the sense that there are points of A that are arbitrarily close to b. Finally, the difference between a limit point and a closure point is that we do not allow the situation where b is a hermit, i.e. we require that every neighborhood of b contains points of A, but we do not allow b to serve as such a "close" point.

Notation 4.1.3. *Let $A \subseteq \mathbb{R}$.*

- *We let* $int(A) = \{b \in \mathbb{R} : b$ *is an interior point of* $A\}$*, and call* $int(A)$ *the* interior *of* A*. Some sources use the notation* A° *for the interior of* A*.*
- *We let* $cl(A) = \{b \in \mathbb{R} : b$ *is a closure point of* $A\}$*, and call* $cl(A)$ *the* closure *of* A*. Some sources use the notation* $\overline{A}$ *for the closure of* A*.*

To get a feel for these definitions, let's consider some examples. Consider the set

$$A = \{x \in \mathbb{R} : 0 \leq x < 1\} \cup \{2\} \cup \left\{4 - \frac{1}{n} : n \in \mathbb{N}^+\right\}.$$

Although we will not work through careful proofs, we have the following:

- $\text{int}(A) = (0, 1)$: Intuitively, we can expand any element of $(0, 1)$ to a neighborhood that will still be a subset of A, but this will not work for 0 or any of the other points.
- $\text{cl}(A) = [0, 1] \cup \{2, 4\} \cup \left\{4 - \frac{1}{n} : n \in \mathbb{N}^+\right\}$: All of the points of A are closure points, and now we include both 1 and 4 because there are points of A that are arbitrarily close to these.
- $\{b \in \mathbb{R} : b$ is a limit point of $A\} = [0, 1] \cup \{4\}$: Here we can find points of A arbitrarily close to 4 that are distinct from 4, but the same can not be said of the points 2 and $4 - \frac{1}{n}$.

If we instead consider the subset $\mathbb{Q} \subseteq \mathbb{R}$, we obtain the following:

- $\text{int}(\mathbb{Q}) = \emptyset$: We know that every open interval contains irrationals.
- $\text{cl}(\mathbb{Q}) = \mathbb{R}$: Given any $b \in \mathbb{R}$ and any $\varepsilon > 0$, we know that the open interval $V_\varepsilon(b)$ contains a rational by the Density of $\mathbb{Q}$ in $\mathbb{R}$.
- $\{b \in \mathbb{R} : b$ is a limit point of $\mathbb{Q}\} = \mathbb{R}$: Using the Density of $\mathbb{Q}$ and $\mathbb{R}$ a couple of times, given any $b \in \mathbb{R}$ and any $\varepsilon > 0$, we know that the open interval $V_\varepsilon(b)$ contains several rationals, so will contain a rational distinct from b.

In order to be able to work with these concepts and prove some of the above computations more easily, we start with some simple properties of the interior and closure.

Proposition 4.1.4. *For every* $A \subseteq \mathbb{R}$*, we have* $int(A) \subseteq A \subseteq cl(A)$*.*

Proof. Let $b \in \text{int}(A)$ be arbitrary. As b is an interior point of A, we can fix $\varepsilon > 0$ with $V_\varepsilon(b) \subseteq A$. Since $b \in V_\varepsilon(b)$ trivially, it follows that $b \in A$.

Now let $a \in A$ be arbitrary. For any $\varepsilon > 0$, we have $a \in V_\varepsilon(a)$, so $a \in V_\varepsilon(b) \cap A$, and hence $V_\varepsilon(b) \cap A \neq \emptyset$. Thus, $a \in \text{cl}(A)$. □

Proposition 4.1.5. *If* $A \subseteq B \subseteq \mathbb{R}$*, then* $int(A) \subseteq int(B)$ *and* $cl(A) \subseteq cl(B)$*.*

Proof. Let $A \subseteq B \subseteq \mathbb{R}$.

- We first show that $\text{int}(A) \subseteq \text{int}(B)$. Let $c \in \text{int}(A)$ be arbitrary. By definition, we can fix $\varepsilon > 0$ such that $V_\varepsilon(c) \subseteq A$. Since $A \subseteq B$, we then have $V_\varepsilon(c) \subseteq B$.
- We next show that $\text{cl}(A) \subseteq \text{cl}(B)$. Let $c \in \text{cl}(A)$ be arbitrary. Let $\varepsilon > 0$. Since $c \in \text{cl}(A)$, we know that $V_\varepsilon(c) \cap A \neq \emptyset$, so we can fix $d \in V_\varepsilon(c) \cap A$. Since $d \in A$ and $A \subseteq B$, it follows that $d \in B$. Therefore, $d \in V_\varepsilon(c) \cap B$, and hence $V_\varepsilon(c) \cap B \neq \emptyset$. Since $\varepsilon > 0$ was arbitrary, we conclude that $c \in \text{cl}(B)$.

□

A point b is in $\text{cl}(A)$ if it can be well-approximated by points of A (because we can find points of A that are arbitrarily close to b). Another way to say that b can be well-approximated by points of A is to say that we can find a sequence from A that converges to b. We now prove that these two concepts coincide.

Proposition 4.1.6. *Let $A \subseteq \mathbb{R}$ and let $b \in \mathbb{R}$. The following are equivalent:*

1. *b is a closure point of A.*
2. *There exists a sequence $\langle a_n \rangle$ with $a_n \in A$ for all $n \in \mathbb{N}^+$ such that $\langle a_n \rangle$ converges to b.*

Proof. First, we assume (1), i.e. that b is a closure point of A. We define a sequence $\langle a_n \rangle$ as follows. Given $n \in \mathbb{N}^+$, we have that $\frac{1}{n} > 0$, so we know that $V_{\frac{1}{n}}(b) \cap A \neq \emptyset$, and we choose a_n to be some element of the nonempty set $V_{\frac{1}{n}}(b) \cap A$. Notice that $a_n \in A$ for all $n \in \mathbb{N}^+$ by definition. We show that $\langle a_n \rangle$ converges to b. Let $\varepsilon > 0$. By Corollary 1.4.7, we can fix $N \in \mathbb{N}^+$ with $\frac{1}{N} < \varepsilon$. Now given any $n \geq N$, we know that $\frac{1}{n} \leq \frac{1}{N}$, so

$$\begin{aligned} |a_n - b| &< \frac{1}{n} && \text{(since } a_n \in V_{\frac{1}{n}}(b)\text{)} \\ &\leq \frac{1}{N} \\ &< \varepsilon. \end{aligned}$$

Therefore, $\langle a_n \rangle$ converges to b.

Conversely, assume (2), and fix a sequence $\langle a_n \rangle$ with $a_n \in A$ for all $n \in \mathbb{N}^+$ such that $\langle a_n \rangle$ converges to b. We show that b is a closure point of A. Let $\varepsilon > 0$. Since $\langle a_n \rangle$ converges to b, we can fix $N \in \mathbb{N}^+$ such that for all $n \geq N$, we have $|a_n - b| < \varepsilon$. In particular, we have $|a_N - b| < \varepsilon$, and hence $a_N \in V_\varepsilon(b)$. Since $a_N \in A$ by assumption, we conclude that $V_\varepsilon(b) \cap A$ is nonempty. We have shown that $V_\varepsilon(b) \cap A \neq \emptyset$ for every $\varepsilon > 0$, so it follows that b is a closure point of A. □

For limit points, there is a similar characterization where we change the sequence to require that we never use the point b itself.

Proposition 4.1.7. *Let $A \subseteq \mathbb{R}$ and let $b \in \mathbb{R}$. The following are equivalent:*

1. *b is a limit point of A.*
2. *There exists a sequence $\langle a_n \rangle$ with $a_n \in A$ and $a_n \neq b$ for all $n \in \mathbb{N}^+$ such that $\langle a_n \rangle$ converges to b.*

Proof. Adapt the proof of Proposition 4.1.6. The details are left as an exercise. □

We also have the following simple result that helps to describe the relationship between closure points and limit points.

Proposition 4.1.8. *For any set $A \subseteq \mathbb{R}$, we have*

$$cl(A) = A \cup \{b \in \mathbb{R} : b \text{ is a limit point of } A\}.$$

Proof. Immediately from the definition we know that any limit point of A is a closure point of A. We also know that $A \subseteq \text{cl}(A)$ from Proposition 4.1.4. Putting these facts together, it follows that $A \cup \{b \in \mathbb{R} : b \text{ is a limit point of } A\} \subseteq \text{cl}(A)$.

For the reverse containment, let $b \in \text{cl}(A)$ be arbitrary. If $b \in A$, then we are done, so suppose that $b \notin A$. We show that b is a limit point of A. Let $\varepsilon > 0$. Since $b \in \text{cl}(A)$, we know that $V_\varepsilon(b) \cap A \neq \emptyset$. Fix some $c \in V_\varepsilon(b) \cap A$. As $b \notin A$, we know that $c \neq b$. It follows that $V_\varepsilon(b) \cap A$ contains a point other than b. Since $\varepsilon > 0$ was arbitrary, we conclude that b is a limit point of A. □

We now work through the details of computing interiors and closures of the basic intervals.

Proposition 4.1.9. *Let $c, d \in \mathbb{R}$ with $c < d$.*

1. *$\text{int}((c,d)) = (c,d)$ and $\text{cl}((c,d)) = [c,d]$.*
2. *$\text{int}([c,d]) = (c,d)$ and $\text{cl}([c,d]) = [c,d]$.*

Proof.

1. We know that $\text{int}((c,d)) \subseteq (c,d)$ by Proposition 4.1.4, so we need only show that $(c,d) \subseteq \text{int}((c,d))$. Let $a \in (c,d)$ be arbitrary, so $c < a < d$. Let $\varepsilon = \min\{a - c, d - a\}$ and notice that $\varepsilon > 0$. We claim that $V_\varepsilon(a) \subseteq (c,d)$. To see this, let $x \in V_\varepsilon(a)$ be arbitrary. By definition, we then have $|x - a| < \varepsilon$. Since $\varepsilon \leq a - c$, we have $|x - a| < a - c$, so $-(a - c) < x - a$, and hence $c < x$. Similarly, since $\varepsilon \leq d - a$, we have $|x - a| < d - a$, so $x - a < d - a$, and hence $x < d$. Therefore, $x \in (c,d)$. Since $x \in V_\varepsilon(a)$ was arbitrary, it follows that $V_\varepsilon(a) \subseteq (c,d)$, and hence $a \in \text{int}((c,d))$. Therefore, $(c,d) \subseteq \text{int}((c,d))$.

   For the closure, first notice that $(c,d) \subseteq \text{cl}((c,d))$ by Proposition 4.1.4. We now argue that $c \in \text{cl}((c,d))$. Let $\varepsilon > 0$ be arbitrary. Let $\delta = \min\{\frac{\varepsilon}{2}, \frac{d-c}{2}\} > 0$. We then have that

$$c < c + \delta \leq c + \frac{\varepsilon}{2} < c + \varepsilon,$$

   so $c + \delta \in V_\varepsilon(c)$. We also have

$$c < c + \delta \leq c + \frac{d-c}{2} = \frac{c+d}{2} < \frac{d+d}{2} = d,$$

   so $c + \delta \in (c,d)$. In particular, we have shown that $V_\varepsilon(c) \cap (c,d) \neq \emptyset$. Since $\varepsilon > 0$ was arbitrary, we conclude that $c \in \text{cl}((c,d))$. A similar argument shows that $d \in \text{cl}((c,d))$. Combining this with the above, it follows that $[c,d] \subseteq \text{cl}((c,d))$.

   We finally show that $\text{cl}((c,d)) \subseteq [c,d]$. Let $b \in \text{cl}((c,d))$ be arbitrary. By Proposition 4.1.6, we can fix a sequence $\langle a_n \rangle$ such that $a_n \in (c,d)$ for all $n \in \mathbb{N}^+$ and such that $\langle a_n \rangle$ converges to b. We then have that $c \leq a_n \leq b$ for all $n \in \mathbb{N}^+$, so by Theorem 2.2.10, it follows that $c \leq \lim\limits_{n\to\infty} a_n \leq d$, which is to say that $c \leq b \leq d$. Therefore, $b \in [c,d]$.

2. We know that $\text{int}([c,d]) \subseteq [c,d]$ by Proposition 4.1.4. Since $(c,d) \subseteq [c,d]$, we can use Proposition 4.1.5 to conclude that $\text{int}((c,d)) \subseteq \text{int}([c,d])$. Using part (1), it follows that $(c,d) \subseteq \text{int}([c,d])$. Putting these together, we have $(c,d) \subseteq \text{int}([c,d]) \subseteq [c,d]$. Thus, to finish the proof that $\text{int}([c,d]) = (c,d)$, we need only show that $c, d \notin \text{int}([c,d])$. Consider c. Given any $\varepsilon > 0$, we have $c - \frac{\varepsilon}{2} \in V_\varepsilon(c)$ and $c - \frac{\varepsilon}{2} \notin [c,d]$, so $V_\varepsilon(c) \not\subseteq (c,d)$. Thus, $c \notin \text{int}([c,d])$. A similar argument shows that $d \notin \text{int}([c,d])$, completing the proof.

   To see that $\text{cl}([c,d]) = [c,d]$, first notice that $[c,d] \subseteq \text{cl}([c,d])$ by Proposition 4.1.4. To see that $\text{cl}([c,d]) \subseteq [c,d]$, simply follow the corresponding proof from part (1), which only used the fact that the sequence $\langle a_n \rangle$ satisfied $c \leq a_n \leq d$ rather than the stronger $c < a_n < d$.

□

We can now use all of our hard work to immediately conclude similar results about the "half-open" intervals $[c,d)$ and $(c,d]$.

Corollary 4.1.10. *Let $c, d \in \mathbb{R}$ with $c < d$.*

1. *$\text{int}([c,d)) = (c,d)$ and $\text{cl}([c,d)) = [c,d]$.*

2. *int*$((c,d]) = (c,d)$ *and* *cl*$((c,d]) = [c,d]$.

Proof. Since $(c,d) \subseteq [c,d) \subseteq [c,d]$, we can use Proposition 4.1.5 to conclude that $\text{int}((c,d)) \subseteq \text{int}([c,d)) \subseteq \text{int}([c,d])$. Using Proposition 4.1.9, it follows that $(c,d) \subseteq \text{int}([c,d)) \subseteq (c,d)$, and hence $\text{int}([c,d)) = (c,d)$. All of the other proofs are completely analogous. □

It is also possible to follow the arguments in Proposition 4.1.9 to show that sets like (c,∞), $[c,\infty)$, $(-\infty,d)$, and $(-\infty,d]$ have the expected interiors and closures. We leave the details as an exercise.

4.2 Open and Closed Sets

Recall from Proposition 4.1.4 that

$$\text{int}(A) \subseteq A \subseteq \text{cl}(A)$$

for every $A \subseteq \mathbb{R}$. Sets A that achieve equality on one of these containments are given a special name.

Definition 4.2.1. *Let* $A \subseteq \mathbb{R}$.

1. *We say that* A *is* open *if* $A \subseteq$ *int*(A)*, i.e. if every element of* A *is an interior point of* A*. Notice that this is equivalent to saying that* $A =$ *int*(A).

2. *We say that* A *is* closed *if* *cl*$(A) \subseteq A$*, i.e. if every closure point of* A *is an element of* A*. Notice that this is equivalent to saying that* $A =$ *cl*(A).

Proposition 4.1.9 tells us that every open interval (c,d) is an open set and that every closed interval $[c,d]$ is a closed set. Notice that $\emptyset$ and $\mathbb{R}$ each trivially satisfy the two definitions, so $\emptyset$ and $\mathbb{R}$ are both open and closed. There also exist sets that are neither open nor closed, such as $[c,d)$ (see Corollary 4.1.10) and $\mathbb{Q}$.

There exist many other sets aside from intervals that are either open or closed. In order to construct some examples, we first show that the collection of open (resp. closed) sets are closed under simple unions and intersections.

Proposition 4.2.2. *Let* $A, B \subseteq \mathbb{R}$.

1. *If* A *and* B *are both open, then both* $A \cup B$ *and* $A \cap B$ *are open.*

2. *If* A *and* B *are both closed, then both* $A \cup B$ *and* $A \cap B$ *are closed.*

Proof.

1. Assume that A and B are both open.

   - $A \cup B$ is open: Let $c \in A \cup B$ be arbitrary. We need to show that $c \in \text{int}(A \cup B)$. Since $c \in A \cup B$, we have two cases:
     - *Case 1:* Suppose that $c \in A$. Since A is open, we know that c is an interior point of A, so we can fix $\varepsilon > 0$ such that $V_\varepsilon(c) \subseteq A$. Since $A \subseteq A \cup B$, it follows that $V_\varepsilon(c) \subseteq A \cup B$. Therefore, c is an interior point of $A \cup B$.
     - *Case 2:* Suppose that $c \in B$. Since B is open, we know that c is an interior point of B, so we can fix $\varepsilon > 0$ such that $V_\varepsilon(c) \subseteq B$. Since $B \subseteq A \cup B$, it follows that $V_\varepsilon(c) \subseteq A \cup B$. Therefore, c is an interior point of $A \cup B$.

   We have shown that $A \cup B \subseteq \text{int}(A \cup B)$, so $A \cup B$ is open.

- $A \cap B$ is open: Let $c \in A \cap B$ be arbitrary. We need to show that $c \in \text{int}(A \cap B)$. Since $c \in A$ and A is open, we know that c is an interior point of A, so we can fix $\varepsilon_1 > 0$ with $V_{\varepsilon_1}(c) \subseteq A$. Since $c \in B$ and B is open, we know that c is an interior point of B, so we can fix $\varepsilon_2 > 0$ with $V_{\varepsilon_2}(c) \subseteq B$. Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$, and notice that $\varepsilon > 0$. As both $\varepsilon_1 \leq \varepsilon$ and $\varepsilon_2 \leq \varepsilon$, it follows that $V_\varepsilon(c) \subseteq V_{\varepsilon_1}(c) \subseteq A$ and $V_\varepsilon(c) \subseteq V_{\varepsilon_2}(c) \subseteq B$. Therefore, $V_\varepsilon(c) \subseteq A \cap B$, and hence c is an interior point of $A \cap B$. We have shown that $A \cap B \subseteq \text{int}(A \cap B)$, so $A \cap B$ is open.

2. Assume that A and B are both closed.

   - $A \cup B$ is closed: Let $c \in \text{cl}(A \cup B)$ be arbitrary. We need to show that $c \in A \cup B$. By Proposition 4.1.6, we can fix a sequence $\langle a_n \rangle$ such that $a_n \in A \cup B$ for all $n \in \mathbb{N}^+$ and such that $\langle a_n \rangle$ converges to c. We now have two cases:
     - *Case 1:* Suppose that $\{n \in \mathbb{N}^+ : a_n \in A\}$ is infinite. In this case, we can take the values of n from this set to extract a subsequence of $\langle a_n \rangle$ consisting of elements of A. Since a subsequence of a convergent sequence must converge to the same limit (this is a nice exercise, but see Theorem 2.5.2 of the book), it follows from Proposition 4.1.6 that $c \in \text{cl}(A)$. Since A is closed, we conclude that $c \in A$, and hence $c \in A \cup B$.
     - *Case 2:* Suppose that $\{n \in \mathbb{N}^+ : a_n \in A\}$ is finite. In this case, we must have that $\{n \in \mathbb{N}^+ : a_n \in B\}$ is infinite. Following the logic in Case 1, we conclude that $c \in \text{cl}(B) = B$, and hence $c \in A \cup B$.

     Thus, in either case, we have shown that $c \in A \cup B$.
   - $A \cap B$ is closed: Let $c \in \text{cl}(A \cap B)$ be arbitrary. We need to show that $c \in A \cap B$. We first show that $c \in \text{cl}(A)$ and $c \in \text{cl}(B)$. Let $\varepsilon > 0$ be arbitrary. Since $c \in \text{cl}(A \cap B)$, we know that $V_\varepsilon(c) \cap (A \cap B)$ is nonempty. Now $V_\varepsilon(c) \cap (A \cap B)$ is a subset of both $V_\varepsilon(c) \cap A$ and $V_\varepsilon(c) \cap B$, so both of these sets must be nonempty. Since $\varepsilon > 0$ was arbitrary, we conclude that $c \in \text{cl}(A)$ and $c \in \text{cl}(B)$. Recall that A and B are both closed, so it follows that $c \in A$ and $c \in B$, and hence $c \in A \cap B$.

□

We can use Proposition 4.2.2 to provide other examples of open and closed sets. For example, since $(0, 1)$ and $(2, 3)$ are both open, it follows that $(0, 1) \cup (2, 3)$ is open. Since $[1, 2]$ and $[2, 3]$ are both closed, we conclude that $\{2\} = [1, 2] \cap [2, 3]$ is closed. More generally, this argument shows that $\{a\}$ is closed for any $a \in \mathbb{R}$. Now given any $a, b \in \mathbb{R}$ with $a \neq b$, we have $\{a, b\} = \{a\} \cup \{b\}$, so any subset of $\mathbb{R}$ with exactly two elements is closed. By repeatedly using Proposition 4.2.2 in this way, it follows that any finite subset of $\mathbb{R}$ is closed.

Our next result provides another way to construct open and closed sets. At first, it might appear immediate from the definition that $\text{int}(A)$ is always an open set. However, there is some real subtlety here. Recall that to show that a set B is open, we have to prove that $B \subseteq \text{int}(B)$. Thus, to show that $\text{int}(A)$ is open, we have to prove that $\text{int}(A) \subseteq \text{int}(\text{int}(A))$. Similarly, to show that $\text{cl}(A)$ is closed, we have to prove that $\text{cl}(\text{cl}(A)) \subseteq \text{cl}(A)$.

Proposition 4.2.3. *For all $A \subseteq \mathbb{R}$, $\text{int}(A)$ is open and $\text{cl}(A)$ is closed.*

Proof. Let $A \subseteq \mathbb{R}$ be arbitrary.

- We first show that $\text{int}(A)$ is open by proving that $\text{int}(A) \subseteq \text{int}(\text{int}(A))$. Let $b \in \text{int}(A)$ be arbitrary. We need to show that b is an interior point of $\text{int}(A)$. Since b is an interior point of A, we can fix $\varepsilon > 0$ such that $V_\varepsilon(b) \subseteq A$. We claim that $V_\varepsilon(b) \subseteq \text{int}(A)$. To see this, first notice that $V_\varepsilon(b) = (b - \varepsilon, b + \varepsilon)$. Now since $(b - \varepsilon, b + \varepsilon) \subseteq A$, we know from Proposition 4.1.5 that $\text{int}((b - \varepsilon, b + \varepsilon)) \subseteq \text{int}(A)$, and hence $(b - \varepsilon, b + \varepsilon) \subseteq \text{int}(A)$ by Proposition 4.1.9. Therefore, $V_\varepsilon(b) \subseteq \text{int}(A)$. Since $\varepsilon > 0$, we conclude that b is an interior point of $\text{int}(A)$, which completes the proof.

- We now show that $\text{cl}(A)$ is closed by proving that $\text{cl}(\text{cl}(A)) \subseteq \text{cl}(A)$. Let $b \in \text{cl}(\text{cl}(A))$ be arbitrary. We then have that b is a closure point of $\text{cl}(A)$. We need to show that b is a closure point of A. Let $\varepsilon > 0$ be arbitrary. Since b is a closure point of $\text{cl}(A)$, we know know that $V_{\varepsilon/2}(b) \cap \text{cl}(A)$ is nonempty, so we fix some c in this intersection. Since $c \in \text{cl}(A)$, we know that c is a closure point of A, so $V_{\varepsilon/2}(c) \cap A$ is nonempty, and hence we can fix some d in this intersection. We then have

$$\begin{aligned} |d - b| &= |d - c + c - b| \\ &\leq |d - c| + |c - b| \\ &< |d - c| + \frac{\varepsilon}{2} && \text{(since } c \in V_{\varepsilon/2}(b)\text{)} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } d \in V_{\varepsilon/2}(c)\text{)} \\ &= \varepsilon, \end{aligned}$$

  so $d \in V_\varepsilon(b)$. Since we also have $d \in A$, we conclude that $V_\varepsilon(b) \cap A$ is nonempty. As $\varepsilon > 0$ was arbitrary, it follows that b is a closure point of A, i.e. that $b \in \text{cl}(A)$, completing the proof.

□

There is a strong complementary relationship between open and closed sets. For a simple example, an interior point is one that satisfies a "there exists" statement, while a closure point is one that satisfies a "for all" statement. More interestingly, in the proof of Proposition 4.2.2, we saw that the union proof for open sets was easier than the intersection proof, while the intersection proof for closed sets was easier than the union proof. We can formalize these vague ideas in the following important result.

Proposition 4.2.4. *Let $A \subseteq \mathbb{R}$. A is open if and only if $A^c = \mathbb{R} \backslash A$ is closed.*

Proof. Suppose first that A^c is closed. We show that A is open. Let $b \in A$ be arbitrary. We then have that $b \notin A^c$, so $b \notin \text{cl}(A^c)$. Thus, we can fix $\varepsilon > 0$ such that $V_\varepsilon(b) \cap A^c = \emptyset$. We then must have $V_\varepsilon(b) \subseteq A$.

Suppose that A^c is not closed. We show that A is not open. Since A^c is not closed, we can then fix $b \in \text{cl}(A^c)$ with $b \notin A^c$. Since $b \notin A^c$, we know that $b \in A$. However, given any $\varepsilon > 0$, we know that $V_\varepsilon(b) \cap A^c \neq \emptyset$, and hence $V_\varepsilon(b) \not\subseteq A$. Therefore, b is not an interior point of A. Thus, A is not open. □

We can use this result together with a few simple set-theoretic facts to save work. For example, suppose that we have proved that whenever A and B are open sets, we have that $A \cup B$ is open (i.e. the very first part of Proposition 4.2.2). We can use this fact together with Proposition 4.2.4 to argue that whenever A and B are closed sets, we have that $A \cap B$ is closed (i.e. the very last part of Proposition 4.2.2). To see this, let $A, B \subseteq \mathbb{R}$ be arbitrary closed sets. We then have that A^c and B^c are both open by Proposition 4.2.4, so we can conclude that $A^c \cup B^c$ is open. Using the fact that

$$(A^c \cup B^c) = (A \cap B)^c,$$

it follows that $(A \cap B)^c$ is open. Since $A \cap B = ((A \cap B)^c)^c$, we can use Proposition 4.2.4 again to conclude that $A \cap B$ is closed.

Although finite unions and intersections are interesting and fundamental, we have also seen how to define countable unions and intersections. Recall that if we have sets $A_1, A_2, A_3, \ldots$, then we defined

$$\bigcup_{n=1}^{\infty} A_n = \{x : \text{There exists } n \in \mathbb{N}^+ \text{ with } x \in A_n\}$$

$$\bigcap_{n=1}^{\infty} A_n = \{x : \text{For all } n \in \mathbb{N}^+, \text{ we have } x \in A_n\}$$

In this case, we have a family of sets that are indexed by positive natural numbers. Can we do the same with other "index sets"? What if we had a set A_r for each $r \in \mathbb{R}$? After a little thought, we see that there is nothing special about using $\mathbb{N}^+$ above. In fact, given *any* set I, if we have a set A_i for each $i \in I$, then we can still talk about the "general union" and "general intersection". When we use a set I to index sets in this way, we naturally call I an *index set*. If we have such a set I, together with sets A_i for each $i \in I$, then we define

$$\bigcup_{i \in I} A_i = \{x : \text{There exists } i \in I \text{ with } x \in A_i\}$$
$$\bigcap_{i \in I} A_i = \{x : \text{For all } i \in I, \text{ we have } x \in A_i\}$$

We now ask whether the general union, or general intersection, of a family of open (resp. closed) sets is still open (resp. closed). Unfortunately, the answer is no, even for families indexed by $\mathbb{N}^+$. For example, we know that $(-\frac{1}{n}, \frac{1}{n})$ is open for each $n \in \mathbb{N}^+$, but

$$\bigcap_{n \in \mathbb{N}^+} (-1/n, 1/n) = \{0\},$$

which is not open. We can even get sets that are neither open nor closed in this way. For example, we have

$$\bigcap_{n \in \mathbb{N}^+} (0, 1 + 1/n) = (0, 1].$$

Similarly, the general union of closed sets need not be closed, even in the case where we index over $\mathbb{N}^+$. For example, we have

$$\bigcup_{n \in \mathbb{N}^+} [1/n, 3 - (1/n)] = (0, 3).$$

For a more interesting example, if we use $\mathbb{Q}$ as our index set, and let $A_q = \{q\}$ for all $q \in \mathbb{Q}$, then each A_q is closed, but

$$\bigcup_{q \in \mathbb{Q}} \{q\} = \mathbb{Q},$$

which is neither open nor closed. Fortunately, we do have the following result which says that open sets behave well under *arbitrary* unions, and closed sets behave well under *arbitrary* intersections.

Proposition 4.2.5. *Let I an index set, and suppose we have sets A_i for each $i \in I$.*

1. *If A_i is open for all $i \in I$, then $\bigcup_{i \in I} A_i$ is open.*
2. *If A_i is closed for all $i \in I$, then $\bigcap_{i \in I} A_i$ is closed.*

Proof. Assume that A_i is open for all i. Let $c \in \bigcup_{i \in I} A_i$ be arbitrary. By definition, we can then fix $j \in I$ with $c \in A_j$. Since A_j is open, we know that c is an interior point of A_j, so we can fix $\varepsilon > 0$ such that $V_\varepsilon(c) \subseteq A_j$. Since $A_j \subseteq \bigcup_{i \in I} A_i$, it follows that $V_\varepsilon(c) \subseteq \bigcup_{i \in I} A_i$. Therefore, c is an interior point of $\bigcup_{i \in I} A_i$.

Assume now that A_i is closed for all i. Let $c \in \text{cl}(\bigcap_{i \in I} A_i)$ be arbitrary. We need to show that $c \in \bigcap_{i \in I} A_i$. We first show that $c \in \text{cl}(A_i)$ for all $i \in I$. So let $j \in I$ be arbitrary, and let $\varepsilon > 0$. Since $c \in \text{cl}(\bigcap_{i \in I} A_i)$, we know that $V_\varepsilon(c) \cap (\bigcap_{i \in I} A_i) \neq \emptyset$ is nonempty. Now $V_\varepsilon(c) \cap (\bigcap_{i \in I} A_i)$ is a subset of both $V_\varepsilon(c) \cap A_j$, so this set must be nonempty. Since $\varepsilon > 0$ was arbitrary, we conclude that $c \in \text{cl}(A_j)$. Recall that A_j is closed, so it follows that $c \in A_j$. Since $j \in I$ was arbitrary, we conclude that $c \in \bigcap_{i \in I} A_i$. □

We can use these ideas to construct an example of a very interesting closed set that has some counterintuitive properties.

Definition 4.2.6. *We define a sequence of sets recursively. We start by letting $C_0 = [0, 1]$. Suppose that C_n is a pairwise disjoint union of 2^n many closed intervals, each of length $\frac{1}{3^n}$. We then let C_{n+1} be the result of removing the open intervals that are the middle third of each interval in C_n, so C_{n+1} is a pairwise disjoint union of 2^{n+1} many closed intervals, each of length $\frac{1}{3^{n+1}}$. This completes the recursive definition of the sets C_n. We then define*

$$C = \bigcap_{n=0}^{\infty} C_n.$$

The set C is called the Cantor set*.*

Notice that each C_n is closed (because it is a finite union of closed intervals), and hence C is closed by Proposition 4.2.5. The set C has many interesting properties. If one tries to determine how "big" C is, one approach would be to determine how much of $[0, 1]$ we have removed at each stage. Notice that we removed an interval of length $\frac{1}{3}$ when constructing C_1, then removed 2 intervals of length $\frac{1}{9}$ when constructing C_2, then removed $4 = 2^2$ intervals of length $(\frac{1}{3})^3$ when constructing C_3, etc. If we follow this logic, the we can determine the total amount removed by looking at the infinite series

$$\frac{1}{3} + 2 \cdot \left(\frac{1}{3}\right)^2 + 2^2 \cdot \left(\frac{1}{3}\right)^3 + 2^3 \cdot \left(\frac{1}{3}\right)^4 + \cdots = \sum_{n=0}^{\infty} \left(\frac{1}{3}\right) \cdot \left(\frac{2}{3}\right)^n.$$

Using Proposition 3.1.2, this series converges and

$$\sum_{n=0}^{\infty} \left(\frac{1}{3}\right) \cdot \left(\frac{2}{3}\right)^n = \frac{1/3}{1 - (2/3)} = \frac{1/3}{1/3} = 1.$$

In other words, it appears that we have removed the full length of $[0, 1]$. However, this does not mean that $C = \emptyset$. In fact, we certainly have $0, 1 \in C$. Moreover, one can prove inductively that the endpoints of any C_n are endpoints of C_m for all $m \geq n$, and hence C contains the endpoints of each of the intervals in C_n. Thus, C is infinite. More surprisingly, C is uncountable!

4.3 Compact Sets

We have everything we need to define one of the most important concepts in analysis and topology: a *compact* set. We begin with the following definition.

Definition 4.3.1. *Let $A \subseteq \mathbb{R}$. An* open cover *of A is a collection of open sets $\{D_i : i \in I\}$, where I is some index set, such that $A \subseteq \bigcup_{i \in I} D_i$.*

We know that $(-n, n)$ is an open set for each $n \in \mathbb{N}^+$ by Proposition 4.1.9. Thus, we can consider the collection $\{(-n, n) : n \in \mathbb{N}^+\}$ of all of these open sets. In this case, our index set is $I = \mathbb{N}^+$, and we are letting $D_i = (-i, i)$ for each $i \in I$. Notice that for every $x \in \mathbb{R}$, we can use Proposition 1.4.5 to fix $n \in \mathbb{N}^+$ with $|x| < n$, from which it follows that $x \in (-n, n)$. Thus, we have $\mathbb{R} \subseteq \bigcup_{n \in \mathbb{N}^+} (-n, n)$, and hence $\{(-n, n) : n \in \mathbb{N}^+\}$ is an open cover of $\mathbb{R}$. In fact, given any $A \subseteq \mathbb{R}$, the collection $\{(-n, n) : n \in \mathbb{N}^+\}$ is an open cover of A because $A \subseteq \mathbb{R} \subseteq \bigcup_{n \in \mathbb{N}^+} (-n, n)$

For another example, consider the collection of open sets $\{(1/n, 1) : n \in \mathbb{N}^+\}$. Here, we are again using $I = \mathbb{N}^+$, but now we are letting $D_i = (1/i, 1)$ for each $i \in \mathbb{N}^+$ (and using $D_1 = (1, 1) = \emptyset$, which is also open). Applying Corollary 1.4.7, it is straightforward to check that $\{(1/n, 1) : n \in \mathbb{N}^+\}$ is an open cover of $(0, 1)$. As above, given any $A \subseteq (0, 1)$, the collection $\{(1/n, 1) : n \in \mathbb{N}^+\}$ is also an open cover of A. For example, $\{(1/n, 1) : n \in \mathbb{N}^+\}$ is an open cover of $[\frac{1}{4}, 1)$. However, notice that $\{(1/n, 1) : n \in \mathbb{N}^+\}$ is *not* an open cover of $[0, 1)$, because $0 \notin (1/n, 1)$ for any $n \in \mathbb{N}^+$.

Definition 4.3.2. *Let $A \subseteq \mathbb{R}$. We say that A is* compact *if every open cover of A has a finite subcover, i.e. whenever $\{D_i : i \in I\}$ is an open cover of A, there exists a finite set $F \subseteq I$ such that $A \subseteq \bigcup_{i \in F} D_i$.*

The definition of compact may be the least intuitive definition that you have seen in mathematics up to this point. It is certainly abstract, and it takes some time to develop an appreciation for its elegance and power. We start by giving a couple of examples of sets that are *not* compact:

- The set $\mathbb{R}$ itself is not compact: As we saw above, the collection of sets $\{(-n, n) : n \in \mathbb{N}^+\}$ is an open cover of $\mathbb{R}$. However, there is no finite $F \subseteq \mathbb{N}^+$ with $\mathbb{R} \subseteq \bigcup_{n \in F}(-n, n)$. To see this, suppose that $F \subseteq \mathbb{N}^+$ is finite. If $F = \emptyset$, then $\bigcup_{n \in F}(-n, n) = \emptyset$, which is certainly not an open cover of $\mathbb{R}$. Suppose then that $F \neq \emptyset$, and let $N = \max(F)$. We then have $(-n, n) \subseteq (-N, N)$ for all $n \in F$, so $\bigcup_{n \in F}(-n, n) = (-N, N)$, which is not an open cover of $\mathbb{R}$. We have shown that the open cover $\{(-n, n) : n \in \mathbb{N}^+\}$ of $\mathbb{R}$ does not have a finite subcover, so $\mathbb{R}$ is not compact.
- The set $(0, 1)$ is not compact: As we saw above, the collection of sets $\{(1/n, 1) : n \in \mathbb{N}^+\}$ is an open cover of $(0, 1)$. However, there is no finite $F \subseteq \mathbb{N}^+$ with $(0, 1) \subseteq \bigcup_{n \in F}(1/n, 1)$. To see this, suppose that $F \subseteq \mathbb{N}^+$ is finite. We may assume that $F \neq \emptyset$ as above, and let $N = \max(F)$. We then have $(1/n, 1) \subseteq (1/N, 1)$ for all $n \in F$, so $\bigcup_{n \in F}(1/n, 1) = (1/N, 1)$, which is not an open cover of $(0, 1)$. We have shown that the open cover $\{(1/n, 1) : n \in \mathbb{N}^+\}$ of $(0, 1)$ does not have a finite subcover, so $(0, 1)$ is not compact.

Consider the set $[\frac{1}{4}, 1)$. As mentioned above, the collection of sets $\{(1/n, 1) : n \in \mathbb{N}^+\}$ is an open cover of $[\frac{1}{4}, 1)$. Now this particular open cover of $[\frac{1}{4}, 1)$ does have a finite subcover. For example, letting $F = \{5\}$, we have $\bigcup_{n \in F}(1/n, 1) = (1/5, 1)$, which is a finite subcover of $[\frac{1}{4}, 1)$. There are also many other finite subsets of $\mathbb{N}^+$ that also work, because any finite $F \subseteq \mathbb{N}^+$ which contains an element greater than 4 will satisfy $[1/4, 1) \subseteq \bigcup_{n \in F}(1/n, 1)$. However, it is important to note that we can *not* conclude from this one collection that $[1/4, 1)$ is compact, as the definition of compact requires that *every* open cover has a finite subcover. In this case, the collection $\{(0, 1 - 1/n) : n \in \mathbb{N}^+\}$ is an open cover of $[1/4, 1)$ that does not have a finite subcover, so in fact $[1/4, 1)$ is not compact.

With this background, it might seem very difficult to prove that a given set $A \subseteq \mathbb{R}$ is compact. In general, this is certainly true, but it is relatively straightforward to handle finite sets.

Proposition 4.3.3. *If $A \subseteq \mathbb{R}$ is finite, then A is compact.*

Proof. Let $A \subseteq \mathbb{R}$ be finite. If $A = \emptyset$, then A is trivially compact because given any open cover $\{D_i : i \in I\}$ of $\emptyset$, we can let $F = \emptyset$. Suppose then that $A \neq \emptyset$, and write $A = \{a_1, a_2, \ldots, a_n\}$. Let $\{D_i : i \in I\}$ be an arbitrary open cover of A. For each k with $1 \leq k \leq n$, we have $a_k \in \bigcup_{i \in I} D_i$, so we can fix $i_k \in I$ with $a_k \in D_{i_k}$. Let $F = \{i_1, i_2, \ldots, i_n\}$. We then have that F is finite and $A \subseteq \bigcup_{i \in F} D_i$, so the open cover $\{D_i : i \in I\}$ of A has a finite subcover. □

Perhaps surprisingly, there do exist infinite compact sets. Before trying to come up with an example, we first prove the following result, which restricts the potential options. The proof is a generalization of the important examples we discussed above.

Proposition 4.3.4. *Every compact set is both closed and bounded.*

Proof. Let A be a compact set.

- We first prove that A is bounded. Consider the collection of sets $\{(-n, n) : n \in \mathbb{N}^+\}$. Since $\mathbb{R} = \bigcup_{n=1}^{\infty}(-n, n)$, we know that $\{(-n, n) : n \in \mathbb{N}^+\}$ is an open cover of A. As A is compact, we can fix a finite set $F \subseteq \mathbb{N}^+$ with $A \subseteq \bigcup_{i \in F}(-n, n)$. If $F = \emptyset$, then $A = \emptyset$, which is trivially bounded. So assume that $F \neq \emptyset$, and let $N = \max(F)$. We then have $(-n, n) \subseteq (-N, N)$ for all $n \in F$, so $\bigcup_{n \in F}(-n, n) = (-N, N)$. Therefore, $A \subseteq (-N, N)$, so $-N < a < N$ for all $a \in A$, and hence A is bounded.

- We next prove that A is closed, which we do by showing that the complement A^c is open. Let $b \in A^c$ be arbitrary. We show that $b \in \text{int}(A^c)$. For each $n \in \mathbb{N}^+$, let

$$D_n = \left(-\infty, b - \frac{1}{n}\right) \cup \left(b + \frac{1}{n}, \infty\right)$$

Notice that each D_n is open and that

$$\bigcup_{n=1}^{\infty} D_n = \mathbb{R} \backslash \{b\}.$$

Since $b \notin A$, we have

$$A \subseteq \bigcup_{n \in \mathbb{N}^+} D_n,$$

and hence $\{D_n : n \in \mathbb{N}^+\}$ is an open cover of A. Using the fact that A is compact, we can fix a finite set $F \subseteq \mathbb{N}^+$ such that $A \subseteq \bigcup_{n \in F} D_n$. If $F = \emptyset$, then $A = \emptyset$, which is trivially closed. So assume that $F \neq \emptyset$, and let $N = \max(F)$. Since $D_m \subseteq D_n$ whenever $m \leq n$, it follows that $\bigcup_{n \in F} D_n = D_N$. Therefore, we have $A \subseteq D_N$, i.e.

$$A \subseteq \left(-\infty, b - \frac{1}{N}\right) \cup \left(b + \frac{1}{N}, \infty\right).$$

It follows that

$$\left(b - \frac{1}{N}, b + \frac{1}{N}\right) \subseteq A^c,$$

which is to say that $V_{1/N}(b) \subseteq A^c$. Hence $b \in \text{int}(A^c)$. We have shown that every element of A^c is an element of $\text{int}(A^c)$, so A^c is open. Using Proposition 4.2.4, we conclude that $A = (A^c)^c$ is closed.

□

Thus, if we are looking for an example of an infinite compact set, then we must look to closed and bounded sets. The simplest example of an infinite closed and bounded set is a closed interval. We now prove that *every* closed interval is compact.

Proposition 4.3.5. *Let $c, d \in \mathbb{R}$ with $c < d$. We then have that $[c, d]$ is compact.*

Proof. Let $\{D_i : i \in I\}$ be an arbitrary open cover of $[c, d]$. Let

$$B = \{x \in [c, d] : \text{There exists a finite set } F \subseteq I \text{ with } [c, x] \subseteq \bigcup_{i \in F} D_i\}.$$

Notice that $c \in B$ trivially because c is an element of some D_i, and we can fix such an i and let $F = \{i\}$. Also, B is bounded above by d. Thus, we can let $s = \sup B$. Since d is an upper bound of B, it follows that $s \leq d$. Also, since $c \in B$, we have $c \leq s$.

We first claim that $s = d$. Suppose instead that $s < d$. Since $c \leq s$, we have $s \in [c, d]$, so we can fix a $j \in I$ with $s \in D_j$. Since D_j is open, we can fix $\varepsilon > 0$ with $V_\varepsilon(s) \subseteq D_j$. Now $s - \varepsilon < s$, so $s - \varepsilon$ is not an upper bound of B, and hence we can fix $x \in B$ with $x > s - \varepsilon$. By definition of B, we can fix a finite set $F \subseteq I$ with $[c, x] \subseteq \bigcup_{i \in F} D_i$. Letting $G = F \cup \{j\}$ and $\delta = \min\{d - c, \frac{\varepsilon}{2}\} > 0$, we then have that G is finite and $s < s + \delta \leq d$. Furthermore, $[c, s + \delta] \subseteq \bigcup_{i \in G} D_i$, so $s + \delta \in B$, contradicting the fact that s is an upper bound of B. Therefore, we must have $s = d$.

We now show that the open cover $\{D_i : i \in I\}$ of $[c, d]$ has a finite subcover. Since $\{D_i : i \in I\}$ is an open cover of $[c, d]$, we can fix $j \in I$ with $d \in D_j$. Since D_j is open, we can fix $\varepsilon > 0$ with $V_\varepsilon(s) \subseteq D_j$. As $d - \varepsilon < d$, we know that $d - \varepsilon$ is not an upper bound of B, and hence we can fix $x \in B$ with $x > d - \varepsilon$. By definition of B, we can fix a finite set $F \subseteq I$ with $[c, x] \subseteq \bigcup_{i \in F} D_i$. Letting $G = F \cup \{j\}$, we then have that G is finite and that $[c, d] \subseteq \bigcup_{i \in G} D_i$. Thus, we have found a finite subcover of $[c, d]$. □

In fact, *every* closed and bounded subset of $\mathbb{R}$ is compact. In order to proof this surprising and fundamental result, we will make use of the following lemma.

Lemma 4.3.6. *Every closed subset of a compact set is compact. In other words, if $A \subseteq \mathbb{R}$ is compact, and $B \subseteq A$ is closed, then B is compact.*

Proof. Suppose that $A \subseteq \mathbb{R}$ is compact and $B \subseteq A$ is closed. Let $\{D_i : i \in I\}$ be an arbitrary open cover of B. Since B is closed, we know from Proposition 4.2.4 that B^c is open. Notice that $\{B^c\} \cup \{D_i : i \in I\}$ is an open cover of $\mathbb{R}$, and so it is certainly an open cover of A. Since A is compact, we can fix a finite set $F \subseteq I$ such that $\{B^c\} \cup \{D_i : i \in F\}$ is an open cover of A. As $B \subseteq A$, we know that $\{B^c\} \cup \{D_i : i \in F\}$ is an open cover of B. But $B \cap B^c = \emptyset$, so $\{D_i : i \in F\}$ is a finite subcover of B. □

We can now put together all of the pieces we have assembled to obtain a nice characterization of the compact sets.

Theorem 4.3.7 (Heine-Borel)**.** *Let $A \subseteq \mathbb{R}$. A is compact if and only if it is closed and bounded.*

Proof. If A is compact, then we know A is closed and bounded from Proposition 4.3.4. Conversely, suppose that A is closed and bounded. Since A is bounded, we can fix $d \in \mathbb{R}$ with $d > 0$ such that $-d \leq a \leq d$ for all $a \in A$. We then have that $A \subseteq [-d, d]$. Now $[-d, d]$ is compact by Proposition 4.3.5. Since A is a closed subset of the compact $[-d, d]$, we can use Lemma 4.3.6 to conclude that A is compact. □

We can also obtain another characterization of compact sets. Notice that Abbott uses this characterization as the *definition* of compact, but our definition is the standard one because it is both more powerful and generalizable.

Proposition 4.3.8. *Let $A \subseteq \mathbb{R}$. A is compact if and only if every sequence $\langle a_n \rangle$ from A (i.e. where $a_n \in A$ for all $n \in \mathbb{N}^+$) has a subsequence that converges to an element of A.*

Proof. By the Heine-Borel Theorem, we know that A is compact if and only if A is closed and bounded. Thus, to prove the result, it suffices to show each of the following:

1. If A is closed and bounded, then every sequence $\langle a_n \rangle$ from A has a subsequence that converges to an element of A.

2. If A is not bounded, then there exists a sequence $\langle a_n \rangle$ from A that has no subsequence converging to a point of A.

3. If A is not closed, then there exists a sequence $\langle a_n \rangle$ from A that has no subsequence converging to a point of A.

We now prove each of these.

1. Suppose that A is closed and bounded. Let $\langle a_n \rangle$ be an arbitrary sequence from A. Since A is a bounded set, we know that $\langle a_n \rangle$ is bounded, so we can use the Bolzano-Weierstrass Theorem (Theorem 2.4.4) to fix a convergent subsequence $\langle a_{n_k} \rangle$ of $\langle a_n \rangle$. Let $b = \lim_{k \to \infty} a_{n_k}$. Since $a_{n_k} \in A$ for all $k \in \mathbb{N}^+$, we know from Proposition 4.1.6 that $b \in \text{cl}(A)$. As A is closed, we have $A = \text{cl}(A)$, so $b \in A$. Therefore, $\langle a_n \rangle$ has a subsequence that converges to an element of A.

2. Suppose that A is not bounded. Define a sequence $\langle a_n \rangle$ as follows. Given $n \in \mathbb{N}^+$, we know that $A \not\subseteq [-n, n]$ because A is not bounded, so can let a_n be some element of A with $|a_n| > n$. Now let $\langle a_{n_k} \rangle$ be an arbitrary subsequence of $\langle a_n \rangle$. Given any $k \in \mathbb{N}^+$, we have $|a_{n_k}| > n_k \geq k$ by Lemma 2.4.2. Therefore, $\langle a_{n_k} \rangle$ is not bounded. Using Proposition 2.2.7, we conclude that $\langle a_{n_k} \rangle$ does not converge at all, let alone to a point of A.

3. Suppose that A is not closed. We then have that $\text{cl}(A) \neq A$. Since we know that $A \subseteq \text{cl}(A)$ from Proposition 4.1.4, we can fix $b \in \text{cl}(A) \backslash A$. As $b \in \text{cl}(A)$, we can use Proposition 4.1.6 to fix a sequence $\langle a_n \rangle$ from A that converges to b. Now every subsequence of $\langle a_n \rangle$ will also converge to b (see Theorem 2.5.2 in the book), so in particular no subsequence of $\langle a_n \rangle$ converges to a point of A.

□

Chapter 5

Functions, Limits, and Continuity

The concept of a limit of a function is fundamental in Calculus, as it underlies the definitions of continuity and differentiability. However, one often treats limits (and the subsequent concepts) somewhat informally in Calculus. For most "natural" functions that are defined by simple formulas, such an intuitive approach rarely leads to problems. In fact, for a long time, when a mathematician used the word *function*, they meant a formulaic expression built up from polynomials, trigonometric functions, and exponentials (together with their inverses) through addition, multiplication, division, and composition. As a result, they were able to work at a more intuitive level just like you did in Calculus.

Eventually, this rigid and simplistic view of functions began to fall out of favor due to the intrinsic limitations of such an approach. For example, Newton (and many mathematicians after him) made extensive use of power series, which we will explore in the next chapter. As mathematicians came up with more intricate ways to describe relationships between two numbers, they were forced to adopt the more modern understanding of a function as *any* rule that assigns a single output to each input.

When exploring these ideas in the 19th century (as the foundations of analysis were being forged), mathematicians stumbled upon some weird and counterintuitive functions. As a simple example, consider the function

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{otherwise.} \end{cases}$$

If one tries to think about limits of this function, or whether it is continuous, it quickly becomes apparent that we need a careful definition of these concepts. In this chapter, we carry out this task, and build up the fundamental ideas from Calculus on the resulting foundation.

5.1 Limits

We start with a careful definition of a limit of a function. Suppose that $A \subseteq \mathbb{R}$ and $f\colon A \to \mathbb{R}$ is a function. We want to make sense of $\lim_{x \to c} f(x) = \ell$. Before jumping into the definition, we should ask whether we need to put any restrictions on c. For example, should we require that c is in the domain of the function, i.e. that $c \in A$? Thinking back to Calculus, most of the interesting limits (such as in the definition of the derivative) arise in setting where c is *not* in the domain. However, we do want c to have some relationship to A. Since we want to think about points of A that are "close" to c, we certainly want such points to exist! As we do no care about c itself, the natural restriction is that we want c to be a limit point of A.

Definition 5.1.1. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let c be a limit point of A. We write $\lim_{x \to c} f(x) = \ell$ to mean that for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $0 < |x - c| < \delta$, we have $|f(x) - \ell| < \varepsilon$. We*

say that $\lim_{x\to c} f(x)$ exists *if there exists* $\ell \in \mathbb{R}$ *such that* $\lim_{x\to c} f(x) = \ell$. *Otherwise, we say that* $\lim_{x\to c} f(x)$ *does not exist.*

We can rephrase the definition of $\lim_{x\to c} f(x) = \ell$ using the language and notation of ε-neighborhoods:

Proposition 5.1.2. *Let* $A \subseteq \mathbb{R}$, *let* $f\colon A \to \mathbb{R}$, *and let* c *be a limit point of* A. *The following are equivalent:*

1. $\lim_{x\to c} f(x) = \ell$.
2. *For all* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for all* $x \in A$ *with* $x \in V_\delta(c)\backslash\{c\}$, *we have* $f(x) \in V_\varepsilon(\ell)$.

Proof. Immediate. □

Before moving on to theoretical results, we work through the definition using a couple of examples. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function $f(x) = 4x - 5$. We claim that

$$\lim_{x\to 2} f(x) = 3.$$

To see this, let $\varepsilon > 0$ be arbitrary. Consider $\delta = \frac{\varepsilon}{4} > 0$. For any x with $0 < |x - 2| < \delta$, we have

$$\begin{aligned} |f(x) - 3| &= |4x - 5 - 3| \\ &= |4x - 8| \\ &= |4 \cdot (x - 2)| \\ &= 4 \cdot |x - 2| \\ &< 4 \cdot \delta \\ &= \varepsilon. \end{aligned}$$

Therefore, $\lim_{x\to 2} f(x) = 3$. Notice that in our argument, we used $\delta = \frac{\varepsilon}{4}$. Algebraically, that was simply what we needed to carry out the above argument. However, it makes sense from a geometric perspective. The graph of f is a line of slope 4, so it rises 4 units for every corresponding unit change in the input. Thus, if we are challenged to get within ε of 3, we would expect that we would have to respond with an input window that is at most one-fourth as large.

For a more complicated example, let $f\colon \mathbb{R} \to \mathbb{R}$ be the function $f(x) = x^2$. We claim that

$$\lim_{x\to 3} f(x) = 9.$$

Before diving into the argument, let's build some intuition. Suppose that we are challenged with some $\varepsilon > 0$, We want to pick $\delta > 0$ such that given any x with $0 < |x - 3| < \delta$, we have $|x^2 - 9| < \varepsilon$. Notice that

$$\begin{aligned} |x^2 - 9| &= |(x + 3)(x - 3)| \\ &= |x + 3| \cdot |x - 3|. \end{aligned}$$

Now we get to pick δ in response to ε. By choosing a sufficiently small δ, we will be able to make $|x - 3|$ as small as we would like. How do we handle that $|x + 3|$ term? Intuitively, if x is close to 3, then $|x + 3|$ will be close to 6. We can force the values of x to be close to 3 by choosing δ small enough. Suppose that we make sure to choose a $\delta \leq 1$ so that we can get a decent upper bound. Notice that if $0 < |x - 3| < 1$, then $|x + 3| < 7$. Now that we have a reasonable upper bound on $|x + 3|$, we can also ensure that δ is at most $\frac{\varepsilon}{7}$ to make the whole product less than ε. With these ideas in mind, we now turn to the argument.

Let $\varepsilon > 0$. Consider $\delta = \min\{1, \frac{\varepsilon}{7}\} > 0$. For any x with $0 < |x-3| < \delta$, we have

$$\begin{aligned} |x+3| &= |x-3+6| \\ &\leq |x-3| + |6| \\ &< 1 + 6 \\ &= 7, \end{aligned}$$

so

$$\begin{aligned} |f(x) - 9| &= |x^2 - 9| \\ &= |(x+3)(x-3)| \\ &= |x+3| \cdot |x-3| \\ &< 7 \cdot \delta \\ &\leq \varepsilon. \end{aligned}$$

Therefore, $\lim\limits_{x\to 3} f(x) = 9$.

We turn now to some fundamental results about limits. We begin by proving an analogue of Proposition 2.1.5 about the uniqueness of limits. The proof is very similar, but we replace the (large) values of N with corresponding (small) values of δ. Notice that we need to use the fact that c is a limit point here in order to ensure that there is at least one point very close to, but not equal to, c.

Proposition 5.1.3. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let c be a limit point of A. If $\ell, m \in \mathbb{R}$ and both $\lim\limits_{x\to c} f(x) = \ell$ and $\lim\limits_{x\to c} f(x) = m$, then $\ell = m$.*

Proof. We show that $|\ell - m| < \varepsilon$ for all $\varepsilon > 0$. Let $\varepsilon > 0$ be arbitrary. Since $\lim\limits_{x\to c} f(x) = \ell$, we can fix $\delta_1 > 0$ such that $|f(x) - \ell| < \frac{\varepsilon}{2}$ whenever $x \in A$ and $0 < |x - c| < \delta_1$. Since $\lim\limits_{x\to c} f(x) = m$, we can fix $\delta_2 > 0$ such that $|f(x) - \ell| < \frac{\varepsilon}{2}$ whenever $x \in A$ and $0 < |x - c| < \delta_2$. Let $\delta = \min\{\delta_1, \delta_2\} > 0$. Since c is a limit point of A, we know that $V_\delta(c) \cap A$ contains at least one point other than c, so fix a point $x \neq c$ in this intersection. We then have

$$\begin{aligned} |\ell - m| &= |\ell - f(x) + f(x) - m| \\ &\leq |\ell - f(x)| + |f(x) - m| && \text{(by the Triangle Inequality)} \\ &= |f(x) - \ell| + |f(x) - m| \\ &< \frac{\varepsilon}{2} + |f(x) - m| && \text{(since } x \in A \text{ and } 0 < |x-c| < \delta_1\text{)} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } x \in A \text{ and } 0 < |x-c| < \delta_2\text{)} \\ &= \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that $|\ell - m| < \varepsilon$ for all $\varepsilon > 0$. Using Proposition 1.6.4, we conclude that $\ell = m$. □

We can similarly switch N's with δ's to prove an analogue of Proposition 2.2.9.

Proposition 5.1.4. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let c be a limit point of A. Suppose that $\lim\limits_{x\to c} f(x)$ exists and let $\ell = \lim\limits_{x\to c} f(x)$.*

1. *If $\ell > 0$, then there exists $\delta > 0$ such that for all $x \in A$ with $0 < |x - c| < \delta$, we have $f(x) > 0$.*

2. *If $\ell < 0$, then there exists $\delta > 0$ such that for all $x \in A$ with $0 < |x - c| < \delta$, we have $f(x) < 0$.*

Proof.

1. Suppose that $\ell > 0$. Since $\lim_{x\to c} f(x) = \ell$, we can fix $\delta > 0$ such that $|f(x) - \ell| < \frac{\ell}{2}$ whenever $x \in A$ and $0 < |x - c| < \delta$. Let $x \in A$ be arbitrary with $0 < |x - c| < \delta$. We then have $|f(x) - \ell| < \frac{\ell}{2}$, so we know from Proposition 1.6.3 that $f(x) - \ell > -\frac{\ell}{2}$. Adding ℓ to both sides, it follows that $f(x) > \frac{\ell}{2}$. Since $\frac{\ell}{2} > 0$, we conclude that $f(x) > 0$.

2. Suppose that $\ell < 0$. Since $\lim_{x\to c} f(x) = \ell$, we can fix $\delta > 0$ such that $|f(x) - \ell| < -\frac{\ell}{2}$ whenever $x \in A$ and $0 < |x - c| < \delta$. Let $x \in A$ be arbitrary with $0 < |x - c| < \delta$. We then have have $|f(x) - \ell| < \frac{-\ell}{2}$, so we know from Proposition 1.6.3 that $f(x) - \ell < -\frac{\ell}{2}$. Adding ℓ to both sides, it follows that $f(x) < \frac{\ell}{2}$. Since $\frac{\ell}{2} < 0$, we conclude that $f(x) < 0$.

□

Similarly, we get the following analogue to Proposition 2.2.1.

Proposition 5.1.5. *Let $A \subseteq \mathbb{R}$, let $f, g\colon A \to \mathbb{R}$, and let c be a limit point of A. Suppose that there exists $\delta > 0$ such that $f(x) = g(x)$ for all $x \in (A \cap V_\delta(c))\backslash\{c\}$. If $\lim\limits_{x\to c} f(x)$ exists, then $\lim\limits_{x\to c} g(x)$ exists, and $\lim\limits_{x\to c} f(x) = \lim\limits_{x\to c} g(x)$.*

Proof. Exercise. □

It is certainly possible to mimic many of the other proofs from Chapter 2, while changing the various N's to δ's, and taking minimums instead of maximums. Instead of working through all of those details, we can instead prove the following result that relates $\lim_{x\to c} f(x)$ to a condition about sequences that converge to c. Intuitively, if $\lim_{x\to c} f(x) = \ell$, and we take a sequence $\langle a_n\rangle$ that converges to c, then the corresponding sequence of function values $\langle f(a_n)\rangle$ should converge to ℓ. However, there is some subtlety here, because we have to ensure that $a_n \neq c$ for all $n \in \mathbb{N}^+$ (since the value of f at c is irrelevant to the limit).

Proposition 5.1.6. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let c be a limit point of A. The following are equivalent:*

1. $\lim\limits_{x\to c} f(x) = \ell$.
2. *For every sequence $\langle a_n\rangle$ where $a_n \in A\backslash\{c\}$ for all $n \in \mathbb{N}^+$ and where $\langle a_n\rangle$ converges to c, we have that $\langle f(a_n)\rangle$ converges to ℓ.*

Proof. Suppose first that $\lim_{x\to c} f(x) = \ell$. Let $\langle a_n\rangle$ be an arbitrary sequence where $a_n \in A\backslash\{c\}$ for all $n \in \mathbb{N}^+$ and where $\langle a_n\rangle$ converges to c. We show that $\langle f(a_n)\rangle$ converges to ℓ. Let $\varepsilon > 0$ be arbitrary. Since $\lim_{x\to c} f(x) = \ell$, we can fix $\delta > 0$ such that for all $x \in A$ with $0 < |x - c| < \delta$, we have $|f(x) - \ell| < \varepsilon$. Since $\langle a_n\rangle$ converges to c, we can fix $N \in \mathbb{N}^+$ such that for all $n \geq N$, we have $|a_n - c| < \delta$. Now let $n \geq N$ be arbitrary. We then have $|a_n - c| < \delta$, and since $a_n \neq c$, it follows that $0 < |a_n - c| < \delta$. Combining this with the fact that $a_n \in A$, we conclude that $|f(a_n) - \ell| < \varepsilon$.

Suppose now that it is not the case that $\lim_{x\to c} f(x) = \ell$. We show that there exists a sequence $\langle a_n\rangle$ where $a_n \in A\backslash\{c\}$ for all $n \in \mathbb{N}^+$ and where $\langle a_n\rangle$ converges to c, but where $\langle f(a_n)\rangle$ does not converge to ℓ. Since it is not the case that $\lim_{x\to c} f(x) = \ell$, we can fix $\varepsilon > 0$ such that for all $\delta > 0$ there exists $x \in A$ with $0 < |x - c| < \delta$ and $|f(x) - \ell| \geq \varepsilon$. We define a sequence $\langle a_n\rangle$ as follows. Given $n \in \mathbb{N}^+$, let a_n be some element of A with both $0 < |a_n - c| < \frac{1}{n}$ and $|f(a_n) - \ell| \geq \varepsilon$. Notice that for all $n \in \mathbb{N}^+$, we have $a_n \neq c$ because $|a_n - c| > 0$. Thus, we have $a_n \in A\backslash\{c\}$ for all $n \in \mathbb{N}^+$. Since $|a_n - c| < \frac{1}{n}$ for all $n \in \mathbb{N}^+$, we have $c - \frac{1}{n} < a_n < c + \frac{1}{n}$ for all $n \in \mathbb{N}^+$. Since $\lim_{n\to\infty}(c - \frac{1}{n}) = 0 = \lim_{n\to\infty}(c + \frac{1}{n})$, we can use the Squeeze Theorem to conclude that $\langle a_n\rangle$ converges to c. Finally, since $|f(a_n) - \ell| \geq \varepsilon$ for all $n \in \mathbb{N}^+$, we know that $\langle f(a_n)\rangle$ does not converge to ℓ. □

Let's use this result to show that $\lim_{x\to 0} \sin(1/x)$ does not exist. One way to prove this is to use Proposition 5.1.6 to find two sequence of nonzero values $\langle a_n \rangle$ and $\langle b_n \rangle$ that both converge to 0, but where $\langle f(a_n) \rangle$ and $\langle f(b_n) \rangle$ converge to different values. To that end, let

$$a_n = \frac{1}{n\pi} \qquad \text{and} \qquad b_n = \frac{1}{2\pi n + (\pi/2)}$$

for all $n \in \mathbb{N}^+$. We then have that $\langle a_n \rangle$ and $\langle b_n \rangle$ both consist of nonzero values and both converge to 0 (the latter because $0 \leq b_n \leq \frac{1}{2\pi n}$ for all $n \in \mathbb{N}^+$). However, we have $f(a_n) = \sin(n\pi) = 0$ for all $n \in \mathbb{N}^+$, and $f(b_n) = \sin(2\pi n + (\pi/2)) = 1$ for all $n \in \mathbb{N}^+$. Thus, $\langle f(a_n) \rangle$ converges to 0 while $\langle f(b_n) \rangle$ converges to 1. Using Proposition 5.1.6, if $\lim_{x\to 0} \sin(1/x)$ existed, then it must be both 0 and 1, contradicting Proposition 5.1.3.

Alternatively, we can argue that $\lim_{x\to 0} \sin(1/x)$ does not exist by finding one example of a sequence $\langle a_n \rangle$ of nonzero values that converges to 0, but where $\langle f(a_n) \rangle$ diverges. To do this, let

$$a_n = \frac{1}{\pi n + (\pi/2)}$$

for all $n \in \mathbb{N}^+$, and notice that $\langle a_n \rangle$ converges to 0. However, we have $f(a_n) = (-1)^n$ for all $n \in \mathbb{N}^+$, so $\langle f(a_n) \rangle$ diverges.

Despite these examples, the immediate value of Proposition 5.1.6 is that it often allows us to directly transfer over results about limits of sequences to corresponding results about limits of functions. Here is perhaps the most important example.

Theorem 5.1.7. *Let $A \subseteq \mathbb{R}$, let $f, g\colon A \to \mathbb{R}$, and let c be a limit point of A. Suppose that $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ both exist, and let $\ell = \lim_{x\to c} f(x)$ and $m = \lim_{x\to c} g(x)$.*

1. *The limit $\lim_{x\to c}(f+g)(x)$ exists, and $\lim_{x\to c}(f+g)(x) = \ell + m$.*
2. *The limit $\lim_{x\to c}(f-g)(x)$ exists, and $\lim_{x\to c}(f-g)(x) = \ell - m$.*
3. *The limit $\lim_{x\to c}(f\cdot g)(x)$ exists, and $\lim_{x\to c}(f\cdot g)(x) = \ell \cdot m$.*
4. *Suppose that $m \neq 0$. We then have that c is a limit point of the domain of $\frac{f}{g}$. Moreover, the limit $\lim_{x\to c}(\frac{f}{g})(x)$ exists, and $\lim_{x\to c}(\frac{f}{g})(x) = \frac{\ell}{m}$.*

Proof. We prove (1) by using Proposition 5.1.6. Let $\langle a_n \rangle$ be arbitrary sequence where $a_n \in A\backslash\{c\}$ for all $n \in \mathbb{N}^+$ and where $\langle a_n \rangle$ converges to c. We show that $\langle (f+g)(a_n) \rangle$ converges to $\ell + m$. Since $\ell = \lim_{x\to c} f(x)$, we can use Proposition 5.1.6 to conclude that $\langle f(a_n) \rangle$ converges to ℓ. Similarly, since $m = \lim_{x\to c} f(x)$, we can use Proposition 5.1.6 to conclude that $\langle g(a_n) \rangle$ converges to m. Using Theorem 2.2.8, we conclude that $\langle f(a_n) + g(a_n) \rangle$ converges to $\ell + m$, i.e. $\langle (f+g)(a_n) \rangle$ converges to $\ell + m$. Since $\langle a_n \rangle$ was an arbitrary sequence with $a_n \in A\backslash\{c\}$ for all $n \in \mathbb{N}^+$ and where $\langle a_n \rangle$ converges to c, we can apply Proposition 5.1.6 again to conclude that $\lim_{x\to c}(f+g)(x) = \ell + m$.

The proofs for (2) and (3) are similar. For (4), we first need to argue that c is a limit point of the domain of $\frac{f}{g}$. To do this, use the fact that $m \neq 0$ together with Proposition 5.1.4 to see that there exists $\delta > 0$ such that $g(x) \neq 0$ for all $x \in (V_\delta(c)\backslash\{c\}) \cap A$, so every point in this set will be in the domain of $\frac{f}{g}$. Now follow the proof of the other parts. □

Theorem 5.1.8 (The Squeeze Theorem)**.** *Let $f, g, h\colon A \to \mathbb{R}$ be functions and let c be a limit point of A. Suppose that*

$$\lim_{x\to c} f(x) = \ell = \lim_{x\to c} h(x).$$

Suppose also that there exists $\delta_0 > 0$ such that $f(x) \leq g(x) \leq h(x)$ for all $x \in (V_{\delta_0}(c)\backslash\{c\}) \cap A$. We then have that $\lim_{x\to c} g(x)$ exists, and that $\lim_{x\to c} g(x) = \ell$.

Proof. Let $\varepsilon > 0$. By assumption, we can fix $\delta_0 > 0$ such that $f(x) \leq g(x) \leq h(x)$ for all $x \in (V_{\delta_0}(c)\backslash\{c\}) \cap A$. Since $\lim_{x\to c} f(x) = \ell$, we can fix $\delta_1 > 0$ such that $|f(x) - \ell| < \varepsilon$ for all $x \in A$ with $0 < |x - c| < \delta_1$. Since $\lim_{x\to c} h(x) = \ell$, we can fix $\delta_2 > 0$ such that $|h(x) - \ell| < \varepsilon$ for all $x \in A$ with $0 < |x - c| < \delta_2$. Let $\delta = \min\{\delta_0, \delta_1, \delta_2\}$. Now let $x \in A$ be arbitrary with $0 < |x - c| < \delta$. Since $0 < |x - c| < \delta_1$, we have $|f(x) - \ell| < \varepsilon$. Since $0 < |x - c| < \delta_2$, we have $|h(x) - \ell| < \varepsilon$. Using Proposition 1.6.3, it follows that

$$-\varepsilon < f(x) - \ell < \varepsilon \qquad \text{and} \qquad -\varepsilon < h(x) - \ell < \varepsilon.$$

In particular, we have both

$$\ell - \varepsilon < f(x) \qquad \text{and} \qquad h(x) < \ell + \varepsilon.$$

Now since $0 < |x - c| < \delta_0$, we know that $f(x) \leq g(x) \leq h(x)$, and therefore we have

$$\ell - \varepsilon < g(x) < \ell + \varepsilon.$$

It follows that $-\varepsilon < g(x) - \ell < \varepsilon$, and so using Proposition 1.6.3 again, we conclude that $|g(x) - \ell| < \varepsilon$. Therefore, $\lim_{x\to c} g(x) = \ell$. □

For a simple example of using the Squeeze Theorem, we show that $\lim_{x\to 0} x \cdot \sin(1/x) = 0$. First, notice that for all $x \in \mathbb{R}\backslash\{0\}$, we have

$$-1 \leq \sin(1/x) \leq 1.$$

Therefore, for any $x \in \mathbb{R}\backslash\{0\}$, we have

$$-x \leq x \sin(1/x) \leq x$$

(for $x < 0$, note that multiplying through by x flips the inequalities, but it still comes out to this). Now it is straightforward to see that $\lim_{x\to 0} x = 0 = \lim_{x\to 0} -x$ by using $\delta = \varepsilon$ in the definition. Using the Squeeze Theorem, we conclude that $\lim_{x\to 0} x \cdot \sin(1/x) = 0$.

5.2 Continuity

Although limits are important, perhaps the most fundamental concept related to functions is continuity.

Definition 5.2.1. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let $c \in A$. We say that f is* continuous *at c if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$, we have $|f(x) - f(c)| < \varepsilon$.*

This definition closely resembles the definition of a limit, but there are a few important differences. First, we require that $c \in A$, rather than assume that c is a limit point of A. We need to assume that c is in the domain of the function so that we can apply f to it. However, it is possible that $c \in A$, and yet c is not a limit point of A. How? If $c \in A$, then we know that $c \in \text{cl}(A)$, so by homework, we can conclude that either c is a limit point of A or c is an isolated point of A. Now the situation of an isolated point is a little strange from an intuitive perspective, but it is straightforward (and a good exercise) to see that c is an isolated point of A, then f is automatically continuous at c.

However, the most interesting situation is when $c \in A$ is also a limit point of A (we really only include that isolated points as possibilities because that lines up with the generalizations of continuity to more abstract spaces). In this case, how does the definition of continuity of c compare with the limit definition? The most obvious change is that we are not considering only $f(c)$ rather than a general ℓ. The other small change is that we are considering all $x \in A$ with $|x - c| < \delta$ rather than all $x \in A$ with $0 < |x - c| < \delta$. However, notice that if $|x - c| = 0$, then we must have $x = c$, and hence $|f(x) - f(c)| = 0$, which will be less than any positive ϵ. Putting these ideas together, we arrive at the following simple result.

Proposition 5.2.2. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let $c \in A$. Assume also that c is a limit point of A. The following are equivalent:*

1. *f is continuous at c.*
2. $\lim_{x \to c} f(x) = f(c)$.

Proof. All of the key ideas are in the above paragraph, but we leave the details as an exercise. □

We extend the definition of continuity at a point to continuity on a set in the obvious way.

Definition 5.2.3. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let $C \subseteq A$. We say that f is continuous on C if f is continuous at every $c \in C$. Unwrapping the definition, f is continuous on C if for all $c \in C$ and all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$, we have $|f(x) - f(c)| < \varepsilon$.*

For a very simple example, notice that if we define $f\colon \mathbb{R} \to \mathbb{R}$ by letting $f(x) = x$, then f is continuous on $\mathbb{R}$ (given any $c \in \mathbb{R}$ and any $\varepsilon > 0$, we can take $\delta = \varepsilon$). We now prove that the square root function is continuous on its natural domain.

Proposition 5.2.4. *Define $f\colon [0, \infty) \to \mathbb{R}$ by letting $f(x) = \sqrt{x}$. We then have that f is continuous on $[0, \infty)$.*

Proof. Let $c \in [0, \infty)$. We show that f is continuous at c by handling two cases:

- *Case 1:* Suppose that $c = 0$. Let $\varepsilon > 0$. Consider $\delta = \varepsilon^2 > 0$. Let $x \geq 0$ be arbitrary with $|x - 0| < \delta$. Since $x \geq 0$, we then have $0 < x < \delta$. Using the solution to Problem 3a on Homework 4, we then have $\sqrt{x} < \sqrt{\delta} = \sqrt{\varepsilon^2} = \varepsilon$, and hence

$$\begin{aligned} |f(x) - f(c)| &= |\sqrt{x} - \sqrt{0}| \\ &= |\sqrt{x}| \\ &= \sqrt{x} && \text{(since } \sqrt{x} \geq 0\text{)} \\ &= \varepsilon && \text{(from above).} \end{aligned}$$

  Thus, f is continuous at $c = 0$ in this case.

- *Case 2:* Suppose that $c > 0$. Let $\varepsilon > 0$. Consider $\delta = \sqrt{c} \cdot \varepsilon > 0$. Let $x \geq 0$ be arbitrary with $|x - c| < \delta$. We then have

$$\begin{aligned} |f(x) - f(c)| &= |\sqrt{x} - \sqrt{c}| \\ &= |\sqrt{x} - \sqrt{c}| \cdot \frac{|\sqrt{x} + \sqrt{c}|}{|\sqrt{x} + \sqrt{c}|} \\ &= \frac{|x - c|}{\sqrt{x} + \sqrt{c}} && \text{(since } \sqrt{x} + \sqrt{c} \geq 0\text{)} \\ &\leq \frac{|x - c|}{\sqrt{c}} && \text{(since } \sqrt{x} \geq 0\text{)} \\ &< \frac{\delta}{\sqrt{c}} \\ &= \frac{\sqrt{c} \cdot \varepsilon}{\sqrt{c}} \\ &= \varepsilon. \end{aligned}$$

  Thus, f is continuous at c in this case as well.

□

In light of Proposition 5.2.2, we can also easily adapt the sequential characterization of a limit from Proposition 5.1.6 to apply to continuity. In this setting, the result is actually easier, because we do not need to assume that the sequence $\langle a_n \rangle$ consists of elements distinct from c.

Proposition 5.2.5. *Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$, and let $c \in A$. The following are equivalent:*

1. *f is continuous at c.*
2. *For all $\varepsilon > 0$, there exists $\delta > 0$, such that for all $x \in V_\delta(c)$, we have $f(x) \in V_\varepsilon(f(c))$.*
3. *For every sequence $\langle a_n \rangle$ where $a_n \in A$ for all $n \in \mathbb{N}^+$ and where $\langle a_n \rangle$ converges to c, we have that $\langle f(a_n) \rangle$ converges to $f(c)$.*

Proof. The first two are just restatements of each other. For the equivalence (1) and (3), simply follow the proof of Proposition 5.1.6, noting that since we replaced $0 < |x - c| < \delta$ in the definition of a limit with $|x - c| < \delta$ in the definition of continuity, we can allow $a_n = c$ for some values of $n \in \mathbb{N}^+$. □

We can use this result to obtain information about limits of sequences. For example, suppose that $\langle a_n \rangle$ is a sequence that converges to ℓ, and assume that $a_n \geq 0$ for all $n \in \mathbb{N}^+$. Using Theorem 2.2.10, we then know that $\ell \geq 0$ as well. Now we know from Proposition 5.2.4 that the function $f\colon [0, \infty) \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$ is continuous on $[0, \infty)$, and hence is continuous at ℓ. Using Proposition 5.2.5, it follows that the sequence $\langle f(a_n) \rangle$ converges to $f(\ell)$, i.e. the sequence $\langle \sqrt{a_n} \rangle$ converges to $\sqrt{\ell}$.

We also have the following analogue of Theorem 5.1.7. In fact, if c is a limit point of A, then this result following immediately from that theorem combined with Proposition 5.2.2. It is possible to handle the isolated points separately (and rather easily), but we can also just follow the proof of Theorem 5.1.7 while replacing applications of Proposition 5.1.6 with applications of Proposition 5.2.5.

Theorem 5.2.6. *Let $A \subseteq \mathbb{R}$, let $f, g\colon A \to \mathbb{R}$, and let $c \in A$. Suppose that f and g are both continuous at c.*

1. *The function $f + g$ is continuous at c.*
2. *The function $f - g$ is continuous at c.*
3. *The function $f \cdot g$ is continuous at c.*
4. *If $g(c) \neq 0$, then $\frac{f}{g}$ is continuous at c.*

Proof. See above. Again, the details are left as an exercise. □

By repeatedly using this theorem, we can now easily see that a wide class of natural functions are continuous on their natural domains.

Corollary 5.2.7.

1. *All polynomial functions are continuous on $\mathbb{R}$. That is, if $a_0, a_1, \ldots, a_n \in \mathbb{R}$, and we define $f\colon \mathbb{R} \to \mathbb{R}$ by letting $f(x) = a_n x^n + \cdots + a_1 x + a_0$, then f is continuous on $\mathbb{R}$.*
2. *All rational functions are continuous on their domain. More formally, we have the following. Let $a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_m \in \mathbb{R}$ with at least one $b_i \neq 0$. Let $T = \{r \in \mathbb{R} : b_m r^m + \cdots + b_1 r + b_0 = 0\}$, and notice that T is finite. If we define $f\colon \mathbb{R} \backslash T \to \mathbb{R}$ by letting*

$$f(x) = \frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0}$$

then f is continuous on $\mathbb{R} \backslash T$.

Proof.

1. We noted above that the identity function $g(x) = x$ is continuous on $\mathbb{R}$. Now for any $k \in \mathbb{N}^+$, we can apply part (3) of Theorem 5.2.6 several times to conclude that the function $g(x) = x^k$ is continuous on $\mathbb{R}$. Also, given any $a \in \mathbb{R}$, the constant function $g(x) = a$ is easily seen to be continuous on $\mathbb{R}$ (given any $\varepsilon > 0$, just let $\delta = 1$). Therefore, by part (3) of Theorem 5.2.6 again, given any $k \in \mathbb{N}^+$ and $a_k \in \mathbb{R}$, the function $g(x) = a_k x^k$ is continuous on $\mathbb{R}$. Finally, we can repeatedly apply part (1) of Theorem 5.2.6 to conclude that any polynomial function is continuous on $\mathbb{R}$.

2. This is immediate by the first part of this result combined with part (4) of Theorem 5.2.6.

□

Although Theorem 5.2.6 tells us that the continuous functions are closed under arithmetic operations on functions, there is another fundamental function operation that we have not considered: composition. For example, consider the function $h(x) = \sqrt{x^2 - 4}$ define on the domain $(-\infty, -2] \cup [2, \infty)$. We know that the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2 - 4$ is continuous on $\mathbb{R}$ by Corollary 5.2.7, and we know that the function $g\colon [0, \infty) \to \mathbb{R}$ defined by $g(x) = \sqrt{x}$ is continuous on $[0, \infty)$ by Proposition 5.2.4. Notice that $h = g \circ f$, assuming that we restrict f to the set $A = (-\infty, -2] \cup [2, \infty)$. In other words, since f is continuous on $\mathbb{R}$, we trivially have that the restricted function $f \restriction A\colon A \to \mathbb{R}$ is continuous on A, and we are really letting h be the composition of g with $f \restriction A$.

We now prove that the composition of continuous functions always results in a continuous function, assuming that the composition makes sense. In order to state this last part precisely, we use the following (terrible) notation. Given a function $f\colon D \to \mathbb{R}$ and a set $A \subseteq D$, we let $f(A) = \{f(a) : a \in A\}$.

Proposition 5.2.8. *Let $f\colon A \to \mathbb{R}$, let $g\colon B \to \mathbb{R}$, and assume that $f(A) \subseteq B$ (i.e. that $f(a) \in B$ for all $a \in A$), so that $g \circ f$ is defined on A. If f is continuous at c, and g is continuous at $f(c)$, then $g \circ f$ is continuous at c.*

We give two proofs of this fundamental result.

Proof 1. We use Proposition 5.2.5. Let $\langle a_n \rangle$ be an arbitrary sequence from A that converges to c. Since f is continuous at c, we know from Proposition 5.2.5 that $\langle f(a_n) \rangle$ converges to $f(c)$. Since g is continuous at $f(c)$, we can apply Proposition 5.2.5 again to conclude that $\langle g(f(a_n)) \rangle$ converges to $g(f(c))$. Therefore, $\langle (g \circ f)(a_n) \rangle$ converges to $(g \circ f)(c)$. Since $\langle a_n \rangle$ was arbitrary, Proposition 5.2.5 tells us that $g \circ f$ is continuous at c. □

Proof 2. We use the definition of continuity. Let $c \in A$ and $\varepsilon > 0$ be arbitrary. Since g is continuous at $f(c)$, we can fix $\alpha > 0$ such that for all $y \in A$ with $|y - f(c)| < \alpha$, we have $|g(y) - g(f(c))| < \varepsilon$. Since f is continuous at c, we can fix $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$, we have $|f(x) - f(c)| < \alpha$. Now let $x \in A$ be arbitrary with $|x - c| < \delta$. We then have $|f(x) - f(c)| < \alpha$, so $|g(f(x)) < g(f(c))| < \varepsilon$, and hence $|(g \circ f)(x) - (g \circ f)(c)| < \varepsilon$. Therefore, $g \circ f$ is continuous at c. □

Before moving on, we investigate a fascinating example of a strange function $f\colon \mathbb{R} \to \mathbb{R}$. It goes by various names, such as Thomae's function, the popcorn function, or the raindrop function. Here is the description. If $x \in \mathbb{R} \backslash \mathbb{Q}$, then $f(x) = 0$. Suppose then that $x \in \mathbb{Q}$. Write $x = \frac{m}{n}$ uniquely in lowest terms where $m, n \in \mathbb{Z}$ and $n > 0$, and then define $f(x) = \frac{1}{n}$. In other words, given a rational, write it in lowest terms, and then change the numerator to 1. Here is the compact description:

$$f(x) = \begin{cases} \frac{1}{n} & \text{if } x = \frac{m}{n} \text{ where } m, n \in \mathbb{Z}, n > 0, \text{ and } \gcd(m, n) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

We argue that f is continuous at each irrational number, and is not continuous at each rational number.

- Let $c \in \mathbb{Q}$. We show that f is not continuous at c. Fix the unique $m, n \in \mathbb{Z}$ with $c = \frac{m}{n}$, $n > 0$, and $\gcd(m, n) = 1$. Let $\varepsilon = \frac{1}{n} > 0$. Given any $\delta > 0$, we know from Corollary 1.5.12 that we can fix an irrational $x \in (c - \delta, c + \delta)$, and then we have

$$\begin{aligned} |f(x) - f(c)| &= \left|0 - \frac{1}{n}\right| \\ &= \frac{1}{n} \\ &= \varepsilon. \end{aligned}$$

  Therefore, f is not continuous at c.

- Let $c \in \mathbb{R}\backslash\mathbb{Q}$. We will show that f is continuous at c. Before getting into the details, we fix argue that given any $n \in \mathbb{N}^+$, the set

$$\left\{m \in \mathbb{Z} : \left|\frac{m}{n} - c\right| < 1\right\}$$

  is finite. To see this, let $n \in \mathbb{N}^+$ be arbitrary. Given any $m \in \mathbb{Z}$ with $|\frac{m}{n} - c| < 1$, we have $-1 < \frac{m}{n} - c < 1$, so $-n < m - cn < n$, and hence $(c - 1)n < m < (c + 1)n$. Thus,

$$\left\{m \in \mathbb{Z} : \left|\frac{m}{n} - c\right| < 1\right\} \subseteq \{m \in \mathbb{Z} : (c - 1)n < m < (c + 1)n\}.$$

  Now it is intuitively clear that the latter set is finite (a careful proof uses Proposition 1.4.5), so our original set is finite.

  We now show that f is continuous at c. Let $\varepsilon > 0$. Using Proposition 1.4.5, fix $k \in \mathbb{N}^+$ with $k > \frac{1}{\varepsilon}$. Now there are only finitely many $n \in \mathbb{N}^+$ with $n < k$, and for each such n, we know from above that there are only finitely many $m \in \mathbb{Z}$ with $|\frac{m}{n} - c| < 1$. Let F be the finite set of all such rational numbers, i.e. let

$$F = \left\{q \in \mathbb{Q} : |q - c| < 1 \text{ and } q = \frac{m}{n} \text{ for some } m, n \in \mathbb{Z} \text{ with } 0 < n < k\right\}$$

  Consider

$$\delta = \min(\{1\} \cup \{|q - c| : q \in F\}).$$

  Notice that $\delta > 0$ because $c \notin \mathbb{Q}$ (and hence $c \neq q$ for all $q \in F$). Let $x \in \mathbb{R}$ be arbitrary with $|x - c| < \delta$. If $x \in \mathbb{R}\backslash\mathbb{Q}$, then

$$\begin{aligned} |f(x) - f(c)| &= |0 - 0| \\ &= 0 \\ &< \varepsilon. \end{aligned}$$

  Suppose then that $x \in \mathbb{Q}$. Write $x = \frac{m}{n}$ where $m, n \in \mathbb{Z}$, $n > 0$, and $\gcd(m, n) = 1$. Notice that $|x - c| < \delta \leq |q - c|$ for all $q \in F$, so $x \notin F$. Since we also have $|x - c| < \delta \leq 1$, it must be the case that $n \geq k$. Therefore, we have $n > \frac{1}{\varepsilon}$, so $\frac{1}{n} < \varepsilon$, and hence

$$\begin{aligned} |f(x) - f(c)| &= \left|\frac{1}{n} - 0\right| \\ &= \frac{1}{n} \\ &< \varepsilon. \end{aligned}$$

  It follows that f is continuous at c.

We now talk about some fundamental results about how continuous functions behave with respect to compact sets. Our first result in this direction is the following.

Proposition 5.2.9. *The continuous image of a compact set is compact. That is, if f is continuous on a compact set K, then $f(K)$ is compact.*

Proof. We use Proposition 4.3.8 to show that $f(K)$ is compact. Let $\langle c_n \rangle$ be an arbitrary sequence from $f(K)$. For each $n \in \mathbb{N}^+$, fix $a_n \in K$ with $f(a_n) = c_n$. Since $\langle a_n \rangle$ is a sequence from the compact set K, we can use Proposition 4.3.8 to fix a subsequence $\langle a_{n_k} \rangle$ of $\langle a_n \rangle$ such that $\langle a_{n_k} \rangle$ converges to a point of K. Let $b = \lim_{k \to \infty} a_{n_k} \in K$. Since f is continuous, we can use Proposition 5.2.5 to conclude that $\langle f(a_{n_k}) \rangle$ converges to $f(b)$. Thus, $\langle c_{n_k} \rangle = \langle f(a_{n_k}) \rangle$ is a subsequence of $\langle c_n \rangle$ that converges to $f(b)$, which is an element of $f(K)$. Using Proposition 4.3.8, it follows that $f(K)$ is compact. □

One of the most important consequences of this result is the following theorem. It plays an important role in Calculus when talking about continuous functions defined on closed bounded intervals. In that setting, we learn techniques to find the absolute maximum and minimum values of the function on the interval. However, those computational techniques depend essentially on the *existence* of such values. Since closed bounded intervals are compact (see Proposition 4.3.5), this result applies there. We will return to discuss these computational techniques after we discuss differentiation.

Theorem 5.2.10 (Extreme Value Theorem)**.** *If K is compact and $f \colon K \to \mathbb{R}$ is continuous, then f achieves a maximum and minimum value on K.*

Proof. By Proposition 5.2.9, we know that $f(K)$ is compact, so it has a maximum and minimum value by the homework. □

Compact sets interact with continuity in another interesting, but much more subtle, way. Before getting into the details, let's examine the function $f(x) = x^2$. Now we know from Corollary 5.2.7 that f is continuous on $\mathbb{R}$. But let's prove this fact directly. Let $c \in \mathbb{R}$ and $\varepsilon > 0$ be arbitrary. Consider $\delta = \min\{1, \frac{\varepsilon}{1+2\cdot|c|}\} > 0$. For any x with $0 < |x - c| < \delta$, we have

$$\begin{aligned} |x + c| &= |x - c + 2c| \\ &\leq |x - c| + |2c| \\ &< 1 + 2 \cdot |c|, \end{aligned}$$

so

$$\begin{aligned} |f(x) - f(c)| &= |x^2 - c^2| \\ &= |(x + c)(x - c)| \\ &= |x + c| \cdot |x - c| \\ &< (1 + 2 \cdot |c|) \cdot \delta \\ &\leq (1 + 2 \cdot |c|) \cdot \frac{\varepsilon}{1 + 2 \cdot |c|} \\ &= \varepsilon. \end{aligned}$$

Therefore, f is continuous on $\mathbb{R}$. Notice that in this argument, our choice of δ depending on *both* c and ε. Of course, this is certainly allowed, but there do exist functions where the choice of δ depends only on ε. For example, consider the function $f(x) = 4x - 5$. We claim that f is continuous on $\mathbb{R}$. To see this, let $c \in \mathbb{R}$

and $\varepsilon > 0$ be arbitrary. Consider $\delta = \frac{\varepsilon}{4} > 0$. For any x with $0 < |x - c| < \delta$, we have

$$\begin{aligned} |f(x) - f(c)| &= |(4x - 5) - (4c - 5)| \\ &= |4x - 4c| \\ &= 4 \cdot |x - c| \\ &< 4 \cdot \delta \\ &= 4 \cdot \frac{\varepsilon}{4} \\ &= \varepsilon. \end{aligned}$$

Notice that δ depended only on ε here.

Definition 5.2.11. *Let $A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$. We say that f is* uniformly continuous on A *if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in A$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.*

Let's spell out the distinction between "continuous on A" and "uniformly continuous on A" in detail. Let $f\colon A \to \mathbb{R}$. Written purely symbolically, the statement that f is continuous on A is

$$(\forall c \in A)(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in A)[|x - c| < \delta \to |f(x) - f(c)| < \varepsilon].$$

Thus, we take a c and an ε, and then our choice of δ can depend on both. In contrast, the statement that f is uniformly continuous on A is

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in A)(\forall y \in A)[|x - y| < \delta \to |f(x) - f(y)| < \varepsilon],$$

which under a simple change of names is just

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall c \in A)(\forall x \in A)[|x - c| < \delta \to |f(x) - f(c)| < \varepsilon].$$

The only difference between this statement and the original one is that we have moved $(\forall c \in A)$ past the $(\exists \delta > 0)$ quantifier. The effect of this shift is precisely that our choice of δ can *only* depend on ε, and can not in any way depend on c.

Of course, if we prove that a function f is continuous on A, and we use a choice of δ that depends on both ε and c, we can *not* necessarily conclude that f fails to be uniformly continuous on A. After all, there might be a different argument that dispenses with the dependence on c. But nonetheless, let's show that $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$. Consider $\varepsilon = 1$. Let $\delta > 0$ be arbitrary. Fix $n \in \mathbb{N}^+$ with $\frac{1}{n} < \delta$, and consider $x = n + \frac{1}{n}$ and $y = n$. We then have $|x - y| = \frac{1}{n} < \delta$, and

$$\begin{aligned} |f(x) - f(y)| &= \left|x^2 - y^2\right| \\ &= \left|n^2 + 2 + \frac{1}{n^2} - n^2\right| \\ &= 2 + \frac{1}{n^2} \\ &> 1 \\ &= \varepsilon. \end{aligned}$$

Theorem 5.2.12. *A continuous function on a compact set is uniformly continuous. That is, if K is compact and $f\colon K \to \mathbb{R}$ is continuous on K, then f is uniformly continuous on K.*

Proof. Let K be compact and let $f\colon K \to \mathbb{R}$. Let $\varepsilon > 0$. For each $c \in K$, since f is continuous at c, we can fix $\delta_c > 0$ such that for all $x \in K \cap (c - \delta_c, c + \delta_c)$, we have $|f(x) - f(c)| < \frac{\varepsilon}{2}$. Consider the collection of

open sets $\{(c - \frac{1}{2} \cdot \delta_c, c + \frac{1}{2} \cdot \delta_c) : c \in K\}$. Notice that this is an open cover of K. Since K is compact, we can fix a finite subcover

$$\{(c_1 - \tfrac{1}{2} \cdot \delta_{c_1}, c_1 + \tfrac{1}{2} \cdot \delta_{c_1}), (c_2 - \tfrac{1}{2} \cdot \delta_{c_2}, c_2 + \tfrac{1}{2} \cdot \delta_{c_2}), \ldots, (c_n - \tfrac{1}{2} \cdot \delta_{c_n}, c_n + \tfrac{1}{2} \cdot \delta_{c_n})\}$$

of K. Let

$$\delta = \min\{\tfrac{1}{2} \cdot \delta_{c_1}, \tfrac{1}{2} \cdot \delta_{c_2}, \ldots, \tfrac{1}{2} \cdot \delta_{c_n}\}.$$

Let $x, y \in K$ be arbitrary with $|x - y| < \delta$. Choose k such that $x \in (c_k - \frac{1}{2} \cdot \delta_{c_k}, c_k + \frac{1}{2} \cdot \delta_{c_k})$. Then $|x - c_k| < \frac{1}{2} \cdot \delta_{c_k}$, so

$$\begin{aligned} |y - c_k| &= |y - x + x - c_k| \\ &\leq |y - x| + |x - c_k| \\ &< \delta + \frac{1}{2} \cdot \delta_{c_k} \\ &\leq \frac{1}{2} \cdot \delta_{c_k} + \frac{1}{2} \cdot \delta_{c_k} \\ &= \delta_{c_k}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} |f(x) - f(y)| &= |f(x) - f(c_k) + f(c_k) - f(y)| \\ &\leq |f(x) - f(c_k)| + |f(c_k) - f(y)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

□

We end by proving a fundamental result about continuous functions that you learn about in Calculus: the Intermediate Value Theorem. Although it is a very intuitive result, it does provide another way to understand that the real have no "holes" in them, and so working through the details is essential. We start with the following analogue of Proposition 5.1.4.

Proposition 5.2.13. *Let $A \subseteq \mathbb{R}$, let $f \colon A \to \mathbb{R}$, and let $c \in A$. Suppose that f is continuous at c.*

1. *If $f(c) > 0$, then there exists $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$, we have $f(x) > 0$.*
2. *If $f(c) < 0$, then there exists $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$, we have $f(x) < 0$.*

Proof. If c is a limit point of A, then this follows from Proposition 5.1.4 and Proposition 5.2.2. One can then handle the case of c being an isolated point separately. Alternatively, just follow the proof of Proposition 5.1.4, using the definition of continuity in place of the definition of a limit. □

We now prove the following weak version of the Intermediate Value Theorem.

Lemma 5.2.14. *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f \colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, and suppose that $f(a) < 0 < f(b)$. There exists $c \in (a, b)$ with $f(c) = 0$.*

Proof. Let $B = \{x \in [a, b] : f(x) < 0\}$. Notice that B is nonempty because $a \in B$, and B is bounded above by b. Thus, we can let $c = \sup B$. We then have $a \leq c \leq b$, so $c \in [a, b]$. We claim that $f(c) = 0$. We prove this by showing that the other two possibilities are impossible:

- *Case 1:* Suppose that $f(c) < 0$. As $f(b) > 0$, we must have $c < b$. Since f is continuous at c, we can use Proposition 5.2.13 to fix $\delta > 0$ such that $f(x) < 0$ for all $x \in [a, b] \cap V_\delta(c)$. Let $d = \min\{c + \frac{\delta}{2}, b\}$, and notice that $d \in [a, b]$ and $c < d$. Since $d \in [a, b] \cap V_\delta(c)$, we have $f(d) < 0$, and hence $d \in B$. However, as $d > c$, this contradictions the fact that c is an upper bound of B.

- *Case 2:* Suppose that $f(c) > 0$. As $f(a) < 0$, we must have $a < c$. Since f is continuous at c, we can use Proposition 5.2.13 to fix $\delta > 0$ such that $f(x) > 0$ for all $x \in [a, b] \cap V_\delta(c)$. Let $d = \max\{a, c - \frac{\delta}{2}\}$, and notice that $d \in [a, b]$ and $d < c$. Now for any $x \in [d, c]$, we know that $f(x) > 0$. Since c is an upper bound for B, and no element of $[d, c]$ is in B, it follows that d is an upper bound for B. However, this contradicts the fact that c is the least upper bound of B.

Since both cases lead to a contradiction, we conclude that we must have $f(c) = 0$. □

We can now prove the full result.

Theorem 5.2.15 (Intermediate Value Theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f\colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$.*

1. *If $f(a) < f(b)$, then for all $r \in \mathbb{R}$ with $f(a) < r < f(b)$, there exists $c \in (a, b)$ with $f(c) = r$.*
2. *If $f(a) > f(b)$, then for all $r \in \mathbb{R}$ with $f(b) < r < f(a)$, there exists $c \in (a, b)$ with $f(c) = r$.*

Proof. Suppose first that $f(a) < f(b)$. Let $r \in \mathbb{R}$ with $f(a) < r < f(b)$. Define $g\colon [a, b] \to \mathbb{R}$ by letting $g(x) = f(x) - r$. Since f is continuous on $[a, b]$, and constant functions are continuous, we can use Theorem 5.2.6 to conclude that g is continuous on $[a, b]$. Now $g(a) = f(a) - r < 0$ and $g(b) = f(b) - r > 0$, so by Lemma 5.2.14, we can fix $c \in (a, b)$ with $g(a) = 0$. We then have $f(c) = r$.

For the second part, simply work with the function $g(x) = r - f(x)$ instead. □

Chapter 6

Differentiation and Integration

The operations of differentiation and integration, and the relationship between them, forms the heart of Calculus. In this chapter, we build up these core concepts using the fundamental ideas and results that we have developed. In the case of differentiation, most of the core material is handled pretty well in Calculus, so we will move quickly (while working out the details of a few interesting proofs). In contrast, integration is much more interesting and subtle, both in working out the definitions and determining which functions are integrable.

6.1 Differentiation

Although it is certainly possible to work with functions on complicated domains, we will typically restrict out attention in this chapter to functions defined on intervals. However, we allow open or closed intervals (or even half-open), as well as both bounded and unbounded intervals. In order to be more careful about what we mean, we can either list all of the possible options, or we can abstract away the fundamental similarity underlying the different types of intervals. The key property is the following.

Definition 6.1.1. *Let $A \subseteq \mathbb{R}$. We say that A is* interval *if whenever $c, d \in A$ with $c < d$, we have $(c, d) \subseteq A$.*

In other words, a set $A \subseteq \mathbb{R}$ is an interval if whenever we have two points in A, every point between them is also in A. Using this definition, it is possible to prove the following. The details are not hard, but there are many cases to consider. For example, we first consider the case where A is nonempty and bounded above, and then break into cases based on whether $\sup A \in A$ or not. We omit the details.

Proposition 6.1.2. *Every interval in $\mathbb{R}$ is of one of the following types: $\mathbb{R}$, $\emptyset$, (a, b), $(a, b]$, $[a, b)$, $[a, b]$, $(-\infty, b)$, $(-\infty, b]$, (a, ∞), and $[a, \infty)$. Moreover, each of these sets are intervals.*

Proof. Exercise. □

Although we will restrict our attention to functions $f\colon A \to \mathbb{R}$ where A is an interval, notice that if B is a finite (or even countable) union of disjoint intervals, and $f\colon B \to \mathbb{R}$, then we can simply work with each interval in the union separately. For example, we can work with the function $f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$ by restricting attention separately to the two intervals $(-\infty, 0)$ and $(0, \infty)$.

So let A be an interval, and let $f\colon A \to \mathbb{R}$. Given $c \in A$, we want to define the derivative of f at c. The idea from Calculus is to consider the difference quotient

$$\frac{f(x) - f(c)}{x - c}$$

and determine what happens as x approaches c. In the formal definition, we simply take the limit as x approaches c. Moreover, we will restrict attention to points $c \in \text{int}(A)$ (i.e. we will not talk about differentiability of a function at the possible endpoints of an interval), so that we have an entire neighborhood of points around c that lie in A. Notice that if $c \in \text{int}(A)$, then we certainly have that c is a limit point of $A\backslash\{c\}$, so we can take a limit.

Definition 6.1.3. *Let A be an interval, let $f\colon A \to \mathbb{R}$ be a function, and let $c \in int(A)$. Define $g\colon A\backslash\{c\} \to \mathbb{R}$ by letting*

$$g(x) = \frac{f(x) - f(c)}{x - c}.$$

We say that f is differentiable *at c if $\lim_{x\to c} g(x)$ exists, i.e. if*

$$\lim_{x\to c} \frac{f(x) - f(c)}{x - c}$$

exists. In this case, we define $f'(c)$ to be this limit.

Let's see a couple of interesting examples. Consider $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Is f differentiable at 0? Notice that for all $x \in \mathbb{R}\backslash\{0\}$, we have

$$\begin{aligned} \frac{f(x) - f(0)}{x - 0} &= \frac{x\sin(1/x) - 0}{x - 0} \\ &= \sin(1/x), \end{aligned}$$

and we know from Section 5.1 that this function has no limit as x approaches 0. Therefore, f is not differentiable at 0.

Now consider $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^2\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

We again ask whether f differentiable at 0. Notice that for all $x \in \mathbb{R}\backslash\{0\}$, we have

$$\begin{aligned} \frac{f(x) - f(0)}{x - 0} &= \frac{x^2\sin(1/x) - 0}{x - 0} \\ &= x\sin(1/x), \end{aligned}$$

and we know that this limit exists and equals 0 by the Squeeze Theorem (see the end of Section 5.1). Therefore, f is differentiable at 0 and $f'(0) = 0$.

In Calculus, one often defines the derivative $f'(c)$ as the limit

$$\lim_{h\to 0} \frac{f(c+h) - f(c)}{h}.$$

Although the structure is a bit different from our definition, they both clearly represent the limit of slopes of secant lines on the graph of f as we move closer to $(c, f(c))$. To formally argue that the definitions are equivalent, we use a result from the homework.

Proposition 6.1.4. *Let A be an interval, let $f\colon A \to \mathbb{R}$ be a function, and let $c \in int(A)$. We then have that f is differentiable at c if and only if*

$$\lim_{h\to 0} \frac{f(c+h)-f(c)}{h}$$

exists, and in this case, we have that $f'(c)$ equals this limit.

Proof. Define $g_1\colon A\backslash\{c\} \to \mathbb{R}$ by letting

$$g_1(x) = \frac{f(x)-f(c)}{x-c}.$$

A straightforward argument shows that $A - \{c\} = \{a - c : a \in A\}$ is also an interval with $0 \in \text{int}(A)$. Now define $g_2\colon (A-\{c\})\backslash\{0\}$ by

$$g_2(x) = \frac{f(c+x)-f(c)}{x}.$$

Notice that $g_2(x) = g_1(x+c)$ for all $x \in (A-\{c\})\backslash\{0\}$ and that $g_1(x) = g_2(x-c) = g_2(x+(-c))$ for all $x \in A\backslash\{c\}$. Thus, by Problem 2a on Homework 9, we know that $\lim_{x\to c} g_1(x)$ exists if and only if $\lim_{x\to 0} g_2(x)$ exists, and in this case the limits are equal. □

We now prove that differentiability is stronger than continuity.

Proposition 6.1.5. *Let A be an interval, let $f\colon A \to \mathbb{R}$ be a function, and let $c \in int(A)$. If f is differentiable at c, then f is continuous at c.*

Proof. Suppose that f is differentiable at c. Define $g\colon A\backslash\{c\} \to \mathbb{R}$ by letting

$$g(x) = \frac{f(x)-f(c)}{x-c}$$

for all $x \in A\backslash\{c\}$. Notice that

$$f(x) = f(c) + (x-c)\cdot g(x)$$

for all $x \in A\backslash\{c\}$. Since f is differentiable at c, we know that $\lim_{x\to c} g(x)$ exists and equals $f'(c)$. Using the simple facts that $\lim_{x\to c} x = c$ and that $\lim_{x\to c} a = a$ for any $a \in \mathbb{R}$, we can use Theorem 5.1.7 to conclude that $\lim_{x\to c} f(x)$ exists, and moreover that

$$\begin{aligned}\lim_{x\to c} f(x) &= \lim_{x\to c}[f(c) + (x-c)\cdot g(x)] \\ &= f(c) + (c-c)\cdot f'(c) \\ &= f(c).\end{aligned}$$

Since $c \in \text{int}(A)$, we know that c is a limit point of A, so we can apply Proposition 5.2.2 to conclude that f is continuous at c. □

The next two results are covered pretty well in Calculus. We omit the proofs, but they are a good exercise (or see the book).

Theorem 6.1.6. *Let A be an interval, let $f, g\colon A \to \mathbb{R}$ be functions, and let $c \in int(A)$.*

1. *If f and g are both differentiable at c, then $f+g$ is differentiable at c, and $(f+g)'(c) = f'(c) + g'(c)$.*
2. *If f is differentiable at c, and $r \in \mathbb{R}$, then $r\cdot f$ is differentiable at c, and $(r\cdot f)'(c) = r\cdot f'(c)$.*

3. *If f and g are both differentiable at c, then $f \cdot g$ is differentiable at c, and $(f \cdot g)'(c) = f'(c) \cdot g(c) + f(c) \cdot g'(c)$.*

4. *If g is differentiable at c, and $g'(c) \neq 0$, then $\frac{1}{g}$ is differentiable at c, and*

$$\left(\frac{1}{g}\right)'(c) = -\frac{g'(c)}{g(c)^2}.$$

5. *If f and g are both differentiable at c, and $g'(c) \neq 0$, then $\frac{f}{g}$ is differentiable at c, and*

$$\left(\frac{f}{g}\right)'(c) = \frac{f'(c) \cdot g(c) - f(c) \cdot g'(c)}{g(c)^2}.$$

Theorem 6.1.7 (Chain Rule)**.** *Let A and B be intervals, Let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$, and assume that $f(A) \subseteq B$, so that $g \circ f$ is defined on A. Let $c \in A$. If f is differentiable at c, g is continuous at $f(c)$, and $f(c) \in int(B)$, then $g \circ f$ is differentiable at c and $(g \circ f)'(c) = g'(f(c)) \cdot f'(c)$.*

We can now follow the proof of Theorem 5.2.6 to obtain the corresponding results about differentiability.

Corollary 6.1.8.

1. *Each polynomial functions is differentiable on $\mathbb{R}$.*

2. *Each rational function is differentiable on any interval contained in its domain.*

Proof. Notice that the identity function and all constant functions are differentiable on $\mathbb{R}$, and then follow the proof of Corollary 6.1.8, replacing applications of Theorem 5.2.6 with applications of Theorem 6.1.6. □

Although we now know that polynomial and rational functions are differentiable, how do we actually compute these derivatives? Once we find the derivatives of functions of the form $f(x) = x^n$ for each $n \in \mathbb{N}^+$, we can apply Theorem 6.1.6. Of course, the answer is the Power Rule, which says that if $n \in \mathbb{N}^+$ and $f(x) = x^n$, then $f'(c) = n \cdot c^{n-1}$ for all $c \in \mathbb{R}$. There are several proofs of this result, but here are outlines of the three most common approaches:

- Notice that if $f(x) = x$, then $f'(c) = 1$ for all $c \in \mathbb{N}^+$. Now use induction on $n \in \mathbb{N}^+$ together with the Product Rule from Theorem 6.1.6.

- Use the fact that

$$x^n - c^n = (x - c)(x^{n-1} + cx^{n-2} + c^2x^{n-3} + \cdots + c^{n-2}x + c^{n-1}),$$

  for all $x, c \in \mathbb{R}$ and $n \in \mathbb{N}^+$, and apply the definition of the derivative.

- Use the Binomial Theorem on $(c + h)^n$, and apply Proposition 6.1.4.

We leave the details of each of these arguments as an exercise, and turn out attention to the relationship between differentiability and optimization.

Definition 6.1.9. *Let A be an interval, let $f\colon A \to \mathbb{R}$, and let $c \in int(A)$.*

- *We say that c is a* local maximum *if there exists $\delta > 0$ such that $V_\delta(c) \subseteq A$ and where $f(x) \leq f(c)$ for all $x \in V_\delta(c)$.*

- *We say that c is a* local minimum *if there exists $\delta > 0$ such that $V_\delta(c) \subseteq A$ and where $f(x) \geq f(c)$ for all $x \in V_\delta(c)$.*

Proposition 6.1.10. *Let A be an interval, let $f\colon A \to \mathbb{R}$, and let $c \in int(A)$. If c is a local maximum or minimum, and if f is differentiable at c, then $f'(c) = 0$.*

Proof. Define $g\colon A\backslash\{c\} \to \mathbb{R}$ by letting

$$g(x) = \frac{f(x) - f(c)}{x - c}.$$

We are assuming that f is differentiable at c, so $\lim_{x\to c} g(x)$ exists. Let $\ell = \lim_{x\to c} g(x)$. Suppose, without loss of generality, that c is a local maximum of f (the case of a local minimum is completely analogous). We show that $\ell = 0$ by ruling out that other two possible possibilities.

- *Case 1:* Suppose that $\ell > 0$. Since c is a local maximum of f, we can fix $\delta_1 > 0$ such that $V_{\delta_1}(c) \subseteq A$ and such that $f(x) \leq f(c)$ for all $x \in V_{\delta_1}(c)$. Since $\ell > 0$, we can use Proposition 5.1.4 to fix $\delta_2 > 0$ such that $g(x) > 0$ for all $x \in A \cap (V_{\delta_2}(c)\backslash\{c\})$. Let $\delta = \min\{\delta_1, \delta_2\}$. Notice that on the one hand we have
$$g(c + \delta/2) = \frac{f(c + \delta/2) - f(c)}{\delta/2} \leq 0$$
because $f(c + \delta/2) \leq f(c)$. However, on the other hand we also must have $g(c + \delta/2) > 0$, which is a contradiction.

- *Case 2:* Suppose that $\ell < 0$. Since c is a local maximum of f, we can fix $\delta_1 > 0$ such that $V_{\delta_1}(c) \subseteq A$ and such that $f(x) \leq f(c)$ for all $x \in V_{\delta_1}(c)$. Since $\ell < 0$, we can use Proposition 5.1.4 to fix $\delta_2 > 0$ such that $g(x) < 0$ for all $x \in A \cap (V_{\delta_2}(c)\backslash\{c\})$. Let $\delta = \min\{\delta_1, \delta_2\}$. Notice that on the one hand we have
$$g(c - \delta/2) = \frac{f(c - \delta/2) - f(c)}{-\delta/2} \geq 0$$
because $f(c + \delta/2) \leq f(c)$. However, on the other hand we also must have $g(c - \delta/2) < 0$, which is a contradiction.

Therefore, we must have $\ell = 0$, i.e. that $f'(c) = 0$. □

Taken together with the Extreme Value Theorem (Theorem 5.2.10), we can use Proposition 6.1.10 as the justification for the process of finding absolute maximum (resp. minimum) that you learned in Calculus. That is, suppose that $f\colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b). Since $[a, b]$ is compact and f is continuous on $[a, b]$, we know from the Extreme Value Theorem that f attains an absolute maximum (resp. minimum) on $[a, b]$. Now if the absolute maximum (resp. minimum) occurs at a point c in the interior (a, b), then c is certainly a local maximum (resp. minimum) of f, so $f'(c) = 0$. Thus, to find that the point(s) c where f attains the absolute maximum (resp. minimum), it suffices to examine only the interior points c where $f'(c) = 0$, as well as the endpoints a and b.

We can also apply these ideas to obtain the following geometrically reasonable, yet fundamental, result.

Theorem 6.1.11 (Rolle's Theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f\colon [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Suppose that $f(a) = f(b)$. There exists $c \in (a, b)$ with $f'(c) = 0$.*

Proof. We have two cases:

- *Case 1:* Suppose that f is constant on $[a, b]$, i.e. there exists $r \in \mathbb{R}$ with $f(x) = r$ for all $x \in [a, b]$. We then have that $f'(x) = 0$ for all $x \in (a, b)$, so we can let c be any element of (a, b), such as $c = \frac{a+b}{2}$.

- *Case 2:* Suppose that f is not constant on $[a, b]$. The set $[a, b]$ is compact by Proposition 4.3.5, so as f is continuous on $[a, b]$, we know from the Extreme Value Theorem that f achieves a maximum and minimum value on $[a, b]$. Since f is not constant on $[a, b]$, we know that
$$\max\{f(x) : x \in [a, b]\} \neq \min\{f(x) : x \in [a, b]\},$$

and thus (at least) one of these values is distinct from the common value $f(a) = f(b)$. Suppose that $\max\{f(x) : x \in [a,b]\} \neq f(a)$. Fix $c \in [a,b]$ with $f(c) = \max\{f(x) : x \in [a,b]\}$. Since $f(a) = f(b)$, we know that $c \neq a$ and $c \neq b$, so $c \in (a,b)$. Moreover, as $f(c) = \max\{f(x) : x \in [a,b]\}$, we know that c is a local maximum of f. Since f is differentiable on (a,b) and $c \in (a,b)$, we can apply Proposition 6.1.10 to conclude that $f'(c) = 0$. The case where $\min\{f(x) : x \in [a,b]\} \neq f(a)$ is similar.

□

Unfortunately, Rolle's Theorem is difficult to use directly. Fortunately, it is the essential stepping stone to the following result, which is almost certainly the most important result about differentiation. Intuitively, it says that if you form the secant line between $(a, f(a))$ and $(b, f(b))$, then you can always find a point $c \in (a,b)$ where the corresponding tangent line has the same slope as this secant line.

Theorem 6.1.12 (Mean Value Theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f\colon [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). There exists $c \in (a,b)$ with*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The key geometric observation is that it appears such a point will happen when the graph of f is furthest away from the graph of the secant line. Now the secant line has equation

$$y = f(a) + \frac{f(b) - f(a)}{b - a} \cdot (x - a),$$

so this suggests looking at the function that gives the difference between $f(x)$ and this line. Here are the details.

Proof. Define a function $h\colon [a,b] \to \mathbb{R}$ by letting

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a} \cdot (x - a).$$

Since f is continuous on $[a,b]$, and we know that polynomials are continuous by Corollary 5.2.7, we can use Theorem 5.2.6 to conclude that h is continuous on $[a,b]$. Also, since f is differentiable on (a,b), we can use Corollary 6.1.8 and Theorem 6.1.6 to conclude that h is differentiable on (a,b) and that

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

for all $x \in (a,b)$. Now $h(a) = 0 = h(b)$, so by Rolle's Theorem, we can fix $c \in (a,b)$ with $h'(c) = 0$. Thus, we have

$$0 = f'(c) - \frac{f(b) - f(a)}{b - a},$$

and hence

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

□

The Mean Value Theorem is the central result that allows one to take information about the derivative of a function f' and use it to draw conclusions about the original function f. Here is a simple example.

Proposition 6.1.13. *Let A be a nonempty interval, and let $f\colon A \to \mathbb{R}$ be continuous on A and differentiable on $\text{int}(A)$. Suppose that $f'(x) = 0$ for all $x \in \text{int}(A)$. We then have that f is constant on A, i.e. there exists $r \in \mathbb{R}$ with $f(x) = r$ for all $r \in A$.*

Proof. Since A is nonempty, we can fix some $a \in A$. We show that $f(x) = f(a)$ for all $x \in A$ (i.e. we can take $r = f(a)$). So let $x \in A$ be arbitrary. We have three cases:

- *Case 1:* Suppose that $x = a$. We then trivially have $f(x) = f(a)$.
- *Case 2:* Suppose that $a < x$. Since A is an interval and $a, x \in A$, we know that $(a, x) \subseteq A$, and hence $(a, x) = \text{int}((a, x)) \subseteq \text{int}(A)$. Thus, by assumption, we know that f is continuous on $[a, x]$ and differentiable on (a, x). Using the Mean Value Theorem, we can fix $c \in (a, x)$ with

  $$f'(c) = \frac{f(x) - f(a)}{x - a}.$$

  By assumption, we know that $f'(c) = 0$, so

  $$\frac{f(x) - f(a)}{x - a} = 0.$$

  Multiplying both sides by $x - a$, and then adding $f(a)$ to both sides, we conclude that $f(x) = f(a)$.
- *Case 3:* Suppose that $x < a$. Since A is an interval and $x, a \in A$, we know that $(x, a) \subseteq A$, and hence $(x, a) = \text{int}((x, a)) \subseteq \text{int}(A)$. Thus, by assumption, we know that f is continuous on $[x, a]$ and differentiable on (x, a). Using the Mean Value Theorem, we can fix $c \in (a, x)$ with

  $$f'(c) = \frac{f(a) - f(x)}{a - x}.$$

  By assumption, we know that $f'(c) = 0$, so

  $$\frac{f(a) - f(x)}{a - x} = 0.$$

  Multiplying both sides by $a - x$, and then adding $f(a)$ to both sides, we conclude that $f(x) = f(a)$.

Thus, in all cases, we have shown that $f(x) = f(a)$. □

Corollary 6.1.14. *Let A be a nonempty interval, and let $f, g \colon A \to \mathbb{R}$ be continuous on A and differentiable on int(A). Suppose that $f'(x) = g'(x)$ for all $x \in$ int(A). There exists $r \in \mathbb{R}$ such that $f(x) = g(x) + r$ for all $r \in A$.*

Proof. Apply the previous result to the function $h \colon A \to \mathbb{R}$ defined by $h(x) = f(x) - g(x)$. □

We can also use the Mean Value Theorem to show how the derivative affects the "shape" of the graph.

Definition 6.1.15. *Let A be a nonempty interval, and let $f \colon A \to \mathbb{R}$ be continuous on A and differentiable on int(A).*

1. *f is* increasing *on A if for all $a, b \in A$ with $a < b$, we have $f(a) \leq f(b)$.*
2. *f is* strictly increasing *on A if for all $a, b \in A$ with $a < b$, we have $f(a) < f(b)$.*
3. *f is* decreasing *on A if for all $a, b \in A$ with $a < b$, we have $f(a) \geq f(b)$.*
4. *f is* strictly decreasing *on A if for all $a, b \in A$ with $a < b$, we have $f(a) > f(b)$.*
5. *f is* monotonic *if it is either increasing or decreasing.*

Proposition 6.1.16. *Let A be a nonempty interval, and let $f \colon A \to \mathbb{R}$ be continuous on A and differentiable on int(A).*

1. *If $f'(x) \geq 0$ for all $x \in A$, then f is increasing on A.*
2. *If $f'(x) > 0$ for all $x \in A$, then f is strictly increasing on A.*
3. *If $f'(x) \leq 0$ for all $x \in A$, then f is decreasing on A.*
4. *If $f'(x) < 0$ for all $x \in A$, then f is strictly decreasing on A.*

Proof. Exercise. □

We can also use the Mean Value Theorem to argue that if the slopes of the tangents lines of a function never become too steep, then the function is uniformly continuous. The main interest here is the case where the interval is either unbounded or not closed, as Theorem 5.2.12 would handle the general case of a closed and bounded interval (even without any assumption on the derivative).

Proposition 6.1.17. *Let A be an interval, and let $f\colon A \to \mathbb{R}$ be continuous on A and differentiable on $\text{int}(A)$. Suppose that f' exists and is bounded on $\text{int}(A)$, i.e. that there exists $d \in \mathbb{R}$ such that $|f'(x)| \leq d$ for all $x \in \text{int}(A)$. We then have that f is uniformly continuous on A.*

Proof. Fix $d \in \mathbb{R}$ with $|f'(x)| \leq d$ for all $x \in A$. Notice that we can assume that $d > 0$, because we can always choose a larger value of d (i.e. the inequality is still true if we increase d). Now let $\varepsilon > 0$ be arbitrary. Consider $\delta = \frac{\varepsilon}{d} > 0$, and let $x, y \in A$ be arbitrary with $|x - y| < \delta$. We show that $|f(x) - f(y)| < \varepsilon$ by considering three cases:

- *Case 1:* Suppose that $x = y$. We then have $|f(x) - f(y)| = 0 < \varepsilon$.
- *Case 2:* Suppose that $y < x$. Since A is an interval and $x, y \in A$, we know that $(y, x) \subseteq A$, and hence $(y, x) = \text{int}((y, x)) \subseteq \text{int}(A)$. Thus, by assumption, we know that f is continuous on $[y, x]$ and differentiable on (y, x). Using the Mean Value Theorem, we can fix $c \in (y, x)$ with

  $$f'(c) = \frac{f(x) - f(y)}{x - y}.$$

  Multiplying both sides by $x - y$ and taking the absolute value, it follows that

  $$\begin{aligned} |f(x) - f(y)| &= |f'(c)| \cdot |x - y| \\ &< d \cdot \frac{\varepsilon}{d} \\ &= \varepsilon. \end{aligned}$$

- *Case 3:* Suppose that $x < y$. Since A is an interval and $x, y \in A$, we know that $(x, y) \subseteq A$, and hence $(x, y) = \text{int}((x, y)) \subseteq \text{int}(A)$. Thus, by assumption, we know that f is continuous on $[x, y]$ and differentiable on (x, y). Using the Mean Value Theorem, we can fix $c \in (x, y)$ with

  $$f'(c) = \frac{f(y) - f(x)}{y - x}.$$

  Multiplying both sides by $y - x$ and taking the absolute value, it follows that

  $$\begin{aligned} |f(x) - f(y)| &= |f(y) - f(x)| \\ &= |f'(c)| \cdot |y - x| \\ &< d \cdot \frac{\varepsilon}{d} \\ &= \varepsilon. \end{aligned}$$

Thus, we have $|f(x) - f(y)| < \varepsilon$ in all cases. Since $\varepsilon > 0$ was arbitrary, we conclude that f is uniformly continuous on A. □

We end this section with a powerful generalization of the Mean Value Theorem.

Theorem 6.1.18 (Cauchy's Mean Value Theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f, g\colon [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). There exists $c \in (a,b)$ with*

$$f'(c) \cdot [g(b) - g(a)] = g'(c) \cdot [f(b) - f(a)].$$

Thus, if $g(a) \neq g(b)$ and $f(a) \neq f(b)$, then there exists $c \in (a,b)$ with

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Notice that if $g(x) = x$, then this result reduces to the Mean Value Theorem. But for a general g, the intuition here is less clear. To view this result geometrically, consider the parametric curve $x = g(t)$, $y = f(t)$ on the interval $[a,b]$, and assume that $g(a) \neq g(b)$ and $f(a) \neq f(b)$. The quantity

$$\frac{f(b) - f(a)}{g(b) - g(a)}$$

is the slope of the line from the initial point $(g(a), f(a))$ to the final point $(g(b), f(b))$. Now the tangent vector to the curve at any point t is given by $\langle g'(t), f'(t)\rangle$, so the quantity $\frac{f'(c)}{g'(c)}$ on the left is the slope of the tangent line at the corresponding point. Thus, Cauchy's Mean Value Theorem is saying that if you form the secant line between $(g(a), f(a))$ and $(g(b), f(b))$, then you can always find a point $c \in (a,b)$ where the corresponding tangent vector is parallel to the secant line.

Proof. Define $h\colon [a,b]$ by letting

$$h(x) = [f(b) - f(a)] \cdot g(x) - [g(b) - g(a)] \cdot f(x).$$

Since f and g are continuous on $[a,b]$, and we know that polynomials are continuous by Corollary 5.2.7, we can use Theorem 5.2.6 to conclude that h is continuous on $[a,b]$. Also, since f and g are differentiable on (a,b), we can use Corollary 6.1.8 and Theorem 6.1.6 to conclude that h is differentiable on (a,b) and that

$$h'(x) = [f(b) - f(a)] \cdot g'(x) - [g(b) - g(a)] \cdot f'(x).$$

Notice that

$$\begin{aligned} h(a) &= [f(b) - f(a)] \cdot g(a) - [g(b) - g(a)] \cdot f(a) \\ &= f(b)g(a) - f(a)g(a) - g(b)f(a) + g(a)f(a) \\ &= f(b)g(a) - f(a)g(b) \end{aligned}$$

and

$$\begin{aligned} h(b) &= [f(b) - f(a)] \cdot g(b) - [g(b) - g(a)] \cdot f(b) \\ &= f(b)g(b) - f(a)g(b) - g(b)f(b) + g(a)f(b) \\ &= f(b)g(a) - f(a)g(b). \end{aligned}$$

Since $h(a) = h(b)$, we know from Rolle's Theorem that we can fix $c \in (a,b)$ with $h'(c) = 0$. Using the above formula for h', it follows that

$$[f(b) - f(a)] \cdot g'(c) - [g(b) - g(a)] \cdot f'(c) = 0.$$

□

6.2 The Riemann Integral

We now move on to a careful definition of the integral of a (bounded) function defined on a closed and bounded interval $[a, b]$. It is possible to handle unbounded functions and a wider class of domains, but we will see that it is much easier to begin in this simple setting. If you have seen improper integrals, that is one way to extend our approach to these other situations using limits.

Suppose then that we have a bounded function $f\colon [a, b] \to \mathbb{R}$. Think about the region R between the graph of f and the x-axis. We would like to define $\int_a^b f$ to be the (signed) area of this region, where we count the parts below the x-axis negatively. Of course, the key issue is that we don't know how to assign area to arbitrary regions.

In Calculus, you likely defined the integral of a function f on an interval $[a, b]$ as a limit of Riemann sums. Although this approach can be turned into a precise definition, the idea of taking "test points" as the heights of the various rectangles adds a nontrivial amount of complexity. Moreover, when using such test points, it is not clear if the resulting Riemann sum is an under-approximation or over-approximation. Darboux adjusted the core ideas behind Riemann sums to create the following approach.

Definition 6.2.1. *Given a closed bounded interval $[a, b]$, a* partition *of $[a, b]$ is a finite sequence $P = (p_0, p_1, p_2, \ldots, p_n)$ with $p_0 = a$, $p_n = b$, and $p_{i-1} < p_i$ for all i with $1 \leq i \leq n$.*

Given a bounded function $f\colon [a, b] \to \mathbb{R}$ together with a partition P of $[a, b]$, we can now take infimums and supremums on each subinterval described by the partition. If we us these values as the "heights" (recalling that for negative values, we assign a negative "height"), we obtain the following values.

Definition 6.2.2. *Let $f\colon [a, b] \to \mathbb{R}$ be a bounded function, and let $P = (p_0, p_1, p_2, \ldots, p_n)$ be a partition of $[a, b]$. For each $i \in \{1, 2, \ldots, n\}$, let*

$$m_i = \inf\{f(x) : x \in [p_{i-1}, p_i]\}$$
$$M_i = \sup\{f(x) : x \in [p_{i-1}, p_i]\}.$$

We then define

$$L(f, P) = \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1})$$

and

$$U(f, P) = \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}).$$

Now if we have a bounded function $f\colon [a, b] \to \mathbb{R}$, then for any partition P of $[a, b]$, we expect $L(f, P)$ to be a lower bound on the elusive (signed) area that we are seeking, and we expect $U(f, P)$ to be an upper bound. However, we want to turn these ideas on their heads and *use* the various $L(f, P)$ and $U(f, P)$ to give a formal way to *define* what we mean by such a (signed) area. In order to do that, we first have to establish a few fundamental inequalities.

Proposition 6.2.3. *Let $f\colon [a, b] \to \mathbb{R}$ be a bounded function, and let P be a partition of $[a, b]$. We then have $L(f, P) \leq U(f, P)$.*

Proof. For each i, the set $\{f(x) : x \in [p_{i-1}, p_i]\}$ is nonempty, so we know that $m_i \leq M_i$. Using the fact that

$p_i - p_{i-1} > 0$ for each i, it follows that $m_i \cdot (p_i - p_{i-1}) \leq M_i \cdot (p_i - p_{i-1})$ for each i. Therefore,

$$\begin{aligned} L(f,P) &= \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\ &\leq \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}) \\ &= U(f,P). \end{aligned}$$

□

Since we always expect each $L(f,P)$ to be a lower bound on the value we seek, and we expect each $U(f,Q)$ to be an upper bound, it is natural to believe that $L(f,P) \leq U(f,Q)$ for any two partitions P and Q of $[a,b]$. In order to prove this, we introduce the notion of a refinement of a partition.

Definition 6.2.4. *Let $P = (p_0, p_1, \ldots, p_n)$ and $R = (r_0, r_1, \ldots, r_m)$ be partitions of an interval $[a,b]$. We say that R is a* refinement *of P if $\{p_0, p_1, \ldots, p_n\} \subseteq \{r_0, r_1, \ldots, r_m\}$.*

In order words, R is a refinement of P if it contains all of the points of the partition P, but might also include some new points. Intuitively, a refinement of P breaks up the interval $[a,b]$ into finer pieces, so both the corresponding lower and upper bounds should only improve. We now verify this intuition.

Proposition 6.2.5. *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded function. Suppose that P and R are partitions, and that R is a refinement of P. We then have that $L(f,P) \leq L(f,R)$ and $U(f,R) \leq U(f,P)$.*

Proof. Let $P = (p_0, p_1, \ldots, p_n)$ and let R be a refinement of P. It suffices to consider the case where R is a refinement of P with just one additional point, because then we can use the one-point result and induction to establish the general case. Suppose then that R contains exactly one more point that P. We can then fix $r \in \mathbb{R}$ and $k \in \{1, 2, \ldots, n\}$ such that $R = (p_0, p_1, \ldots, p_{k-1}, r, p_k, \ldots, p_n)$. Define the m_i and M_i for the partition P as in Definition 6.2.2. But now also define m', m'', M', M'' as follows:

$$\begin{aligned} m' &= \inf\{f(x) : x \in [p_{k-1}, r]\} \\ M' &= \sup\{f(x) : x \in [p_{k-1}, r]\} \\ m'' &= \inf\{f(x) : x \in [r, p_k]\} \\ M'' &= \sup\{f(x) : x \in [r, p_k]\} \end{aligned}$$

Since

$$\{f(x) : x \in [p_{k-1}, r]\} \subseteq \{f(x) : x \in [p_{k-1}, p_k]\},$$

we have $m_k \leq m' \leq M' \leq M_k$. Similarly, since

$$\{f(x) : x \in [r, p_k]\} \subseteq \{f(x) : x \in [p_{k-1}, p_k]\},$$

we have $m_k \leq m'' \leq M'' \leq M_k$. Therefore,

$$\begin{aligned}
L(f,R) &= \left(\sum_{i=1}^{k-1} m_i \cdot (p_i - p_{i-1})\right) + m' \cdot (r - p_{k-1}) + m'' \cdot (p_k - r) + \left(\sum_{i=k+1}^{n} m_i \cdot (p_i - p_{i-1})\right) \\
&\geq \left(\sum_{i=1}^{k-1} m_i \cdot (p_i - p_{i-1})\right) + m_k \cdot (r - p_{k-1}) + m_k \cdot (p_k - r) + \left(\sum_{i=k+1}^{n} m_i \cdot (p_i - p_{i-1})\right) \\
&= \left(\sum_{i=1}^{k-1} m_i \cdot (p_i - p_{i-1})\right) + m_k \cdot (r - p_{k-1} + p_k - r) + \left(\sum_{i=k+1}^{n} m_i \cdot (p_i - p_{i-1})\right) \\
&= \left(\sum_{i=1}^{k-1} m_i \cdot (p_i - p_{i-1})i\right) + m_k \cdot (p_k - p_{k-1}) + \left(\sum_{i=k+1}^{n} m_i \cdot (p_i - p_{i-1})\right) \\
&= \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\
&= L(f,P),
\end{aligned}$$

and

$$\begin{aligned}
U(f,R) &= \left(\sum_{i=1}^{k-1} M_i \cdot (p_i - p_{i-1})\right) + M' \cdot (r - p_{k-1}) + M'' \cdot (p_k - r) + \left(\sum_{i=k+1}^{n} M_i \cdot (p_i - p_{i-1})\right) \\
&\leq \left(\sum_{i=1}^{k-1} M_i \cdot (p_i - p_{i-1})\right) + M_k \cdot (r - p_{k-1}) + M_k \cdot (p_k - r) + \left(\sum_{i=k+1}^{n} M_i \cdot (p_i - p_{i-1})\right) \\
&= \left(\sum_{i=1}^{k-1} M_i \cdot (p_i - p_{i-1})\right) + M_k \cdot (r - p_{k-1} + p_k - r) + \left(\sum_{i=k+1}^{n} M_i \cdot (p_i - p_{i-1})\right) \\
&= \left(\sum_{i=1}^{k-1} M_i \cdot (p_i - p_{i-1})i\right) + M_k \cdot (p_k - p_{k-1}) + \left(\sum_{i=k+1}^{n} M_i \cdot (p_i - p_{i-1})\right) \\
&= \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}) \\
&= U(f,P).
\end{aligned}$$

This completes the proof. □

We can now prove the result that we are after.

Proposition 6.2.6. *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded function. If P and Q are both partitions of $[a,b]$, then $L(f,P) \leq U(f,Q)$.*

Proof. Let $P = (p_0, p_1, \ldots, p_n)$ and $Q = (q_0, q_1, \ldots, q_m)$ be partitions of $[a,b]$. Place the elements of the set $\{p_0, p_1, \ldots, p_n\} \cup \{q_0, q_1, \ldots, q_m\}$ in increasing order to from a new partition R of $[a,b]$. Notice that R is a common refinement of P and Q, so

$$\begin{aligned}
L(f,P) &\leq L(f,R) && \text{(by Proposition 6.2.5)} \\
&\leq U(f,R) && \text{(by Proposition 6.2.3)} \\
&\leq U(f,Q) && \text{(by Proposition 6.2.5).}
\end{aligned}$$

□

With all of that hard work in hand, we consider the set of *all* lower approximations, and the set of *all* upper approximations. That is, we look at the following two sets:

$$A = \{L(f,P) : P \text{ is a partition of } [a,b]\}$$
$$B = \{U(f,P) : P \text{ is a partition of } [a,b]\}.$$

Notice that A and B are both nonempty set of reals, and we know from Proposition 6.2.6 that $a \leq b$ whenever $a \in A$ and $b \in B$. In particular, A is bounded above and B is bounded below. Moreover, by Problem 3 on Homework 2, we know that $\sup A \leq \inf B$, i.e. we have

$$\sup\{L(f,P) : P \text{ is a partition of } [a,b]\} \leq \inf\{U(f,P) : P \text{ is a partition of } [a,b]\}.$$

Now recall that $A = \{L(f,P) : P \text{ is a partition of } [a,b]\}$ is the set of all lower approximations of the value that we seek, and the set $B = \{U(f,P) : P \text{ is a partition of } [a,b]\}$ is the set of all upper approximations. Thus, if $\sup A = \inf B$, then we have isolated one number that sits between every $L(f,P)$ and every $U(f,Q)$, and that common value would be a natural candidate for the number

Definition 6.2.7. *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded function. We define*

$$\underline{\int_a^b} f = L(f) = \sup\{L(f,P) : P \textit{ is a partition of } [a,b]\}$$

and

$$\overline{\int_a^b} f = U(f) = \inf\{U(f,P) : P \textit{ is a partition of } [a,b]\}$$

From above, we know that $L(f) \leq U(f)$. If these two values are equal, then we say that f is Riemann integrable*, or simple* integrable*, and we define $\int_a^b f$ to be this common value.*

Let's see some examples. Consider the function $f(x) = x$ on the interval $[0,1]$. Let $n \in \mathbb{N}^+$, and consider the partition $P_n = (0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1)$, i.e. P_n is the partition where $p_i = \frac{i}{n}$. Since f is increasing on $[0,1]$, we have that $m_i = f(p_{i-1})$ and $M_i = f(p_i)$ for each i, so

$$\begin{aligned}
L(f,P_n) &= \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\
&= \sum_{i=1}^{n} \frac{i-1}{n} \cdot \left(\frac{i}{n} - \frac{i-1}{n}\right) \\
&= \sum_{i=1}^{n} \frac{i-1}{n} \cdot \frac{1}{n} \\
&= \frac{1}{n^2} \sum_{i=1}^{n} (i-1) \\
&= \frac{1}{n^2} \sum_{i=1}^{n-1} i \\
&= \frac{1}{n^2} \cdot \frac{(n-1)\cdot n}{2} \\
&= \frac{n-1}{2n} \\
&= \frac{1}{2} - \frac{1}{2n}.
\end{aligned}$$

and

$$\begin{aligned} U(f, P_n) &= \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}) \\ &= \sum_{i=1}^{n} \frac{i}{n} \cdot \left(\frac{i}{n} - \frac{i-1}{n} \right) \\ &= \sum_{i=1}^{n} \frac{i}{n} \cdot \frac{1}{n} \\ &= \frac{1}{n^2} \sum_{i=1}^{n} i \\ &= \frac{1}{n^2} \cdot \frac{n \cdot (n+1)}{2} \\ &= \frac{n+1}{2n} \\ &= \frac{1}{2} + \frac{1}{2n}. \end{aligned}$$

We have shown that

$$\left\{ \frac{1}{2} - \frac{1}{2n} : n \in \mathbb{N}^+ \right\} \subseteq \{L(f, P) : P \text{ is a partition of } [a, b]\},$$

so it follows that

$$\begin{aligned} L(f) &= \sup\{L(f, P) : P \text{ is a partition of } [a, b]\} \\ &\geq \sup \left\{ \frac{1}{2} - \frac{1}{2n} : n \in \mathbb{N}^+ \right\} \\ &= \frac{1}{2}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} U(f) &= \inf\{L(U, P) : P \text{ is a partition of } [a, b]\} \\ &\leq \inf \left\{ \frac{1}{2} + \frac{1}{2n} : n \in \mathbb{N}^+ \right\} \\ &= \frac{1}{2}. \end{aligned}$$

Combining this with the general fact that $L(f) \leq U(f)$, we have

$$L(f) \leq U(f) \leq \frac{1}{2} \leq L(f),$$

so $L(f) = U(f) = \frac{1}{2}$. Therefore, f is integrable on $[0, 1]$, and $\int_0^1 f = \frac{1}{2}$.

Now consider the more exotic function $f \colon [0, 1] \to \mathbb{R}$ defined by letting

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{otherwise.} \end{cases}$$

Let $P = (p_0, p_1, \ldots, p_n)$ be an arbitrary partition of $[0,1]$. For any i, we know that (p_{i-1}, p_i) contains both rationals and irrationals (by Theorem 1.4.8 and Corollary 1.5.12), so we have $m_i = 0$ and $M_i = 1$. Thus, $L(f,P) = 0$ and $U(f,P) = 1$. Therefore, we have

$$\{L(f,P) : P \text{ is a partition of } [a,b]\} = \{0\}$$
$$\{U(f,P) : P \text{ is a partition of } [a,b]\} = \{1\}.$$

It follows that $L(f) = 0 < 1 = U(f)$, and hence f is not integrable on $[0,1]$.

Instead of working to determine the actual values of $L(f)$ and $U(f)$ for a given function (which in general is incredibly difficult), we will typically use the following result to argue that a function if integrable.

Proposition 6.2.8. *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded function. The following are equivalent:*

1. *f is Riemann integrable.*
2. *For all $\varepsilon > 0$, there exists partitions P and Q of $[a,b]$ with $U(f,Q) - L(f,P) < \varepsilon$.*
3. *For all $\varepsilon > 0$, there exists a partition R of $[a,b]$ with $U(f,R) - L(f,R) < \varepsilon$.*

Proof. It suffices to prove the following implications:

- (1) $\Rightarrow$ (2): Suppose first that f is Riemann integrable, and let ℓ be the common value $L(f) = U(f)$. Let $\varepsilon > 0$. Since $\ell - \frac{\varepsilon}{2} < \ell = \sup\{L(f,P) : P \text{ is a partition of } [a,b]\}$, we know that $\ell - \frac{\varepsilon}{2}$ is not an upper bound of $\{L(f,P) : P \text{ is a partition of } [a,b]\}$, and hence we can fix a partition P of $[a,b]$ with $L(f,P) > \ell - \frac{\varepsilon}{2}$. Similarly, since $\ell + \frac{\varepsilon}{2} > \ell = \inf\{U(f,Q) : Q \text{ is a partition of } [a,b]\}$, we know that $\ell + \frac{\varepsilon}{2}$ is not a lower bound of $\{U(f,Q) : Q \text{ is a partition of } [a,b]\}$, and hence we can fix a partition Q of $[a,b]$ with $U(f,P) < \ell + \frac{\varepsilon}{2}$. Since $L(f,P) > \ell - \frac{\varepsilon}{2}$, we know that $-L(f,P) < \frac{\varepsilon}{2} - \ell$, and hence

$$\begin{aligned} U(f,Q) - L(f,P) &< \left(\ell + \frac{\varepsilon}{2}\right) + \left(\frac{\varepsilon}{2} - \ell\right) \\ &= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

- (2) $\Rightarrow$ (3): Assume (2). Let $\varepsilon > 0$. By assumption, we can fix partitions P and Q of $[a,b]$ with $U(f,Q) - L(f,P) < \varepsilon$. Place the elements of the set $\{p_0, p_1, \ldots, p_n\} \cup \{q_0, q_1, \ldots, q_m\}$ in increasing order to from a new partition R of $[a,b]$. Notice that R is a common refinement of P and Q, so

$$\begin{aligned} L(f,P) &\leq L(f,R) && \text{(by Proposition 6.2.5)} \\ &\leq U(f,R) && \text{(by Proposition 6.2.3)} \\ &\leq U(f,Q) && \text{(by Proposition 6.2.5).} \end{aligned}$$

  Therefore, we have

$$U(f,R) - L(f,R) \leq U(f,Q) - L(f,P) < \varepsilon.$$

- (3) $\Rightarrow$ (1): We prove the contrapositive. Suppose that f is not Riemann integrable. Since we know that $L(f) \leq U(f)$, we must then have $L(f) < U(f)$. Let $\varepsilon = U(f) - L(f) > 0$. For any partition R of $[a,b]$, we then have $L(f,R) \leq L(f) < U(f) \leq U(f,R)$, so

$$U(f,R) - L(f,R) \geq U(f) - L(f) \geq \varepsilon.$$

□

Our first application of this technique will be to show that every continuous function $f\colon [a,b] \to \mathbb{R}$ is integrable. Before stating this result, we need to be a bit careful. Recall that we only consider integration for bounded functions. However, every continuous $f\colon [a,b] \to \mathbb{R}$ must be bounded by Proposition 5.2.9 (since $[a,b]$ is compact, this result simple that range$(f) = f([a,b])$ is compact, and hence bounded).

Theorem 6.2.9. *If $f\colon [a,b] \to \mathbb{R}$ is continuous, then f is Riemann integrable.*

Proof. Since $[a,b]$ is compact, and f is continuous on $[a,b]$, we know from Theorem 5.2.12 that f is uniformly continuous on $[a,b]$. We use Proposition 6.2.8 to argue that f is integrable. Let $\varepsilon > 0$ be arbitrary. Since f is uniformly continuous, we can fix $\delta > 0$ such that for all $x, y \in [a,b]$ with $|x-y| < \delta$, we have $|f(x) - f(y)| < \frac{\varepsilon}{b-a}$. Create a partition P of $[a,b]$ with increments of size $\frac{\delta}{2}$. That is, let $p_0 = a$, $p_1 = a + \frac{\delta}{2}$, and in general $p_i = a + i \cdot \frac{\delta}{2}$ until we reach an n where $a + n \cdot \frac{\delta}{2} \geq b$, at which point we let $p_n = b$. For each i, we know that f is continuous on the interval $[p_{i-1}, p_i]$, so f attains a maximum and minimum value on $[p_{i-1}, p_i]$ by the Extreme Value Theorem. Thus, for each i, we can fix $x_i, y_i \in [p_{i-1}, p_i]$ with $f(x_i) = m_i$ and $f(y_i) = M_i$. For each i, notice that $|x_i - y_i| \leq \frac{\delta}{2} < \delta$, so

$$\begin{aligned} M_i - m_i &= |M_i - m_i| \\ &= |f(x_i) - f(y_i)| \\ &< \frac{\varepsilon}{b-a}, \end{aligned}$$

Therefore,

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}) - \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\ &= \sum_{i=1}^{n} (M_i - m_i) \cdot (p_i - p_{i-1}) \\ &< \sum_{i=1}^{n} \frac{\varepsilon}{b-a} \cdot (p_i - p_{i-1}) \\ &= \frac{\varepsilon}{b-a} \sum_{i=1}^{n} (p_i - p_{i-1}) \\ &= \frac{\varepsilon}{b-a} \cdot (p_n - p_0) \\ &= \frac{\varepsilon}{b-a} \cdot (b-a) \\ &= \varepsilon. \end{aligned}$$

Using Proposition 6.2.8, we conclude that f is integrable on $[a,b]$. □

Although continuous functions are probably the most important subclass of the integrable functions, there is another natural subclass: the monotonic functions. See Definition 6.1.15 for the relevant definitions. Since there are monotonic functions that are not continuous, this result shows that the converse of Theorem 6.2.9 is false.

Proposition 6.2.10. *If $f\colon [a,b] \to \mathbb{R}$ is monotonic, then f is Riemann integrable.*

Proof. Suppose that f is increasing on $[a,b]$ (the decreasing case is similar). We again use Proposition 6.2.8 to argue that f is integrable. Let $\varepsilon > 0$. By Proposition 1.4.5, we can fix $n \in \mathbb{N}^+$ with

$$n > \frac{(b-a)(f(b) - f(a))}{\varepsilon}.$$

Multiplying both sides by $\frac{\varepsilon}{n}$, we then have

$$\frac{(b-a)(f(b)-f(a))}{n} < \varepsilon.$$

Consider the partition P of $[a,b]$ obtained by dividing $[a,b]$ into n pieces of equal length $\frac{b-a}{n}$, so $p_i = a + i \cdot \frac{b-a}{n}$ for $0 \leq i \leq n$. Since f is increasing, we know that for each i, we have $m_i = f(p_{i-1})$ and $M_i = f(p_i)$. Therefore,

$$\begin{aligned}
U(f,P) - L(f,P) &= \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}) - \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\
&= \sum_{i=1}^{n} (M_i - m_i) \cdot (p_i - p_{i-1}) \\
&= \sum_{i=1}^{n} \left[(f(p_i) - f(p_{i-1})) \cdot \frac{b-a}{n} \right] \\
&= \frac{b-a}{n} \sum_{i=1}^{n} (f(p_i) - f(p_{i-1})) \\
&= \frac{b-a}{n} \cdot (f(p_n) - f(p_0)) \\
&= \frac{b-a}{n} \cdot (f(b) - f(a)) \\
&< \varepsilon.
\end{aligned}$$

Using Proposition 6.2.8, we conclude that f is integrable on $[a,b]$. □

We now prove a simple property of the integral that says that we can break up an integral over an interval into the integrals over the two resulting subintervals. Although the proof is long, the core ideas behind the argument are relatively simple.

Proposition 6.2.11. *Let $f\colon [a,b] \to \mathbb{R}$ be bounded, and let $c \in (a,b)$.*

- *If f is integrable on $[a,b]$, then f is integrable on $[a,c]$ and f is integrable on $[c,b]$.*
- *If f is integrable on both $[a,c]$ and $[c,b]$, then f is integrable on $[a,b]$.*

Moreover, in this case, we have

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

Proof. Suppose first that f is integrable on $[a,b]$. Let $c \in (a,b)$ be arbitrary. We use Proposition 6.2.8 to show that f is integrable on $[a,c]$ (the proof that f is integrable on $[c,b]$ is completely analogous). Let $\varepsilon > 0$. Since f is integrable on $[a,b]$, we can use Proposition 6.2.8 to fix a partition P of $[a,b]$ with $U(f,P) - L(f,P) < \varepsilon$. Let R be the partition obtained by adding c to P (if c is already a point in P, then let $R = P$). We then have that R is a refinement of P, so using Proposition 6.2.3 and Proposition 6.2.5 we conclude that $L(f,P) \leq L(f,R) \leq U(f,R) \leq U(f,P)$, and hence

$$\begin{aligned}
U(f,R) - L(f,R) &\leq U(f,P) - L(f,P) \\
&< \varepsilon.
\end{aligned}$$

Now let Q be the partition of $[a,c]$ obtained by cutting off the partition R at c, i.e. if $R=(p_0,p_1,\ldots,p_n)$ and $p_k=c$, then let $Q=(p_0,p_1,\ldots,p_k)$. Since $k\leq n$ and $(M_i-m_i)\cdot(p_i-p_{i-1})\geq 0$ for all i, we have

$$\begin{aligned}
U(f,Q)-L(f,Q) &= \sum_{i=1}^{k} M_i\cdot(p_i-p_{i-1}) - \sum_{i=1}^{n} m_i\cdot(p_i-p_{i-1}) \\
&= \sum_{i=1}^{k}(M_i-m_i)\cdot(p_i-p_{i-1}) \\
&= \sum_{i=1}^{n}(M_i-m_i)\cdot(p_i-p_{i-1}) \\
&= \sum_{i=1}^{n} M_i\cdot(p_i-p_{i-1}) - \sum_{i=1}^{n} m_i\cdot(p_i-p_{i-1}) \\
&= U(f,R)-L(f,R) \\
&< \varepsilon.
\end{aligned}$$

Therefore, f is integrable on $[a,c]$ by Proposition 6.2.8.

Suppose now that f is integrable on both $[a,c]$ and $[c,b]$. We use Proposition 6.2.8 to show that f is integrable on $[a,b]$. Let $\varepsilon>0$. Since f is integrable on $[a,b]$, we can use Proposition 6.2.8 to fix a partition P of $[a,c]$ with $U(f,P)-L(f,P)<\frac{\varepsilon}{2}$. Similarly, we can fix a partition Q of $[c,b]$ with $U(f,Q)-L(f,Q)<\frac{\varepsilon}{2}$. Let R be the partition of $[a,b]$ obtained by concatenating P with Q, but omitting the double c (which occurs once at the end of P, and once at the beginning of Q). A simple calculation shows that $L(f,R)=L(f,P)+L(f,Q)$ and $U(f,R)=U(f,P)+U(f,Q)$. Thus,

$$\begin{aligned}
U(f,R)-L(f,R) &= [U(f,P)+U(f,Q)]-[L(f,P)+L(f,Q)] \\
&= U(f,P)-L(f,P)+U(f,Q)-L(f,Q) \\
&< \frac{\varepsilon}{2}+\frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

Therefore, f is integrable on $[a,b]$ by Proposition 6.2.8.

We now prove the last part. Suppose that f is integrable on both $[a,c]$ and $[c,b]$. We show that

$$\int_a^b f \leq \int_a^c f+\int_c^b f \qquad \text{and} \qquad \int_a^c f+\int_c^b f \leq \int_a^b f$$

by showing that for all $\varepsilon>0$, we have both

$$\int_a^b f \leq \int_a^c f+\int_c^b f+\varepsilon \qquad \text{and} \qquad \int_a^c f+\int_c^b f \leq \int_a^b f+\varepsilon.$$

So let $\varepsilon>0$. Fix partitions P and Q, of $[a,c]$ and $[c,b]$ respectively, as in the second paragraph above, so that $U(f,P)-L(f,P)<\frac{\varepsilon}{2}$ and $U(f,Q)-L(f,Q)<\frac{\varepsilon}{2}$. Define the partition R of $[a,b]$ by concatenating P and Q as above, and recall that $U(f,R)-L(f,R)<\varepsilon$. We then have

$$L(f,P)\leq \int_a^c f\leq U(f,P) \qquad \text{and} \qquad L(f,Q)\leq \int_c^b f\leq U(f,Q).$$

by definition, so

$$\begin{aligned}\int_a^b f &\leq U(f,R)\\ &< L(f,R)+\varepsilon\\ &= L(f,P)+L(f,Q)+\varepsilon\\ &\leq \int_a^c f+\int_c^b f+\varepsilon\end{aligned}$$

and

$$\begin{aligned}\int_a^c f+\int_c^b f &\leq U(f,P)+U(f,Q)\\ &\leq L(f,P)+L(f,Q)+\varepsilon\\ &= L(f,R)+\varepsilon\\ &\leq \int_a^b f+\varepsilon.\end{aligned}$$

This completes the proof. □

Suppose that $f\colon [a,b]\to\mathbb{R}$ is bounded. We have some general conditions that guarantee that f is integrable on $[a,b]$, such as if f is continuous or if f is monotonic. We now prove an interesting result that says that if f is integrable on arbitrarily large initial (or terminal) subintervals of $[a,b]$, then f is integrable on all of $[a,b]$. As we will see, this little result allows us to "quarantine" a potential bad point.

Proposition 6.2.12. *Let $f\colon [a,b]\to\mathbb{R}$ be bounded.*

1. *If f is integrable on $[a,c]$ for all $c\in(a,b)$, then f is integrable on $[a,b]$.*
2. *If f is integrable on $[c,b]$ for all $c\in(a,b)$, then f is integrable on $[a,b]$.*

Proof. We prove the first (the second is completely analogous). We are assuming that f is bounded, so we can fix $d>0$ such that $|f(x)|\leq d$ for all $x\in[a,b]$. Suppose that f is integrable on $[a,c]$ for all $c\in(a,b)$. We use Proposition 6.2.8 to show that f is integrable on $[a,b]$. Let $\varepsilon>0$. Let $c=\max\{b-\frac{\varepsilon}{4d},\frac{a+b}{2}\}$, and notice that we have both $c\in(a,b)$ and $b-c<\frac{\varepsilon}{4d}$. By assumption, we know that f is integrable on $[a,c]$. Using Proposition 6.2.8, we can fix a partition $P=(p_0,p_1,\ldots,p_n)$ of $[a,c]$ such that $U(f,P)-L(f,P)<\frac{\varepsilon}{2}$. Now let Q be the partition $(p_0,p_1,\ldots,p_n,b)$ of $[a,b]$, and let

$$\begin{aligned}m &= \inf\{f(x) : x\in[p_n,b]\}\\ M &= \sup\{f(x) : x\in[p_n,b]\}.\end{aligned}$$

Since $|f(x)|\leq d$ for all $x\in[a,b]$, we have $-d\leq m\leq M\leq d$, so $M-m\leq 2d$. Therefore,

$$\begin{aligned}U(f,Q)-L(f,Q) &= [U(f,P)+M\cdot(b-p_n)]-[L(f,P)+m\cdot(b-p_n)]\\ &= U(f,P)-L(f,P)+M\cdot(b-c)-m\cdot(b-c)\\ &= U(f,P)-L(f,P)+(M-m)\cdot(b-c)\\ &< \frac{\varepsilon}{2}+2d\cdot\frac{\varepsilon}{4d}\\ &= \varepsilon.\end{aligned}$$

Using Proposition 6.2.8, we conclude that f is integrable on $[a,b]$. □

We now use these results to show that if $f\colon [a,b] \to \mathbb{R}$ is bounded and has at most one point of discontinuity, then f is integrable on $[a,b]$.

Corollary 6.2.13. *Let $f\colon [a,b] \to \mathbb{R}$ be bounded, and let $d \in [a,b]$. If f is continuous at every point of $[a,b]$ with the possible exception of d, then f is integrable on $[a,b]$.*

Proof. We have three cases:

- *Case 1:* Suppose that $d = a$. For any $c \in (a,b)$, we know that f is continuous on $[c,b]$, so f is integrable on $[c,b]$ by Theorem 6.2.9. Using Proposition 6.2.12, we conclude that f is integrable on $[a,b]$.
- *Case 2:* Suppose that $d = b$. For any $c \in (a,b)$, we know that f is continuous on $[a,c]$, so f is integrable on $[a,c]$ by Theorem 6.2.9. Using Proposition 6.2.12, we conclude that f is integrable on $[a,b]$.
- *Case 3:* Suppose that $d \in (a,b)$. For any $c \in (a,d)$, we know that f is continuous on $[a,c]$, so f is integrable on $[a,c]$ by Theorem 6.2.9. Using Proposition 6.2.12, we conclude that f is integrable on $[a,d]$. Similarly, f is integrable on $[d,b]$. Using Proposition 6.2.11, we conclude that f is integrable on $[a,b]$.

Thus, in all cases, f is integrable on $[a,b]$. □

The above result (and its proof) naturally generalize to show that if $f\colon [a,b] \to \mathbb{R}$ is bounded and has only finitely many points of discontinuity, then f is integrable on $[a,b]$.

We list several other fundamental properties of the Riemann integral, but we leave the proofs as an exercise.

Proposition 6.2.14. *Let $f, g\colon [a,b] \to \mathbb{R}$ be bounded functions that are both Riemann integrable on $[a,b]$.*

1. *The function $f+g$ is Riemann integrable on $[a,b]$ and $\int_a^b (f+g) = \int_a^b f + \int_a^b g$.*
2. *For all $k \in \mathbb{R}$, the function $k \cdot f$ is Riemann integrable on $[a,b]$ and $\int_a^b k \cdot f = k \cdot \int_a^b f$.*
3. *If $m \le f(x) \le M$ for all $x \in [a,b]$, then $m(b-a) \le \int_a^b f \le M(b-a)$.*
4. *If $f(x) = c$ for all $x \in [a,b]$, then $\int_a^b f = c \cdot (b-a)$.*
5. *If $f(x) \le g(x)$ for all $x \in [a,b]$, then $\int_a^b f \le \int_a^b g$.*
6. *The function $|f|$ is Riemann integrable on $[a,b]$ and $|\int_a^b f| \le \int_a^b |f|$.*

Proof. Exercise. □

We now have developed all of the tools that we need for the Fundamental Theorem of Calculus.

Theorem 6.2.15 (Fundamental Theorem of Calculus)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be integrable.*

1. *Suppose that $F\colon [a,b] \to \mathbb{R}$ is continuous on $[a,b]$, is differentiable on (a,b), and satisfies $F'(t) = f(t)$ for all $t \in (a,b)$. We then have $\int_a^b f = F(b) - F(a)$.*
2. *Define $G\colon [a,b] \to \mathbb{R}$ by letting $G(x) = \int_a^x f$. We then have that G is continuous on $[a,b]$. Moreover, if f is continuous at a point $c \in (a,b)$, then G is differentiable at c, and $G'(c) = f(c)$.*

Proof. 1. We first show that for all partitions P of $[a,b]$, we have $L(f,P) \leq F(b) - F(a)$. Let $P = (p_0, p_1, \ldots, p_n)$ be an arbitrary partition of $[a,b]$. By the Mean Value Theorem, for each i, we can fix $x_i \in (p_{i-1}, p_i)$ with

$$F'(x_i) = \frac{F(p_i) - F(p_{i-1})}{p_i - p_{i-1}}.$$

We then have $m_i \leq f(x_i)$ for all i, so

$$\begin{aligned} L(f,P) &= \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\ &\leq \sum_{i=1}^{n} f(x_i) \cdot (p_i - p_{i-1}) \\ &= \sum_{i=1}^{n} F'(x_i) \cdot (p_i - p_{i-1}) \\ &= \sum_{i=1}^{n} (F(p_i) - F(p_{i-1})) \\ &= F(p_n) - F(p_0) \\ &= F(b) - F(a). \end{aligned}$$

Therefore, for every partition P of $[a,b]$, we have $L(f,P) \leq F(b) - F(a)$. It follows that $F(b) - F(a)$ is an upper bounded of $\{L(f,P) : P \text{ is a partition of } [a,b]\}$. Since $L(f)$ is the least upper bounded of this set, we conclude that $L(f) \leq F(b) - F(a)$.

Now a similar argument shows that $F(b) - F(a) \leq U(f,P)$ for all partition P (simply switch the m_i with M_i and the $\leq$ with $\geq$ in the above calculation), and hence $F(b) - F(a) \leq U(f)$. Since $\int_a^b f = L(f) = U(f)$, this common value must be $F(b) - F(a)$.

2. Since f is bounded, we can fix a $d > 0$ with $|f(x)| \leq d$ for all $x \in [a,b]$. Notice that if $x, y \in [a,b]$ and $y < x$, then

$$\begin{aligned} |G(x) - G(y)| &= \left| \int_a^x f - \int_a^y f \right| \\ &= \left| \int_a^y f + \int_y^x f - \int_a^y f \right| && \text{(by Proposition 6.2.11)} \\ &= \left| \int_y^x f \right| \\ &\leq \int_y^x |f| && \text{(by Proposition 6.2.14)} \\ &\leq d \cdot (x - y) && \text{(by Proposition 6.2.14)} \\ &= d \cdot |x - y| && \text{(since } y < x\text{)}. \end{aligned}$$

Now if $x, y \in [a,b]$ and $x < y$, then

$$\begin{aligned} |G(x) - G(y)| &= |G(y) - G(x)| \\ &\leq d \cdot |y - x| && \text{(from above since } x < y\text{)} \\ &= d \cdot |x - y|. \end{aligned}$$

Finally, if $x, y \in [a, b]$ and $x = y$, then we trivially have $|G(x) - G(y)| = 0 \leq d \cdot |x - y|$. Therefore, for any $x, y \in [a, b]$, we have $|G(x) - G(y)| \leq d \cdot |x - y|$.

We use this fact to prove that G is uniformly continuous on $[a, b]$. Let $\varepsilon > 0$. Consider $\delta = \frac{\varepsilon}{d} > 0$. Let $x, y \in [a, b]$ be arbitrary with $|x - y| < \delta$. We then have

$$\begin{aligned} |G(x) - G(y)| &\leq d \cdot |x - y| && \text{(from above)} \\ &< d \cdot \frac{\varepsilon}{d} \\ &= \varepsilon. \end{aligned}$$

Therefore, G is uniformly continuous on $[a, b]$, and hence continuous on $[a, b]$.

Suppose now that f is continuous at a point c in (a, b). To show that G is differentiable at c and that $G'(c) = f(c)$, we want to show that

$$\lim_{x \to c} \frac{G(x) - G(c)}{x - c} = f(c).$$

Let $\varepsilon > 0$. Since f is continuous at c, we can fix $\delta > 0$ such that for all $t \in [a, b]$ with $|t - c| < \delta$, we have $|f(t) - f(c)| < \frac{\varepsilon}{2}$. We may assume, by making δ smaller if necessary, that $\delta \leq c - a$ and $\delta \leq b - c$ (so that if $|t - c| < \delta$, then $t \in (a, b)$). We claim that for all x with $0 < |x - c| < \delta$, we have

$$\left| \frac{G(x) - G(c)}{x - c} - f(c) \right| < \varepsilon.$$

So let x be arbitrary with $0 < |x - c| < \delta$. We have two cases:

- *Case 1:* Suppose that $c < x$. Using Proposition 6.2.14 several times, we have

$$\begin{aligned} \left| \frac{G(x) - G(c)}{x - c} - f(c) \right| &= \left| \frac{1}{x - c} \cdot \left(\int_a^x f - \int_a^c f \right) - f(c) \right| \\ &= \left| \frac{1}{x - c} \cdot \left(\int_a^c f + \int_c^x f - \int_a^c f \right) - f(c) \right| \\ &= \left| \left(\frac{1}{x - c} \cdot \int_c^x f \right) - f(c) \right| \\ &= \left| \left(\frac{1}{x - c} \cdot \int_c^x f(t)\ dt \right) - \left(\frac{1}{x - c} \cdot \int_c^x f(c)\ dt \right) \right| \\ &= \left| \left(\frac{1}{x - c} \cdot \int_c^x (f(t) - f(c))\ dt \right) \right| \\ &\leq \frac{1}{x - c} \cdot \int_c^x |f(t) - f(c)|\ dt \\ &\leq \frac{1}{x - c} \cdot \frac{\varepsilon}{2} \cdot (x - c) \\ &= \frac{\varepsilon}{2} \\ &< \varepsilon. \end{aligned}$$

- *Case 2:* Suppose that $x < c$. Using Proposition 6.2.14 several times, we have

$$\begin{aligned}
\left|\frac{G(x)-G(c)}{x-c} - f(c)\right| &= \left|\frac{1}{x-c}\cdot\left(\int_a^x f - \int_a^c f\right) - f(c)\right| \\
&= \left|\frac{1}{x-c}\cdot\left(\int_a^x f - \left(\int_a^x f + \int_x^c f\right)\right) - f(c)\right| \\
&= \left|\left(\frac{1}{x-c}\cdot\left(-\int_x^c f\right)\right) - f(c)\right| \\
&= \left|\left(\frac{1}{c-x}\cdot\int_x^c f\right) - f(c)\right| \\
&= \left|\left(\frac{1}{c-x}\cdot\int_x^c f(t)\ dt\right) - \left(\frac{1}{c-x}\cdot\int_x^c f(c)\ dt\right)\right| \\
&= \left|\left(\frac{1}{c-x}\cdot\int_x^c (f(t)-f(c))\ dt\right)\right| \\
&\leq \frac{1}{c-x}\cdot\int_x^c |f(t)-f(c)|\ dt \\
&\leq \frac{1}{c-x}\cdot\frac{\varepsilon}{2}\cdot(c-x) \\
&= \frac{\varepsilon}{2} \\
&< \varepsilon.
\end{aligned}$$

□

Chapter 7

Sequences and Series of Functions

We've given a lot of attention to sequences and series of (real) numbers, but there's no reason to confine ourselves to these humble situations. We could of course make sequences of any mathematical object: triangles, complex numbers, functions, etc. However, if we want to do anything of interest with these new types of sequences, we need some way of telling when a sequence of these objects approach another so that we can define a notion of convergence. Also, if we want to define series of these objects, we need a way to "add" them. Fortunately, we have ways of doing these for functions. Unfortunately, there is more than one way to give a coherent meaning to "the sequence of functions $\langle f_n \rangle$ approaches the function h".

7.1 Pointwise Convergence

Here is the precise definition of a sequence of functions.

Definition 7.1.1. *Let $A \subseteq \mathbb{R}$ and let $\mathcal{F}$ be the set of all functions $g\colon A \to \mathbb{R}$. A* sequence of functions on A *is a function $f\colon \mathbb{N}^+ \to \mathcal{F}$.*

As in the case of sequences of real numbers, we do not usually think of a sequence of functions as a function that outputs functions. Instead, we think of a sequence of functions on A as an infinite list of functions

$$f_1 \quad f_2 \quad f_3 \quad f_4 \quad \cdots \tag{7.1}$$

where $f_n\colon A \to \mathbb{R}$ for every $n \in \mathbb{N}^+$, and we use the notation $\langle f_n \rangle$ to denote this sequence.

Suppose that $\langle f_n \rangle$ is a sequence of functions on A, and h is a function on A. We need to somehow make sense of the statement that "$\langle f_n \rangle$ approaches h". Well, a function on A is determined by where it sends each point in A, and so perhaps we should define "$\langle f_n \rangle$ approaches h" to simply mean that for every $x \in A$, the sequence $\langle f_n(x) \rangle$ of real numbers converges to the real number $h(x)$, i.e. that for every $x \in A$, we have $\lim\limits_{n\to\infty} f_n(x) = h(x)$. Let's codify this as a definition where we explicitly reference that we're thinking of convergence on a point-by-point basis.

Definition 7.1.2. *Let $A \subseteq \mathbb{R}$, let $\langle f_n \rangle$ be a sequence of functions on A and let h be a function on A. We say that $\langle f_n \rangle$* converges pointwise to h on A *if for all $x \in A$, the sequence $\langle f_n(x) \rangle$ of real numbers converges to $h(x)$. Unwrapping this definition, the sequence $\langle f_n \rangle$ converges pointwise to h means that for all $x \in A$ and all $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$ such that for all $n \geq N$, we have $|f_n(x) - h(x)| < \varepsilon$.*

Definition 7.1.3. *Let $A \subseteq \mathbb{R}$ and let $\langle f_n \rangle$ be a sequence of functions on A. We say that $\langle f_n \rangle$* converges pointwise on A *if there exists a function h on A such that $\langle f_n \rangle$ converges pointwise to h on A.*

Let consider a few examples illustrating pointwise convergence.

- For each $n \in \mathbb{N}^+$, let $f_n \colon \mathbb{R} \to \mathbb{R}$ be the function given by $f_n(x) = \frac{x}{n}$. Let $h \colon \mathbb{R} \to \mathbb{R}$ be the zero function, i.e. $h(x) = 0$ for all $x \in \mathbb{R}$. We claim that $\langle f_n \rangle$ converges pointwise to h on $\mathbb{R}$. To see this, let $x \in \mathbb{R}$ be arbitrary. Since $\lim\limits_{n\to\infty} \frac{1}{n} = 0$, it follows that

$$\begin{aligned} \lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} \frac{x}{n} \\ &= x \cdot 0 && \text{(by Theorem 2.2.8)} \\ &= 0 \\ &= h(x). \end{aligned}$$

- For each $n \in \mathbb{N}^+$, let $f_n \colon [0,1] \to \mathbb{R}$ be the function given by $f_n(x) = x^n$. Let $h \colon [0,1] \to \mathbb{R}$ be the function

$$h(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{otherwise} \end{cases}$$

  We claim $\langle f_n \rangle$ converges pointwise to h on $[0,1]$. Let $x \in [0,1]$ be arbitrary. If $x = 1$, then

$$\begin{aligned} \lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} x^n \\ &= \lim_{n\to\infty} 1^n \\ &= 1 \\ &= h(x). \end{aligned}$$

  On the other hand, if $0 \leq x < 1$ then

$$\begin{aligned} \lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} x^n \\ &= 0 && \text{(by Proposition 2.3.11)} \\ &= h(x). \end{aligned}$$

- For each $n \in \mathbb{N}^+$, let $f_n \colon \mathbb{R} \to \mathbb{R}$ be the function given by $f_n(x) = \frac{1}{n}\sin(2\pi n^2 x)$. Let $h \colon \mathbb{R} \to \mathbb{R}$ be the zero function, i.e. $h(x) = 0$ for all $x \in \mathbb{R}$. We claim $\langle f_n \rangle$ converges pointwise to h on $\mathbb{R}$. To see this, let $x \in \mathbb{R}$ be arbitrary. Since $-1 \leq \sin(2\pi n^2 x) \leq 1$ for all $n \in \mathbb{N}^+$, we have $-\frac{1}{n} \leq \frac{1}{n}\sin(2\pi n^2 x) \leq \frac{1}{n}$ for all $n \in \mathbb{N}^+$, and hence $-\frac{1}{n} \leq f_n(x) \leq \frac{1}{n}$ for all $n \in \mathbb{N}^+$. Using the Squeeze Theorem and the fact that $\lim\limits_{n\to\infty} (-\frac{1}{n}) = 0 = \lim\limits_{n\to\infty} \frac{1}{n}$, we conclude that $\lim\limits_{n\to\infty} f_n(x) = 0 = h(x)$.

We saw in the second example above that it is possible for a sequence of continuous functions to converge pointwise to a discontinuous function. We give another example of this phenomenon since it is probably the biggest failing of the notion of pointwise convergence. Before jumping into the example, we state a simple result about piecewise functions.

Proposition 7.1.4. *Let A be an interval, let $b_1, b_2, \ldots, b_n \in A$, and let $g_0, g_1, \ldots, g_n \colon A \to \mathbb{R}$ be functions. Define $f \colon A \to \mathbb{R}$ by letting*

$$f(a) = \begin{cases} g_0(a) & \text{if } a < b_1 \\ g_1(a) & \text{if } b_1 \leq a < b_2 \\ \quad\vdots & \\ g_{n-1}(a) & \text{if } b_{n-1} \leq a < b_n \\ g_n(a) & \text{if } b_n \leq a. \end{cases}$$

1. *Suppose that each g_i is continuous on A and that $g_{i-1}(b_i) = g_i(b_i)$ whenever $1 \leq i \leq n$. We then have that f is continuous on A.*

2. *Suppose that each g_i is differentiable on A and that both $g_{i-1}(b_i) = g_i(b_i)$ and $g'_{i-1}(b_i) = g'_i(b_i)$ whenever $1 \leq i \leq n$. We then have that f is differentiable on A.*

Proof. Exercise. □

Now define a sequence of functions $\langle f_n \rangle$ on $\mathbb{R}$ by letting

$$f_n(x) = \begin{cases} -1 & \text{if } x < -\frac{1}{n} \\ nx & \text{if } -\frac{1}{n} \leq x < \frac{1}{n} \\ 1 & \text{if } \frac{1}{n} \leq x \end{cases}$$

for every $n \in \mathbb{N}^+$ and define h on $\mathbb{R}$ by

$$h(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

We claim that $\langle f_n \rangle$ is a sequence of continuous functions that converges pointwise to h on $\mathbb{R}$. To see that each f_n is continuous, let $n \in \mathbb{N}^+$ be arbitrary. Define $g_0(x) = -1$, $g_1(x) = nx$, and $g_2(x) = 1$. Each of g_0, g_1, and g_2 is continuous on $\mathbb{R}$, and we have $g_0(-\frac{1}{n}) = -1 = g_1(-\frac{1}{n})$ and also $g_1(\frac{1}{n}) = 1 = g_2(\frac{1}{n})$. Therefore, f_n is continuous.

We now show that $\langle f_n \rangle$ converges pointwise to h on $\mathbb{R}$. Notice that for $x = 0$, we have $f_n(0) = n \cdot 0 = 0$ for all $n \in \mathbb{N}^+$, hence $\lim\limits_{n\to\infty} f_n(0) = 0 = h(0)$. Suppose that $x > 0$. Fix $N \in \mathbb{N}^+$ with $\frac{1}{N} < x$. We then have $f_n(x) = 1$ for all $n \geq N$, hence $\lim\limits_{n\to\infty} f_n(x) = 1 = h(x)$. Suppose that $x < 0$. Fix $N \in \mathbb{N}^+$ with $x < -\frac{1}{N}$. We then have $f_n(x) = -1$ for all $n \geq N$, hence $\lim\limits_{n\to\infty} f_n(x) = -1 = h(x)$. Thus, in all cases, we have $\lim\limits_{n\to\infty} f_n(x) = h(x)$.

It is also possible to "round off the corners" in the previous example to get a sequence of differentiable functions which converges pointwise to a function which is not continuous. Define a sequence of functions $\langle f_n \rangle$ on $\mathbb{R}$ by letting

$$f_n(x) = \begin{cases} -1 & \text{if } x < -\frac{1}{n} \\ \sin(\frac{n\pi x}{2}) & \text{if } -\frac{1}{n} \leq x < \frac{1}{n} \\ 1 & \text{if } x \geq \frac{1}{n} \end{cases}$$

for every $n \in \mathbb{N}^+$ and define h on $\mathbb{R}$ by

$$h(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Using the above proposition, it is straightforward to see that each f_n is differentiable on $\mathbb{R}$. The proof that $\langle f_n \rangle$ converges pointwise to h on $\mathbb{R}$ is the same as above except for at the point 0 where we have $f_n(0) = \sin(\frac{n\pi \cdot 0}{2}) = \sin 0 = 0$ for all $n \in \mathbb{N}^+$, hence $\lim\limits_{n\to\infty} f_n(0) = \lim\limits_{n\to\infty} 0 = 0$.

One thing that you might suspect is that if $\langle f_n \rangle$ converges pointwise to h on $[a, b]$ and each f_n is integrable on $[a, b]$, then h is integrable on $[a, b]$, and

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b h.$$

In fact, we can't even ensure that h is integrable on $[a, b]$, even if it is bounded. To see this, consider the following intriguing example. Since $\mathbb{Q} \cap [0, 1]$ is countable, we can fix a listing

$$q_1 \quad q_2 \quad q_3 \quad q_4 \quad \dots \tag{7.2}$$

of the rational numbers in $[0, 1]$. For each $n \in \mathbb{N}^+$, let $f_n \colon [0, 1] \to \mathbb{R}$ be the function

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{q_1, q_2, \dots, q_n\} \\ 0 & \text{otherwise} \end{cases}$$

Let $h \colon [0, 1] \to \mathbb{R}$ be the function

$$h(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{otherwise} \end{cases}$$

We know that h is not integrable on $[0, 1]$, and for every $n \in \mathbb{N}$ we have that f_n is integrable on $[0, 1]$ because it is 0 at all but finitely many points (and hence has only finitely many discontinuities). We claim $\langle f_n \rangle$ converges pointwise to h on $[0, 1]$. Fix $x \in [0, 1]$. For any $x \in \mathbb{R} \backslash \mathbb{Q}$, we have $f_n(x) = 0$ for all $n \in \mathbb{N}$, hence $\lim\limits_{n\to\infty} f_n(x) = 0 = h(x)$. Suppose then that $x \in \mathbb{Q}$. Fix $k \in \mathbb{N}$ with $x = q_k$. For every $n \geq k$, we have $f_n(x) = 1$, hence $\lim\limits_{n\to\infty} f_n(x) = h(x)$. Therefore, $\langle f_n \rangle$ converges pointwise to h on $[0, 1]$.

All right, if $\langle f_n \rangle$ converges pointwise to h on $[a, b]$, and h happens to be integrable on $[a, b]$, then surely we have

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b h,$$

right? No. The basic idea is that we can put taller and taller, yet narrower and narrower spikes, into a function so that at every point we eventually settle down to 0 but every function in the sequence has a large spike which gives an integral 1. Here are the details. Define a sequence of functions $\langle f_n \rangle$ on $[0, 1]$ by letting

$$f_n(x) = \begin{cases} 4n^2 x & \text{if } 0 \leq x < \frac{1}{2n} \\ 4n - 4n^2 x & \text{if } \frac{1}{2n} \leq x < \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \leq x \leq 1 \end{cases}$$

for every $n \in \mathbb{N}$ and let $h \colon [0, 1] \to \mathbb{R}$ be the zero function. Using the above result, it is straightforward to check that each f_n is continuous. Of course, h is continuous as well. We next claim that $\langle f_n \rangle$ converges pointwise to h. First notice that $\lim\limits_{n\to\infty} f_n(0) = \lim\limits_{n\to\infty} (4n^2 \cdot 0) = 0 = h(0)$. Suppose now that $0 < x \leq 1$. Fix $N \in \mathbb{N}^+$ with $\frac{1}{N} < x$. For every $n \geq N$, we then have $\frac{1}{n} \leq \frac{1}{N} < x$, hence $f_n(x) = 0$. It follows that $\lim\limits_{n\to\infty} f_n(x) = 0 = h(x)$.

We next look at the integrals. Given any $n \in \mathbb{N}^+$, we have

$$\begin{aligned}
\int_0^1 f_n &= \int_0^{\frac{1}{2n}} f_n + \int_{\frac{1}{2n}}^{\frac{1}{n}} f_n + \int_{\frac{1}{n}}^1 f_n \\
&= \int_0^{\frac{1}{2n}} 4n^2 x \; dx + \int_{\frac{1}{2n}}^{\frac{1}{n}} (4n - 4n^2 x) \; dx + \int_{\frac{1}{n}}^1 0 \; dx \\
&= 2n^2 x^2 \Big|_0^{\frac{1}{2n}} + (4nx - 2n^2 x^2) \Big|_{\frac{1}{2n}}^{\frac{1}{n}} + 0 \\
&= 2n^2 \cdot \left(\frac{1}{2n}\right)^2 - 0 + \left(4n \cdot \frac{1}{n} - 2n^2 \cdot \left(\frac{1}{n}\right)^2\right) - \left(4n \cdot \frac{1}{2n} - 2n^2 \cdot \left(\frac{1}{2n}\right)^2\right) \\
&= \frac{1}{2} + 4 - 2 - 2 + \frac{1}{2} \\
&= 1.
\end{aligned}$$

However, we clearly have $\int_0^1 h = 0$. Therefore, we have $\lim\limits_{n\to\infty} \int_0^1 f_n = 1 \neq \int_0^1 h$.

Again, if you want differentiability of the f_n's, you can get it by rounding off the corners. Define a sequence of functions $\langle f_n \rangle$ on $[0, 1]$ by letting

$$f_n(x) = \begin{cases} n(1 - \cos(2\pi n x)) & \text{if } 0 \leq x < \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \leq x \leq 1 \end{cases}$$

and let $h\colon [0, 1] \to \mathbb{R}$ be the zero function. A similar argument to the above shows that each f_n is differentiable, that $\langle f_n \rangle$ converges pointwise to h, and that $\lim\limits_{n\to\infty} \int_0^1 f_n = 1 \neq \int_0^1 h$.

7.2 Uniform Convergence

The examples above illustrating the problems with pointwise convergence exploited the point-by-point basis of the definition in such a way that every individual point eventually behaved, but in a global sense the functions in the sequence were getting worse and worse (having very sharp spikes at points that change with n, for example). As a result, we'd like to come up with a new definition for "the sequence $\langle f_n \rangle$ converges to h" which takes a more global approach. The idea is that for every $\varepsilon > 0$, if we put a little ε window around the whole function h, then f_n will lie entirely in that window for sufficiently large n.

Definition 7.2.1. *Let $A \subseteq \mathbb{R}$, let $\langle f_n \rangle$ be a sequence of functions on A and let h be a function on A. We say that $\langle f_n \rangle$* converges uniformly to h on A *if:*

For all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $x \in A$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \varepsilon$.

Definition 7.2.2. *Let $A \subseteq \mathbb{R}$ and let $\langle f_n \rangle$ be a sequence of functions on A. We say that $\langle f_n \rangle$* converges uniformly on A *if there exists a function h on A such that $\langle f_n \rangle$ converges uniformly to h on A.*

Notice the crucial distinction between the definitions of pointwise convergence and uniform convergence. Here they are side-by-side for you to examine. The first one is "$\langle f_n \rangle$ converges pointwise to h on A" and the second is "$\langle f_n \rangle$ converges uniformly to h on A"

1. For all $x \in A$ and all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, we have $|f_n(x) - h(x)| < \varepsilon$.
2. For all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $x \in A$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \varepsilon$.

Here's the key distinction in more words than symbols. In pointwise convergence, we have a fixed $\varepsilon > 0$ (the error region) *and also* a fixed x (the point in mind) in hand when we have to choose the N (i.e. how far out to go) beyond which we're in the clear for this particular fixed ε and x. On the other hand, in uniform convergence we *only* have the fixed $\varepsilon > 0$ and we need to make our choice of N beyond which we're in the clear for this particular fixed ε and *for all* $x \in A$ simultaneously. Thus, if we're able to succeed for $\varepsilon = \frac{1}{2}$ for each $x \in A$, but different x's require us to choose larger and larger N's without bound (say one x requires that N be at least 100, another at least 1000, etc.), then we would get pointwise convergence but not uniform convergence.

We obtain the following immediately from the definitions.

Proposition 7.2.3. *If $\langle f_n \rangle$ converges uniformly to h on A, then $\langle f_n \rangle$ converges pointwise to h on A.*

There is another, slightly more abstract, way to define uniform convergence. The idea is to assign a number to any bounded function f that somehow represents how "far away" f is from the zero function. We will use the notation $||f||$ to denote this number to remind us of the absolute value. The goal is to use this number in the way we use absolute values since when x is a number, $|x|$ is a number which represents how "far away" x is from 0. Using the absolute value, we then are able to say how far away x is from y by looking at $|x - y|$. Thus, if we succeed in giving a good definition of $||f||$ for every bounded function f, we can use it to assign a number which represents how "far away" f is from g by examining $||f - g||$. There are many ways to intelligently assign such a number $||f||$, but the following definition is the one that jives with uniform convergence.

Definition 7.2.4. *Let $A \subseteq \mathbb{R}$ and let f be a bounded function on A. We define $||f|| = \sup\{|f(x)| : x \in A\}$ and call $||f||$ the* norm *of f on A.*

Now if $\varepsilon > 0$, f and g are two function on a set A, and $f - g$ is bounded on A, then saying that $||f - g|| < \varepsilon$ is equivalent to saying that g lies entirely in a ε-window around the function f. Thus, the following proposition shouldn't be too surprising.

Proposition 7.2.5. *Let $A \subseteq \mathbb{R}$, let $\langle f_n \rangle$ be a sequence of functions on A, and let h be a function on A. The following are equivalent:*

1. *$\langle f_n \rangle$ converges uniformly to h on A.*
2. *The sequence of functions $\langle f_n - h \rangle$ is eventually bounded, and $\lim_{n\to\infty} ||f_n - h|| = 0$.*

Proof. • (1) $\to$ (2): Suppose that $\langle f_n \rangle$ converges uniformly to h on A. We first show that $\langle f_n - h \rangle$ is eventually bounded. Since $\langle f_n \rangle$ converges uniformly to h on A, we can fix $N \in \mathbb{N}^+$ such that for all $x \in A$ and all $n \geq N$, we have $|f_n(x) - h(x)| < 1$. For any $n \geq N$, we then have that $|(f_n - h)(x)| = |f_n(x) - h(x)| < 1$ for all $x \in A$. Thus, for any $n \geq N$, the function $f_n - h$ is bounded.

We now show that $\lim_{n\to\infty} ||f_n - h|| = 0$. Let $\varepsilon > 0$. Since $\langle f_n \rangle$ converges uniformly to h on A, we can fix $N \in \mathbb{N}^+$ such that for all $x \in A$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \frac{\varepsilon}{2}$. Now let $n \geq N$ be arbitrary. We then have that the set $\{|f_n(x) - h(x)| : x \in A\}$ is bounded above by $\frac{\varepsilon}{2}$, hence

$$\begin{aligned} ||f_n - h|| &= \sup\{|f_n(x) - h(x)| : x \in A\} \\ &\leq \frac{\varepsilon}{2} \\ &< \varepsilon. \end{aligned}$$

It follows that $|\ ||f_n - h|| - 0| = ||f_n - h|| < \varepsilon$. Therefore, $\lim_{n\to\infty} ||f_n - h|| = 0$.

- (2) → (1): Suppose that the sequence of functions $\langle f_n - h \rangle$ is eventually bounded, and that $\lim_{n\to\infty} ||f_n - h|| = 0$. We show that $\langle f_n \rangle$ converges uniformly to h on A. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} ||f_n - h|| = 0$, we can fix $N \in \mathbb{N}$ such that for all $n \geq N$, we have $|\, ||f_n - h|| - 0| < \varepsilon$. Now let $x \in A$ and $n \geq N$ be arbitrary. We then have

$$\begin{aligned} |f_n(x) - h(x)| &\leq ||f_n - h|| \\ &= |\, ||f_n - h|| - 0| \\ &< \varepsilon. \end{aligned}$$

  Therefore, $\langle f_n \rangle$ converges uniformly to h on A.

□

This proposition is helpful in that we can often break up the complicated definition of uniform convergence into two steps. First, somehow calculate or at least estimate the value $||f_n - h||$. Second, take the resulting sequence of numbers $\langle ||f_n - h|| \rangle$ and use all of our tools about sequences to determine whether it converges to 0.

For example, let $\langle f_n \rangle$ be the sequence of functions on $\mathbb{R}$ defined by letting $f_n(x) = \frac{x}{n}$, and let $h\colon \mathbb{R} \to \mathbb{R}$ be the zero function, i.e. $h(x) = 0$ for all $x \in \mathbb{R}$. We saw above that $\langle f_n \rangle$ converges pointwise to h. We claim that $\langle f_n \rangle$ does *not* converge uniformly to h on $\mathbb{R}$. We give two arguments:

- Negating the definition of uniform convergence, we need to exhibit an $\varepsilon > 0$ such that for all $N \in \mathbb{N}^+$, there exists $x \in \mathbb{R}$ and $n \geq N$ such that $|f(x) - h(x)| \geq \varepsilon$. Consider $\varepsilon = 1$. Let $N \in \mathbb{N}^+$ be arbitrary. Notice that

$$\begin{aligned} |f_n(x) - h(x)| &= |f_N(N) - h(N)| \\ &= \left| \frac{N}{N} - 0 \right| \\ &= 1 \\ &\geq \varepsilon. \end{aligned}$$

  Therefore, $\langle f_n \rangle$ does not converges uniformly to h on $\mathbb{R}$.

- We use Proposition 7.2.5. Let $n \in \mathbb{N}^+$ be arbitrary. We have that $f_n - h = f_n$ is not bounded on $\mathbb{R}$, because for any $d > 0$, we have $|f_n(dn + 1)| = d + \frac{1}{n} > d$.

Let's consider another example. For each $n \in \mathbb{N}$, let $f_n\colon \mathbb{R} \to \mathbb{R}$ be the function given by $f_n(x) = \frac{1}{n}\sin(2\pi n^2 x)$. Let $h\colon \mathbb{R} \to \mathbb{R}$ be the zero function, i.e. $h(x) = 0$ for all $x \in \mathbb{R}$. We claim that $\langle f_n \rangle$ converges uniformly to h on $\mathbb{R}$. Again, we give two arguments:

- We verify the definition. Let $\varepsilon > 0$. Fix $N \in \mathbb{N}^+$ with $\frac{1}{N} < \varepsilon$. Let $x \in \mathbb{R}$ and $n \geq N$ be arbitrary. We

have

$$\begin{aligned} |f_n(x) - h(x)| &= \left| \frac{1}{n} \sin(2\pi n^2 x) - 0 \right| \\ &= \left| \frac{1}{n} \sin(2\pi n^2 x) \right| \\ &= \frac{1}{n} \cdot |\sin(2\pi n^2 x)| \\ &\leq \frac{1}{n} \\ &\leq \frac{1}{N} && \text{(since } n \geq N\text{)} \\ &< \varepsilon. \end{aligned}$$

It follows that $\langle f_n \rangle$ converges uniformly to h on $\mathbb{R}$.

- We use Proposition 7.2.5. Let $n \in \mathbb{N}^+$ be arbitrary. For any $x \in \mathbb{R}$, we have

$$\begin{aligned} |f_n(x) - h(x)| &= |\frac{1}{n} \sin(2\pi n^2 x) - 0| \\ &= \frac{1}{n} \cdot |\sin(2\pi n^2 x)| \\ &\leq \frac{1}{n}, \end{aligned}$$

hence $||f_n - h|| \leq \frac{1}{n}$ (and in fact $||f_n - h|| = \frac{1}{n}$, as you can easily verify). Since $0 \leq ||f_n - h|| \leq \frac{1}{n}$ for all $n \in \mathbb{N}$ and $\lim\limits_{n\to\infty} \frac{1}{n} = 0$, it follows from the Squeeze Theorem that $\lim\limits_{n\to\infty} ||f_n - h|| = 0$. Therefore, $\langle f_n \rangle$ converges uniformly to h on $\mathbb{R}$.

Now consider the following sequence of functions. For each $n \in \mathbb{N}^+$, let $f_n \colon [0,1] \to \mathbb{R}$ be the function given by $f_n(x) = x^n$. Let $h \colon [0,1] \to \mathbb{R}$ be the function

$$h(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

We claim that $\langle f_n \rangle$ does not converge uniformly to h on $[0,1]$. Again, we give two arguments:

- We prove the negation of the definition. Consider $\varepsilon = \frac{1}{2}$. Let $N \in \mathbb{N}^+$ be arbitrary. Now f_N is continuous on $[0,1]$ and is such that $f_N(0) = 0$ and $f_N(1) = 1$, so by the Intermediate Value Theorem, we can fix $y \in (0,1)$ with $f_N(y) = \frac{1}{2}$ (in fact, we have $y = \sqrt[N]{1/2}$). We now have

$$\begin{aligned} |f_N(y) - h(y)| &= \left| \frac{1}{2} - 0 \right| \\ &= \frac{1}{2} \\ &\geq \varepsilon. \end{aligned}$$

- We use Proposition 7.2.5. Let $n \in \mathbb{N}^+$ be arbitrary. We seek to calculate (or at least estimate) $||f_n - h||$. Notice that $\{x^n : x \in [0,1)\} \subseteq [0,1)$. Moreover, since $f_n(x) = x^n$ is continuous with $f_n(0) = 0$ and

$f_n(1) = 1$, we know from the Intermediate Value Theorem that for all $y \in [0,1)$, there exists $x \in [0,1)$ with $x^n = y$. Thus, $\{x^n : x \in [0,1)\} = [0,1)$. Therefore, we have

$$\begin{aligned} ||f_n - h|| &= \sup\{|f_n(x) - h(x)| : 0 \leq x \leq 1\} \\ &= \sup(\{0\} \cup \{|f_n(x)| : 0 \leq x < 1)\} \\ &= \sup(\{0\} \cup \{x^n : 0 \leq x < 1)\} \\ &= \sup\{[0,1)\} \\ &= 1. \end{aligned}$$

We have show that $||f_n - h|| = 1$ for all $n \in \mathbb{N}^+$, and hence

$$\begin{aligned} \lim_{n\to\infty} ||f_n - h|| &= \lim_{n\to\infty} 1 \\ &= 1 \\ &\neq 0, \end{aligned}$$

so $\langle f_n \rangle$ does not converge uniformly to h on $[0,1]$.

We consider one more example of how to compute $||f_n - h||$. For each $n \in \mathbb{N}^+$, let $f_n \colon [0,1] \to \mathbb{R}$ be the function given by $f_n(x) = x^n(1-x^n)$. Let $h \colon [0,1] \to \mathbb{R}$ be the zero function. We first show that $\langle f_n \rangle$ converges pointwise to h on $[0,1]$. Notice that $f_n(1) = 0$ for all $n \in \mathbb{N}^+$, so clearly $\lim\limits_{n\to\infty} f_n(1) = 0 = h(1)$. Let $x \in \mathbb{R}$ be arbitrary with $0 \leq x < 1$. Using Proposition 2.3.11, we know that $\lim\limits_{n\to\infty} x^n = 0$. By Theorem 2.2.8, it follows that $\lim\limits_{n\to\infty} (1 - x^n) = 1 - 0 = 1$, and hence

$$\begin{aligned} \lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} x^n(1-x^n) \\ &= 0 \cdot 1 \\ &= 0 \\ &= h(x). \end{aligned}$$

Therefore, $\langle f_n \rangle$ converges pointwise to h on $[0,1]$.

In order to determine whether $\langle f_n \rangle$ converges uniformly to h on $[0,1]$, we want to calculate (or at least estimate) $||f_n - h|| = ||f_n||$. Let $n \in \mathbb{N}^+$ be arbitrary. Notice that f_n is continuous on $[0,1]$ and differentiable on $(0,1)$ by Corollary 5.2.7 and Corollary 6.1.8. Since f_n is continuous on the compact set $[0,1]$, we know from the Extreme Value Theorem that f_n achieves a maximum and a minimum on $[0,1]$. By the discussion after Proposition 6.1.10, these extreme values must occur either at an endpoints, or at a point $x \in (0,1)$ with $f_n'(x) = 0$. Since $f_n(x) = x^n - x^{2n}$ for every $x \in [0,1]$, we have

$$\begin{aligned} f_n'(x) &= nx^{n-1} - 2nx^{2n-1} \\ &= nx^{n-1}(1 - 2x^n) \end{aligned}$$

for every $x \in (0,1)$. Therefore, if $x \in (0,1)$, then $f_n'(x) = 0$ if and only if $x = \sqrt[n]{1/2}$. Plugging in these critical values and the endpoints, we see that $f_n(0) = 0$, $f_n(1) = 0$, and $f_n(\sqrt[n]{1/2}) = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4}$. Thus, $0 \leq f_n(x) \leq \frac{1}{4}$ for every $x \in [0,1]$ and $f_n(\sqrt[n]{1/2}) = \frac{1}{4}$, so $||f_n - h|| = \sup\{|f_n(x)| : x \in [0,1]\} = \frac{1}{4}$. Since $\lim\limits_{n\to\infty} ||f_n - h|| = \lim\limits_{n\to\infty} \frac{1}{4} = \frac{1}{4} \neq 0$, it follows that $\langle f_n \rangle$ does not converge uniformly to h on $[0,1]$.

It's time to move from examples to theoretical results about uniform convergence. Fortunately, uniform convergence is well-suited to analysis. We first show that if $\langle f_n \rangle$ is a sequence continuous functions that converges uniformly to h, then h is also continuous. The proof is an $\frac{\varepsilon}{3}$ argument whereby you drop from h down to a suitable f_n, move over across f_n which you know is continuous, and move back up to h. It's quite cute.

Theorem 7.2.6. *Let A be an interval. Suppose that $\langle f_n \rangle$ is a sequence of continuous functions on A and that $\langle f_n \rangle$ converges uniformly to h on A. We then have that h is continuous on A.*

Proof. Let $c \in A$ be arbitrary. We show that h is continuous at c. Let $\varepsilon > 0$. Since $\langle f_n \rangle$ converges to h uniformly on A, we can fix an $N \in \mathbb{N}^+$ such that for all $x \in A$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \frac{\varepsilon}{3}$. Since f_N is continuous at c, we can fix $\delta > 0$ such that for all $x \in A$ with $|x - c| < \delta$, we have $|f_N(x) - f_N(c)| < \frac{\varepsilon}{3}$. Now let $x \in A$ be arbitrary with $|x - c| < \delta$. We then have

$$\begin{aligned}
|h(x) - h(c)| &= |h(x) - f_N(x) + f_N(x) - f_N(c) + f_N(c) - h(c)| \\
&\leq |h(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - h(c)| && \text{(by the Triangle Inequality)} \\
&\leq |f_N(x) - h(x)| + |f_N(x) - f_N(c)| + |f_N(c) - h(c)| \\
&< \frac{\varepsilon}{3} + |f_N(x) - f_N(c)| + \frac{\varepsilon}{3} && \text{(since } x, c \in A \text{ and } N \geq N\text{)} \\
&< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} && \text{(since } |x - c| < \delta\text{)} \\
&= \varepsilon.
\end{aligned}$$

Therefore, h is continuous at c. Since $c \in A$ was arbitrary, it follows that h is continuous on A. □

Also, all is right in the world of integration.

Theorem 7.2.7. *Suppose that $\langle f_n \rangle$ is a sequence of functions which are integrable on $[a, b]$ and that $\langle f_n \rangle$ converges uniformly to h on $[a, b]$. We then have that h is integrable on $[a, b]$ and that*

$$\int_a^b h = \lim_{n \to \infty} \int_a^b f_n$$

Proof. We first show that h is integrable on $[a, b]$. Let $\varepsilon > 0$. Since $\langle f_n \rangle$ converges uniformly to h on $[a, b]$, we can fix $N \in \mathbb{N}^+$ such that for all $x \in [a, b]$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \frac{\varepsilon}{3(b-a)}$. Since f_N is integrable on $[a, b]$, we can use Proposition 6.2.8 to fix a partition $P = (p_0, p_1, \ldots, p_n)$ of $[a, b]$ such that $U(f_N, P) - L(f_N, P) < \frac{\varepsilon}{3}$. For each i with $1 \leq i \leq n$, let

$$\begin{aligned}
m_i &= \inf\{h(x) : p_{i-1} \leq x \leq p_i\} \\
m_i' &= \inf\{f_N(x) : p_{i-1} \leq x \leq p_i\} \\
M_i &= \sup\{h(x) : p_{i-1} \leq x \leq p_i\} \\
M_i' &= \sup\{f_N(x) : p_{i-1} \leq x \leq p_i\}.
\end{aligned}$$

Since $|f_n(x) - h(x)| < \frac{\varepsilon}{3(b-a)}$ for all $x \in [a, b]$, we have

$$f_N(x) - \frac{\varepsilon}{3(b-a)} < h(x) < f_N(x) + \frac{\varepsilon}{3(b-a)}$$

for all $x \in [a,b]$, and thus both $M_i \leq M_i' + \frac{\varepsilon}{3(b-a)}$ and $m_i \geq m_i' - \frac{\varepsilon}{3(b-a)}$ for all i. Therefore,

$$\begin{aligned}
U(h,P) - L(h,P) &= \sum_{i=1}^{n} M_i \cdot (p_i - p_{i-1}) - \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) \\
&\leq \sum_{i=1}^{n} \left(M_i' + \frac{\varepsilon}{3(b-a)} \right) \cdot (p_i - p_{i-1}) - \sum_{i=1}^{n} \left(m_i - \frac{\varepsilon}{3(b-a)} \right) \cdot (p_i - p_{i-1}) \\
&= \sum_{i=1}^{n} M_i' \cdot (p_i - p_{i-1}) - \sum_{i=1}^{n} m_i \cdot (p_i - p_{i-1}) + 2 \cdot \sum_{i=1}^{n} \frac{\varepsilon}{3(b-a)} \cdot (p_i - p_{i-1}) \\
&= U(f_N, P) - L(f_N, P) + \frac{2\varepsilon}{3(b-a)} \sum_{i=1}^{n} (p_i - p_{i-1}) \\
&= U(f_N, P) - L(f_N, P) + \frac{2\varepsilon}{3(b-a)} \cdot (p_n - p_0) \\
&= U(f_N, P) - L(f_N, P) + \frac{2\varepsilon}{3(b-a)} \cdot (b - a) \\
&< \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

Using Proposition 6.2.8, we conclude that h is integrable on $[a,b]$.

We now show that $\int_a^b h = \lim\limits_{n\to\infty} \int_a^b f_n$. Let $\varepsilon > 0$. Since $\langle f_n \rangle$ converges uniformly to h on $[a,b]$, we can fix $N \in \mathbb{N}^+$ such that for all $x \in [a,b]$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \frac{\varepsilon}{2(b-a)}$. Thus, for any $n \geq N$, we have

$$\begin{aligned}
\left| \int_a^b f_n - \int_a^b h \right| &= \left| \int_a^b (f_n - h) \right| && \text{(by Proposition 6.2.14)} \\
&\leq \int_a^b |f_n - h| && \text{(by Proposition 6.2.14)} \\
&\leq \frac{\varepsilon}{2(b-a)} \cdot (b-a) && \text{(by Proposition 6.2.14)} \\
&\leq \frac{\varepsilon}{2} \\
&< \varepsilon.
\end{aligned}$$

Therefore, $\lim\limits_{n\to\infty} \int_a^b f_n = \int_a^b h$. □

With continuity and integrability saved by the switch to uniform convergence over convergence, it may be disappointing to learn that differentiability doesn't fare as well. We start with an example where convergence is uniform and everything in sight is differentiable, but we don't get $h'(x) = \lim\limits_{n\to\infty} f_n'(x)$. The idea is to use more and more rapidly oscillating sine curves which go to zero nicely, but at a slower rate than the oscillations.

For each $n \in \mathbb{N}^+$, let $f_n \colon \mathbb{R} \to \mathbb{R}$ be the function given by $f_n(x) = \frac{1}{n}\sin(2\pi n^2 x)$. Let $h \colon \mathbb{R} \to \mathbb{R}$ be the zero function, i.e. $h(x) = 0$ for all $x \in \mathbb{R}$. We showed about that $\langle f_n \rangle$ converges uniformly to h on $\mathbb{R}$. Notice that each f_n is differentiable on $\mathbb{R}$, and of course h is also differentiable on $\mathbb{R}$. Now for each $n \in \mathbb{N}^+$, we have

$$\begin{aligned}
f_n'(x) &= \frac{1}{n} \cos(2\pi n^2 x) \cdot 2\pi n^2 \\
&= 2\pi n \cos(2\pi n^2 x)
\end{aligned}$$

for all $x \in \mathbb{R}$, hence

$$\begin{aligned} f_n'(0) &= 2\pi n \cos(0) \\ &= 2\pi n \end{aligned}$$

for all $n \in \mathbb{N}^+$. It follows that $\lim_{n\to\infty} f_n'(0) = \infty$, so we certainly do not have $\lim_{n\to\infty} f_n'(0) = h'(0)$.

It is even possible to introduce a "corner" in a uniform limit of sequence of differentiable functions so that the resulting function is not differentiable. Define a sequence of functions $\langle f_n \rangle$ on $\mathbb{R}$ by letting

$$f_n(x) = \begin{cases} -x & \text{if } x < -\frac{1}{n} \\ \frac{1}{2}(nx^2 + \frac{1}{n}) & \text{if } -\frac{1}{n} \leq x < \frac{1}{n} \\ x & \text{if } x \geq \frac{1}{n} \end{cases}$$

for every $n \in \mathbb{N}^+$. Let $h\colon \mathbb{R} \to \mathbb{R}$ be the absolute value function, i.e. $h(x) = |x|$ for all $x \in \mathbb{R}$. We know that h is not differentiable at 0. Using Proposition 7.1.4, it is straightforward to check that each f_n is differentiable on $\mathbb{R}$. We claim that $\langle f_n \rangle$ converges uniformly to h on $\mathbb{R}$. To see this, let $n \in \mathbb{N}^+$. We seek to calculate $||f_n - h||$. Notice that

$$(f_n - h)(x) = \begin{cases} 0 & \text{if } x < -\frac{1}{n} \\ \frac{1}{2}(nx^2 + \frac{1}{n}) + x & \text{if } -\frac{1}{n} \leq x < 0 \\ \frac{1}{2}(nx^2 + \frac{1}{n}) - x & \text{if } 0 \leq x < \frac{1}{n} \\ 0 & \text{if } x \geq \frac{1}{n} \end{cases}$$

We find the maximum and minimum of $f_h - h$ on $[-\frac{1}{n}, 0]$ and $[0, \frac{1}{n}]$. Now $f_n - h$ is continuous on $[-\frac{1}{n}, 0]$ and differentiable on $(-\frac{1}{n}, 0)$ by Corollary 5.2.7 and Corollary 6.1.8. Since f_n is continuous on the compact set $[-\frac{1}{n}, 0]$, we know from the Extreme Value Theorem that f_n achieves a maximum and a minimum on $[-\frac{1}{n}, 0]$. By the discussion after Proposition 6.1.10, these extreme values must occur either at an endpoints, or at a point $x \in (-\frac{1}{n}, 0)$ with $(f_n - h)'(x) = 0$. Now for all $x \in (-\frac{1}{n}, 0)$, we have $(f_n - h)'(x) = nx + 1$, so $(f_n - h)'(x) \neq 0$ for all $x \in (-\frac{1}{n}, 0)$. Thus, the maximum and minimum values of $f_n - h$ on the interval $[-\frac{1}{n}, 0]$ must occur at the endpoints. Since

$$(f_n - h)\left(-\frac{1}{n}\right) = \frac{1}{2} \cdot \left(\frac{1}{n} + \frac{1}{n}\right) - \frac{1}{n} = 0$$

and

$$(f_n - h)(0) = \frac{1}{n},$$

we conclude that $0 \leq (f_n - h)(x) \leq \frac{1}{n}$ for all $x \in [-\frac{1}{n}, 0]$. A similar calculation shows that $0 \leq (f_n - h)(x) \leq \frac{1}{n}$ for all $x \in [0, \frac{1}{n}]$. It follows that $|f_n(x) - h(x)| \leq \frac{1}{n}$ for all $x \in \mathbb{R}$, and hence $0 \leq ||f_n - h|| \leq \frac{1}{n}$. Since $\lim_{n\to\infty} \frac{1}{n} = 0$, we conclude from the Squeeze Theorem that $\lim_{n\to\infty} ||f_n - h|| = 0$, and thus $\langle f_n \rangle$ converges uniformly to h on $\mathbb{R}$.

At this point, you may think that differentiability is a lost cause. However, we do know the Fundamental Theorem of Calculus, and we know that integration works out. Thus, with suitable assumptions, we should be able to make differentiability work nicely. In this light, the key thing that we need is not that $\langle f_n \rangle$ converges uniformly to h, but instead that $\langle f_n' \rangle$ converges uniformly to *something*. Throw in a few tame assumptions to make the argument work, and we get the following.

Theorem 7.2.8. *Let $\langle f_n \rangle$ be a sequence of functions on $[a, b]$, and let $h\colon [a, b] \to \mathbb{R}$. Assume the following:*

1. *For each $n \in \mathbb{N}^+$, f_n is continuous on $[a, b]$ and differentiable on (a, b).*

2. *$\langle f_n \rangle$ converges pointwise to h on $[a, b]$.*

3. *There exists a continuous function $g\colon [a, b] \to \mathbb{R}$ such that $\langle f'_n \rangle$ converges uniformly to g on $[a, b]$.*

4. *For each $n \in \mathbb{N}^+$, the function f'_n is integrable on $[a, b]$.*

We then have that h is continuous on $[a, b]$, is differentiable on (a, b) and that $h' = g$ on (a, b). Thus, in particular, we have $h'(x) = \lim_{n\to\infty} f'_n(x)$ for all $x \in (a, b)$.

Before moving on to the proof, we pause to point out a few technicalities. Since we only define the derivative of a function at an interior point of the interval on which it is defined, technically, f'_n doesn't make sense at either a or b. Thus, assumptions (3) and (4) above do not make literal sense. We can handle this in a couple of ways. One simple thing to do would be to change (3) to say that $\langle f'_n \rangle$ converges uniformly to g on (a, b), and then notice that the value of f'_n on the endpoints does not affect whether the resulting function is integrable on $[a, b]$ (by Corollary 6.2.13). Another approach would be to assume that everything is happening on an open interval containing $[a, b]$. A third approach would be to just work with left-hand and right-hand derivatives at the endpoints. Finally, we could instead simply define $f'_n(a) = g(a)$ and $f'_n(b) = g(b)$. Just note the the proof does not depend on which approach you take.

Proof. For every $x \in [a, b]$, we know that $\langle f'_n \rangle$ converges uniformly to g on $[a, x]$, so we can apply Theorem 7.2.7 to conclude that

$$\int_a^x g = \lim_{n\to\infty} \int_a^x f'_n.$$

Hence, for any $x \in [a, b]$, we can use the Fundamental Theorem of Calculus to conclude that

$$\begin{aligned}
\int_a^x g &= \lim_{n\to\infty} \int_a^x f'_n \\
&= \lim_{n\to\infty} (f_n(x) - f_n(a)) \\
&= \lim_{n\to\infty} f_n(x) - \lim_{n\to} f_n(a) \\
&= h(x) - h(a).
\end{aligned}$$

Adding $h(a)$ to both sides, we conclude that

$$h(x) = h(a) + \int_a^x g$$

for every $x \in [a, b]$. By the Fundamental Theorem of Calculus, we know that $x \mapsto \int_a^x g$ is continuous, so h is continuous on $[a, b]$. Also, for any $c \in (a, b)$, since g is continuous at c, we may use the Fundamental Theorem of Calculus to conclude that h is differentiable at c, and that $h'(c) = 0 + g(c) = g(c)$. Therefore, $h' = g$ on (a, b). □

In fact, if you are willing to work a lot harder and use the Mean Value Theorem (instead of the Fundamental Theorem of Calculus), you can get rid of several of the assumptions. See Theorem 6.3.1 in the book.

Given a sequence of functions $\langle f_n \rangle$ on a set A, how do we determine whether it converges if we do not have a potential limit in hand? For pointwise convergence, we can simply use the fact that a sequence of numbers converges if and only if it is Cauchy. Thus, if we know that for every $x \in A$, the sequence $\langle f_n(x) \rangle$ is Cauchy, then we can define $h\colon A \to \mathbb{R}$ by letting $h(x) = \lim_{n\to\infty} f_n(x)$, from which it follows immediately that $\langle f_n \rangle$ converges pointwise to h. What if we instead consider uniform convergence? The following definition provides a uniform twist for the Cauchy criterion.

Definition 7.2.9. *Let $A \subseteq \mathbb{R}$ and let $\langle f_n \rangle$ be a sequence of functions on A. We say that $\langle f_n \rangle$ is* uniformly Cauchy *on A if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}^+$ such that for all $x \in A$ and all $n, m \geq N$, we have $|f_n(x) - f_m(x)| < \varepsilon$.*

Intuitively, give a sequence of bounded functions $\langle f_n \rangle$ on a set A, the sequence $\langle f_n \rangle$ is uniformly Cauchy if and only if the sequence $\langle ||f_n|| \rangle$ of is a Cauchy sequence of real numbers. Working out the details here is a good exercise.

Proposition 7.2.10. *Let $A \subseteq \mathbb{R}$ and let $\langle f_n \rangle$ be a sequence of functions on A. The sequence $\langle f_n \rangle$ is uniformly convergent on A if and only if $\langle f_n \rangle$ is uniformly Cauchy on A.*

Proof. Suppose first that $\langle f_n \rangle$ is uniformly convergent on A. Fix $h \colon A \to \mathbb{R}$ such that $\langle f_n \rangle$ converges uniformly to h on A. Let $\varepsilon > 0$. Since $\langle f_n \rangle$ converges uniformly to h on A, we can fix $N \in \mathbb{N}^+$ such that for all $x \in A$ and all $n \geq N$, we have $|f_n(x) - h(x)| < \frac{\varepsilon}{2}$. Let $x \in A$ and $n, m \geq N$ be arbitrary. We then have

$$\begin{aligned} |f_n(x) - f_m(x)| &= |f_n(x) - h(x) + h(x) - f_m(x)| \\ &\leq |f_n(x) - h(x)| + |h(x) - f_m(x)| \\ &= |f_n(x) - h(x)| + |f_m(x) - h(x)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } n, m \geq N\text{)} \\ &= \varepsilon. \end{aligned}$$

Therefore, $\langle f_n \rangle$ is uniformly Cauchy on A.

Suppose now that $\langle f_n \rangle$ is uniformly Cauchy on A. Notice that for each fixed $x \in A$, the sequence of real numbers $\langle f_n(x) \rangle$ is a Cauchy sequence, hence converges. Define $h \colon A \to \mathbb{R}$ by letting $h(x) = \lim\limits_{n \to \infty} f_n(x)$ for every $x \in A$. We show that $\langle f_n \rangle$ converges uniformly to h on A. Let $\varepsilon > 0$. Since $\langle f_n \rangle$ is uniformly Cauchy on A, we can fix $N \in \mathbb{N}^+$ such that for all $x \in A$ and all $n, m \geq N$, we have $|f_n(x) - f_m(x)| < \frac{\varepsilon}{2}$. Let $x \in A$ and $m \geq N$ be arbitrary. We show that $|f_m(x) - h(x)| < \varepsilon$. Notice that for all $n \geq N$, we have $|f_m(x) - f_n(x)| < \frac{\varepsilon}{2}$, so

$$f_n(x) - \frac{\varepsilon}{2} < f_m(x) < f_n(x) + \frac{\varepsilon}{2}.$$

Since $f_n(x) - \frac{\varepsilon}{2} < f_m(x)$ for all $m \in \mathbb{N}^+$, and we know that both $\lim\limits_{n \to \infty} (f_n(x) - \frac{\varepsilon}{2}) = h(x) - \frac{\varepsilon}{2}$ and $\lim\limits_{n \to \infty} f_m(x) = f_m(x)$, we can apply Theorem 2.2.10 to conclude that $h(x) - \frac{\varepsilon}{2} \leq f_m(x)$. Similarly, we have $f_m(x) - \frac{\varepsilon}{2} \leq h(x)$. Combining these inequalities, it follows that

$$h(x) - \frac{\varepsilon}{2} \leq f_m(x) \leq h(x) + \frac{\varepsilon}{2},$$

so

$$h(x) - \varepsilon < f_m(x) < h(x) + \varepsilon,$$

and hence $|f_m(x) - h(x)| < \varepsilon$. Therefore, $\langle f_n \rangle$ converges uniformly to h on A. □

7.3 Series of Functions

Since we know how to add functions, we can define series of functions in the obvious way.

Definition 7.3.1. *Let $A \subseteq \mathbb{R}$, $\langle f_n \rangle$ be a sequence of functions on A, and let $h \colon A \to \mathbb{R}$. Define a new sequence of functions $\langle g_n \rangle$ on A by letting $g_n = f_1 + f_2 + \cdots + f_n$ for every $n \in \mathbb{N}^+$. We call $\langle g_n \rangle$ the* sequence of partial sums of $\langle f_n \rangle$ on A.

1. *We say that* $\sum_{n=1}^{\infty} f_n$ converges pointwise to h on A *if the sequence* $\langle g_n \rangle$ *of functions converges pointwise to* h *on* A.

2. *We say that* $\sum_{n=1}^{\infty} f_n$ converges uniformly to h on A *if the sequence* $\langle g_n \rangle$ *of functions converges uniformly to* h *on* A.

Definition 7.3.2. *Let* $A \subseteq \mathbb{R}$ *and let* $\langle f_n \rangle$ *be a sequence of functions on* A. *Let* $\langle g_n \rangle$ *be the sequence of partial sums of* $\langle f_n \rangle$ *on* A.

1. *We say that* $\sum_{n=1}^{\infty} f_n$ converges pointwise on A *if there exists a function* $h \colon A \to \mathbb{R}$ *such that* $\sum_{n=1}^{\infty} f_n$ *converges pointwise to* h *on* A.

2. *We say that* $\sum_{n=1}^{\infty} f_n$ converges uniformly on A *if there exists a function* $h \colon A \to \mathbb{R}$ *such that* $\sum_{n=1}^{\infty} f_n$ *converges uniformly to* h *on* A.

In either case, we write $\sum_{n=1}^{\infty} f_n$ *to denote the unique such* h.

Since continuity, integrability, differentiability all behave well with respect to both finite sums and uniform limits, we immediately obtain the following three propositions.

Proposition 7.3.3. *Let* $A \subseteq \mathbb{R}$, *let* $\langle f_n \rangle$ *be a sequence of continuous function on* A, *and let* $h \colon [a, b] \to \mathbb{R}$. *Suppose that* $\sum_{n=1}^{\infty} f_n$ *converges uniformly to* h *on* A. *We then have that* h *is continuous on* A.

Proof. Let $\langle g_n \rangle$ be the sequence of partial sums of $\langle f_n \rangle$. Since $g_n = f_1 + f_2 + \cdots + f_n$ for every $n \in \mathbb{N}^+$, and each f_n is continuous on A, it follows from Theorem 5.2.6 that each g_n is continuous on A. Since $\langle g_n \rangle$ converges uniformly to h on A, we may use Theorem 7.2.6 to conclude that h is continuous on A. □

Proposition 7.3.4. *Let* $\langle f_n \rangle$ *be a sequence of integrable function on* $[a, b]$ *and let* $h \colon [a, b] \to \mathbb{R}$. *Suppose that* $\sum_{n=1}^{\infty} f_n$ *converges uniformly to* h *on* $[a, b]$. *We then have that* h *is integrable on* $[a, b]$ *and that*

$$\int_a^b h = \sum_{n=1}^{\infty} \int_a^b f_n.$$

Proof. Let $\langle g_n \rangle$ be the sequence of partial sums of $\langle f_n \rangle$ and let $\langle s_n \rangle$ be the sequence of partial sums of $\langle \int_a^b f_n \rangle$. Since $g_n = f_1 + f_2 + \cdots + f_n$ for every $n \in \mathbb{N}^+$, and each f_n is integrable on $[a, b]$, it follows from Proposition 6.2.14 that each g_n is integrable on $[a, b]$ and that

$$\begin{aligned} \int_a^b g_n &= \int_a^b (f_1 + f_2 + \cdots + f_n) \\ &= \int_a^b f_1 + \int_a^b f_2 + \cdots + \int_a^b f_n \\ &= s_n \end{aligned}$$

for all $n \in \mathbb{N}^+$. Since $\langle g_n \rangle$ converges uniformly to h on $[a, b]$, we may use Theorem 7.2.7 to conclude that h is integrable on $[a, b]$ and that

$$\int_a^b h = \lim_{n \to \infty} \int_a^b g_n = \lim_{n \to \infty} s_n.$$

Since we have

$$\sum_{n=1}^{\infty} \int_a^b f_n = \lim_{n \to \infty} s_n$$

by definition, it follows that

$$\int_a^b h = \sum_{n=1}^{\infty} \int_a^b f_n.$$

□

Proposition 7.3.5. *Let $\langle f_n \rangle$ be a sequence of functions on $[a,b]$ and let $h\colon [a,b] \to \mathbb{R}$. Assume the following:*

1. *For each $n \in \mathbb{N}^+$, f_n is continuous on $[a,b]$ and differentiable on (a,b).*
2. *$\sum_{n=1}^{\infty} f_n$ converges pointwise to h on $[a,b]$.*
3. *There exists a continuous function $g\colon [a,b] \to \mathbb{R}$ such that $\sum_{n=1}^{\infty} f_n'$ converges uniformly to g on $[a,b]$.*
4. *For each $n \in \mathbb{N}^+$, the function f_n' is integrable on $[a,b]$.*

We then have that h is continuous on $[a,b]$, is differentiable on (a,b) and that $h' = g$ on (a,b). Thus, in particular, we have $h'(x) = \sum_{n=1}^{\infty} f_n'(x)$ for all $x \in (a,b)$.

Proof. Let $\langle g_n \rangle$ be the sequence of partial sums of $\langle f_n \rangle$. Since $g_n = f_1 + f_2 + \cdots + f_n$ for every $n \in \mathbb{N}$ and each f_n is differentiable on $[a,b]$, it follows from Theorem 6.1.6 that each g_n is differentiable on (a,b) and that $g_n' = f_1' + f_2' + \cdots + f_n'$ on $[a,b]$. Furthermore, since each f_n' is integrable on $[a,b]$, it follows that each g_n' is integrable on $[a,b]$. Now using the facts that $\langle g_n \rangle$ converges pointwise to h on $[a,b]$ and that $\langle g_n' \rangle$ converges uniformly to g on $[a,b]$, we may use Theorem 7.2.8 to conclude that h is differentiable on $[a,b]$ and that $h' = g$ for all $x \in [a,b]$. □

We can now ask a question analogous to the one we asked for a sequence of functions. How can we determine when a given series of functions converges uniformly if we do not have a potential limit in hand? The following test is extremely useful. The idea is that as long as the sequence of functions $\langle f_n \rangle$ are getting small *fast* in the sense that $\sum_{n=1}^{\infty} ||f_n||$ converges (not just that $\lim_{n\to\infty} ||f_n|| = 0$), then we should be able to add the functions up nicely and in fact get uniform convergence to the resulting limit.

Theorem 7.3.6 (Weierstrass M-test)**.** *Let $A \subseteq \mathbb{R}$, and $\langle f_n \rangle$ be a sequence of functions on A, and let $\langle M_n \rangle$ be sequence of numbers. Suppose that $||f_n|| \leq M_n$ for all $n \in \mathbb{N}^+$, and that $\sum_{n=1}^{\infty} M_n$ converges. We then have the following:*

1. *For all $x \in A$, the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely (and hence converges).*
2. *$\sum_{n=1}^{\infty} f_n$ converges uniformly on A.*

Proof. For each $x \in A$, we have $0 \leq |f_n(x)| \leq ||f_n|| \leq M_n$, hence $\sum_{n=1}^{\infty} |f_n(x)|$ converges by the Comparison Test. Thus, $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely for each $x \in A$, and hence converges for each $x \in A$.

Let $\langle g_n \rangle$ be the sequence of partial sums of $\langle f_n \rangle$, i.e. $g_n = f_1 + f_2 + \cdots + f_n$ for all $n \in \mathbb{N}^+$. To show that $\langle g_n \rangle$ converges uniformly, we show that $\langle g_n \rangle$ is uniformly Cauchy. Let $\varepsilon > 0$. Since $\sum_{n=1}^{\infty} M_n$ converges, we can use Proposition 3.1.7 to fix $N \in \mathbb{N}^+$ such that $|M_{k+1} + M_{k+2} + \cdots + M_n| < \varepsilon$ whenever $n > k \geq N$. Let $x \in A$ and $n > k \geq N$ be arbitrary. Now let $x \in A$ and $n > k \geq N$ be arbitrary. We then have

$$\begin{aligned} |g_n(x) - g_k(x)| &= |f_{k+1}(x) + f_{k+2}(x) + \cdots + f_n(x)| \\ &= |f_{k+1}(x)| + |f_{k+2}(x)| + \cdots + |f_n(x)| \\ &\leq ||f_{k+1}|| + ||f_{k+2}|| + \cdots + ||f_n|| \\ &\leq M_{k+1} + M_{k+2} + \cdots + M_n \\ &< \varepsilon. \end{aligned}$$

□

7.4 Power Series

We now begin a study of the most important class of series of functions. The idea is to generalize polynomials to "infinite polynomials". Intuitively, we want to look at a function like

$$h(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots.$$

In other words, for each real number x, we look at the infinite series

$$\sum_{n=0}^{\infty} a_nx^n = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots,$$

and let $h(x)$ be the resulting value (assuming that the corresponding series converges). However, this is a point-by-point way to think about power series. A more abstract, but more unified and powerful, approach is to view the above scenario as a series of functions. That is, for each $n \in \mathbb{N}^+$, let $f_n(x) = a_nx^n$, and then consider the series $\sum_{n=0}^{\infty} f_n$. Here is the formal definition, which in general allows you to change to "center" of the polynomial.

Definition 7.4.1. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence of real numbers and let $c \in \mathbb{R}$. For each $n \in \mathbb{N}$, let $f_n : \mathbb{R} \to \mathbb{R}$ be given by $f_n(x) = a_n(x-c)^n$ (so $f_0(x) = a_0$). We call $\sum_{n=0}^{\infty} f_n$ a* power series centered at c.

We will focus almost entirely on the case when $c = 0$, but know that essentially everything carries over to the case where $c \neq 0$. Since a series is defined in terms of partial sums, we are really defining a sequence $\langle g_n \rangle$ of functions by letting $g_n = f_1 + f_2 + \dots + f_n$ for all $n \in \mathbb{N}$, which means that we are letting

$$g_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n.$$

In this way, we are looking at a sequence of polynomial functions that "cohere" in that we are just adding on another higher-power term at each step.

As mentioned above, a power series (centered at 0) looks like an "infinite polynomial":

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \dots.$$

Of course, just because we can write this down doesn't mean that it makes sense. It is therefore of great interest to us to determine for which values of x a given power series $\sum_{n=0}^{\infty} a_nx^n$ converges. Any power series will certainly converge when $x = 0$. Intuitively, as we move x further and further away from 0, it seems that it will be less and less likely that $\sum_{n=0}^{\infty} a_nx^n$ will converge because the terms a_nx^n will be larger. Let's see a few examples:

- Consider the series:
$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \dots.$$
We claim that this series converges if and only if $x \in (-1, 1)$. First notice that if $x \in (-1, 1)$, then $|x| < 1$, so $\sum_{n=0}^{\infty} x^n$ converges by Proposition 3.1.2. If $x \notin (-1, 1)$, then it is not the case that $\lim_{n\to\infty} x^n = 0$ (by Proposition 2.3.11), hence $\sum_{n=0}^{\infty} x^n$ diverges by Proposition 3.1.3.

- Consider the series
$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots.$$
By Proposition 3.3.3, we know that this series converges for all $x \in \mathbb{R}$.

- Consider the series

$$\sum_{n=0}^{\infty} n^n x^n = 0 + x + 4x^2 + 27x^3 + 256x^4 + \dots.$$

We claim that this series converges if and only if $x = 0$. Clearly it converges when $x = 0$. Suppose then that $x \neq 0$. For any $n \in \mathbb{N}^+$ with $n > \frac{1}{|x|}$, we have $|nx| > 1$, so $|n^n x^n| = |nx|^n > 1$. It particular, we certainly do not have $\lim_{n\to\infty} n^n x^n = 0$. Therefore, $\sum_{n=0}^{\infty} n^n x^n$ diverges if $x \neq 0$.

From these examples, it appears that as we move away from zero, we (may) reach a breaking point before which we get convergence and after which we get divergence. At this breaking point, interesting things can happen as witnessed in the following example. Consider

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots.$$

We claim that this power series converges if and only if $x \in [-1, 1)$. As always, the series converges when $x = 0$. For any $x \neq 0$, we have

$$\begin{aligned}
\lim_{n\to\infty} \left| \frac{x^{n+1}/(n+1)}{x^n/n} \right| &= \lim_{n\to\infty} \left| \frac{x^{n+1}}{n+1} \cdot \frac{n}{x^n} \right| \\
&= \lim_{n\to\infty} |x| \cdot \frac{n}{n+1} \\
&= |x| \cdot \lim_{n\to\infty} \cdot \frac{1}{1 + (1/n)} \\
&= |x| \cdot \frac{1}{1+0} \\
&= |x|.
\end{aligned}$$

Therefore, by the Ratio Test, $\sum_{n=0}^{\infty} \frac{x^n}{n}$ converges if $|x| < 1$ and diverges if $|x| > 1$. We need to handle the cases $x = 1$ and $x = -1$ separately. When $x = 1$, we obtain the Harmonic Series $\sum_{n=0}^{\infty} \frac{1}{n}$, which diverges. When $x = -1$, we obtain that Alternating Harmonic Series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$, which converges. Putting it all together, we conclude that $\sum_{n=1}^{\infty} \frac{x^n}{n}$ converges if and only if $x \in [-1, 1)$.

We now prove the fundamental theorem that as soon as we know convergence of $\sum_{n=0}^{\infty} a_n x_0^n$ for a particular x_0, then we get convergence for all x with $|x| < |x_0|$. In fact, we show that if we move back just a little from x_0, then we get the best kind of convergence that we can hope for. That is, if we take a positive r closer to 0 than x_0, then for each fixed $x \in [-r, r]$ we get pleasing absolute convergence of the series $\sum_{n=0}^{\infty} a_n x^n$, and we also get nice global behavior in that the the series of functions $\sum_{n=0}^{\infty} a_n x^n$ convergence uniformly on $[-r, r]$. Study this result carefully, because it is the basis of everything that follows.

Proposition 7.4.2. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence and let $x_0 \in \mathbb{R}$. Suppose that $\sum_{n=0}^{\infty} a_n x_0^n$ converges. Let $r \in \mathbb{R}$ with $0 < r < |x_0|$. We then have that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in [-r, r]$ and converges uniformly on $[-r, r]$.*

Proof. Since $\sum_{n=0}^{\infty} a_n x_0^n$ converges, we know from Proposition 3.1.3 that $\lim_{n\to\infty} a_n x_0^n = 0$, hence the sequence $\langle a_n x_0^n \rangle_{n=0}^{\infty}$ is bounded by Proposition 2.2.7. Fix $d \in \mathbb{R}$ such that $|a_n| \cdot |x_0|^n = |a_n x_0^n| \leq d$ for all $n \geq 0$. For

any $x \in [-r, r]$ and any $n \in \mathbb{N}$, we have

$$\begin{aligned} |a_n x^n| &= |a_n| \cdot |x|^n \\ &\leq |a_n| \cdot |r|^n \\ &= |a_n| \cdot |x_0|^n \cdot \left|\frac{r}{x_0}\right|^n \\ &\leq d \cdot \left|\frac{r}{x_0}\right|^n \end{aligned}$$

Since $|\frac{r}{x_0}| < 1$, we know from Proposition 3.1.2 that $\sum_{n=0}^{\infty} d \cdot |\frac{r}{x_0}|^n$ converges. Using the Weierstrass M-test, we conclude that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for each $x \in [-r, r]$ and also converges uniformly on $[-r, r]$. □

We that result in hand, we can now formalize our intuition that every power series comes equipped with a "breaking point" inside which everything is beautiful and outside which everything is awful (at the breaking point itself, weird stuff can happen).

Corollary 7.4.3. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence. Exactly one of the following holds:*

1. *$\sum_{n=0}^{\infty} a_n x^n$ converges if and only if $x = 0$.*
2. *$\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in \mathbb{R}$. In this case, for each $r > 0$, the series of functions $\sum_{n=0}^{\infty} a_n x^n$ converges uniformly on $[-r, r]$.*
3. *There exists $R > 0$ such that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in (-R, R)$ and diverges for all x with $|x| > R$. In this case, for each r with $0 < r < R$, the series of functions $\sum_{n=0}^{\infty} a_n x^n$ converges uniformly on $[-r, r]$.*

Proof. Clearly at most one of the three possibilities occurs, so we just need to show that at least one happens. We also know that $\sum_{n=0}^{\infty} a_n x^n$ trivially converges when $x = 0$. If $\sum_{n=0}^{\infty} a_n x^n$ does not converge for any $x \neq 0$, then we are in case (1). Suppose then that $\sum_{n=0}^{\infty} a_n x^n$ converges for some $x \neq 0$. We now have two cases:

- *Case 1:* Assume first that $\sum_{n=0}^{\infty} a_n x^n$ converges for all $x \in \mathbb{R}$. Let $r > 0$ be arbitrary. Let $x_0 = r + 1$. Since $\sum_{n=0}^{\infty} a_n x_0^n$ converges, we know from Proposition 7.4.2 that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in [-r, r]$ and converges uniformly on $[-r, r]$. Since $r > 0$ was arbitrary, it follows that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in \mathbb{R}$.

- *Case 2:* Assume that $\sum_{n=0}^{\infty} a_n x^n$ diverges for some x. Fix some $y \in \mathbb{R}$ such that $\sum_{n=0}^{\infty} a_n y^n$ diverges. For any x_0 with $|x_0| > |y|$, Proposition 7.4.2 implies that $\sum_{n=0}^{\infty} a_n x_0^n$ diverges. Therefore, the set $\{|x| : \sum_{n=0}^{\infty} a_n x^n \text{ converges}\}$ is bounded above by $|y|$. Let $R = \sup\{|x| : \sum_{n=0}^{\infty} a_n x^n \text{ converges}\}$. Since we're not in case (1), we know that $R > 0$.

  For any x with $|x| > R$, we know from Proposition 7.4.2 that $\sum_{n=0}^{\infty} a_n x^n$ diverges. Let $r \in \mathbb{R}$ be arbitrary with $0 < r < R$. Since r is not an upper bound for $\{|x| : \sum_{n=0}^{\infty} a_n x^n \text{ converges}\}$, we can fix $x_0 \in \mathbb{R}$ with $r < |x_0|$ such that $\sum_{n=0}^{\infty} a_n x_0^n$ converges. Using Proposition 7.4.2, it follows that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in [-r, r]$ and converges uniformly on $[-r, r]$. Since r satisfying $0 < r < R$ was arbitrary, it follows that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in (-R, R)$.

□

Definition 7.4.4. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence. We define the* radius of convergence *of the power series $\sum_{n=0}^{\infty} a_n x^n$ as follows. If case (1) of Corollary 7.4.3 occurs, we say that the radius of convergence is 0. If case (2) occurs, we say that the radius of convergence is ∞. If case (3) occurs, we call the corresponding unique R the radius of convergence.*

In what follows, we don't want to handle the case $R = \infty$ differently from the case $R > 0$ constantly, so we agree that if $R = \infty$, then $-R$ denotes $-\infty$, and hence $(-R, R) = (-\infty, \infty) = \mathbb{R}$. Moreover, if $R = \infty$ and we write $r < R$, just interpret that to mean that $r \in \mathbb{R}$.

Now consider an arbitrary power series. Since we have uniform convergence on any slightly cut-off piece inside the radius of convergence, and we're summing polynomials which are continuous functions, it follows that the function defined by the power series inside the radius of convergence is always continuous.

Proposition 7.4.5. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence, let R be the radius of convergence of the power series $\sum_{x=0}^{\infty} a_n x^n$, and suppose $R \neq 0$. Define $h\colon (-R, R) \to \mathbb{R}$ by letting $h(x) = \sum_{n=0}^{\infty} a_n x^n$. We then have that h is continuous on $(-R, R)$.*

Proof. It suffice to show that f is continuous on $(-r, r)$ for all r with $0 < r < R$. So let r be arbitrary with $0 < r < R$. Fix $x_0 \in \mathbb{R}$ with $r < x_0 < R$. We then have that $\sum_{x=0}^{\infty} a_n x_0^n$ converges, so using Proposition 7.4.2, it follows that $\sum_{x=0}^{\infty} a_n x^n$ converges uniformly to h on $[-r, r]$. Now each function function $f_n(x) = a_n x^n$ is continuous on $[-r, r]$ by Corollary 5.2.7, so it follows from Proposition 7.3.3 that h is continuous on $[-r, r]$. Since r with $0 < r < R$ was arbitrary, we conclude that h is continuous on $(-R, R)$. □

With continuity taken care of, the next question is whether the function defined by a power series (inside the radius of convergence) is always differentiable. Suppose then that we have a power series

$$\sum_{x=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

with radius of convergence $R \neq 0$. Intuitively, we might expect that we can differentiate the "infinite polynomial" term-by-term to obtain the derivative

$$\begin{aligned} \sum_{x=1}^{\infty} n a_n x^{n-1} &= \sum_{x=0}^{\infty} (n+1) a_{n+1} x^n \\ &= a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots. \end{aligned}$$

In order to justify this intuition, we look toward Proposition 7.3.5. When examining the hypotheses of this result, we notice that we need to argue that this new "formally differentiated" power series converges uniformly on $(-R, R)$ as well. In order to prove this result, we first establish the following lemma.

Lemma 7.4.6. *For any $r \in \mathbb{R}$ with $|r| < 1$, the series $\sum_{n=0}^{\infty} n r^n$ converges.*

Proof. Let $r \in \mathbb{R}$ be arbitrary with $|r| < 1$. If $r = 0$, then the series trivially converges. Suppose then that $r \neq 0$. Notice that

$$\begin{aligned} \lim_{n\to\infty} \left| \frac{(n+1) r^{n+1}}{n r^n} \right| &= \lim_{n\to\infty} \frac{n+1}{n} \cdot |r| \\ &= |r| \cdot \lim_{n\to\infty} \frac{1 + \frac{1}{n}}{1} \\ &= |r| \cdot \frac{1+0}{1} \\ &= |r| \\ &< 1. \end{aligned}$$

Therefore, $\sum_{n=1}^{\infty} n r^n$ converges by the Ratio Test. □

We now use this lemma to argue that if x_0 is a point of convergence of the original power series, then the formally differentiated power series converges at any point x that is strictly closer to 0. We also do the same for in the reverse direction, i.e. if the formally differentiated power series converges at at x_0, then the original power series converges at any point x that is strictly closer to 0.

Proposition 7.4.7. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence.*

1. *Suppose that $x_0 \in \mathbb{R}$ and $\sum_{n=0}^{\infty} a_n x_0^n$ converges. For any $x \in \mathbb{R}$ with $|x| < |x_0|$, the series*

$$\sum_{n=1}^{\infty} n a_n x^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n$$

converges.

2. *Suppose that $x_0 \in \mathbb{R}$ and*

$$\sum_{n=1}^{\infty} n a_n x_0^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} x_0^n$$

converges. For any $x \in \mathbb{R}$ with $|x| < |x_0|$, the series $\sum_{n=0}^{\infty} a_n x^n$ converges.

Proof. Notice that in both parts, we may assume that $x_0 \neq 0$.

1. Since $\sum_{n=0}^{\infty} a_n x_0^n$ converges, we know from Proposition 3.1.3 that $\lim\limits_{n\to\infty} a_n x_0^n = 0$, hence the sequence $\langle a_n x_0^n \rangle_{n=0}^{\infty}$ is bounded by Proposition 2.2.7. Fix $d \in \mathbb{R}$ such that $|a_n x_0^n| \leq d$ for all $n \in \mathbb{N}$. Let $x \in \mathbb{R}$ be arbitrary with $|x| < |x_0|$. If $x = 0$, then the series converges trivially, so assume that $x \neq 0$. For any $n \in \mathbb{N}^+$, we have

$$\begin{aligned} |n a_n x^{n-1}| &= n \cdot |a_n| \cdot |x|^{n-1} \\ &= n \cdot |a_n| \cdot \frac{|x|^n}{|x|} \cdot \frac{|x_0|^n}{|x_0|^n} \\ &= \frac{|a_n x_0^n|}{|x|} \cdot n \cdot \left| \frac{x}{x_0} \right|^n \\ &\leq \frac{d}{|x|} \cdot n \cdot \left| \frac{x}{x_0} \right|^n \end{aligned}$$

   Now $|\frac{x}{x_0}| < 1$, so by the lemma, we know that $\sum_{n=0}^{\infty} n \cdot |\frac{x}{x_0}|^n$ converges, and hence $\sum_{n=0}^{\infty} \frac{d}{|x|} \cdot n \cdot |\frac{x}{x_0}|^n$ converges. Using the Comparison Test, it follows that $\sum_{n=1}^{\infty} n a_n x^{n-1}$ converges absolutely, and hence converges.

2. Since $\sum_{n=1}^{\infty} n a_n x_0^{n-1}$ converges, we know that $\sum_{n=1}^{\infty} n a_n x_0^n$ converges as well. Using Proposition 3.1.3, it follows that $\lim\limits_{n\to\infty} n a_n x_0^n = 0$, hence the sequence $\langle n a_n x_0^n \rangle_{n=1}^{\infty}$ is bounded by Proposition 2.2.7. Fix $d \in \mathbb{R}$ such that $|n a_n x_0^n| \leq d$ for all $n \in \mathbb{N}^+$. Let $x \in \mathbb{R}$ be arbitrary with $|x| < |x_0|$. For any $n \in \mathbb{N}^+$,

we have

$$\begin{aligned}
|a_n x^n| &= |a_n| \cdot |x|^n \\
&= |a_n| \cdot |x|^n \cdot \frac{|x_0|^n}{|x_0|^n} \\
&= |a_n x_0^n| \cdot \left|\frac{x}{x_0}\right|^n \\
&\leq |n a_n x_0^n| \cdot \left|\frac{x}{x_0}\right|^n \\
&\leq d \cdot \left|\frac{x}{x_0}\right|^n .
\end{aligned}$$

Now $|\frac{x}{x_0}| < 1$, so we know that $\sum_{n=0}^{\infty} d \cdot |\frac{x}{x_0}|^n$ converges. Using the Comparison Test, it follows that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely, and hence converges.

□

The following corollary is now immediate (for latter, notice that if you differentiate term-by-term, you arrive at the original power series).

Corollary 7.4.8. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence, and let R be the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$. Then R is the radius of convergence of the power series*

$$\sum_{n=1}^{\infty} n a_n x^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n$$

and also of the power series

$$\sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1} = \sum_{n=1}^{\infty} \frac{a_{n-1}}{n} x^n.$$

In other words, if we have a power series, and we either formally differentiate the it term-by-term, or we formally take an anti-derivative term-by-term, then we do not affect the radius of convergence. We can now prove the result that motivated our desire to understand that radius of convergence of the "formally differentiated" power series.

Theorem 7.4.9. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence, let R be the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$, and suppose that $R \neq 0$. Define $h \colon (-R, R) \to \mathbb{R}$ by letting*

$$h(x) = \sum_{n=0}^{\infty} a_n x^n$$

for all $x \in (-R, R)$. Since R is also the radius of convergence of $\sum_{n=1}^{\infty} n a_n x^{n-1}$, we can define $g \colon (-R, R) \to \mathbb{R}$ by letting

$$\begin{aligned}
g(x) &= \sum_{n=1}^{\infty} n a_n x^{n-1} \\
&= \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n
\end{aligned}$$

for all $x \in (-R, R)$. Then h is differentiable on $(-R, R)$ and $h' = g$ on $(-R, R)$.

Proof. As in the proof of Proposition 7.4.5, it suffices to prove that for all r with $0 < r < R$ we have that h is differentiable on $[-r, r[$ and that $h' = g$ on $[-r, r]$. So let $r \in \mathbb{R}$ be arbitrary with $0 < r < R$. For each $n \geq 0$, define $f_n \colon [-r, r] \to \mathbb{R}$ by letting $f_n(x) = a_n x^n$ for all $x \in [-r, r]$. We have the following:

1. For each $n \in \mathbb{N}$, f_n is continuous on $[-r, r]$ and is differentiable on $(-r, r)$ (since each f_n is a polynomial).
2. $\sum_{n=0}^{\infty} f_n$ converges pointwise to h on $[-r, r]$.
3. $\sum_{n=0}^{\infty} f_n'$ converges uniformly to g on $[-r, r]$ (since R is also the radius of convergence of the power series $\sum_{n=1}^{\infty} n a_n x^{n-1}$). Also, g is continuous on $[-r, r]$ by Proposition 7.4.5.
4. f_n' is integrable on $[-r, r]$ for every $n \in \mathbb{N}$ (again since each f_n' is a polynomial).

Therefore, by Proposition 7.3.5, h is differentiable on $[-r, r]$ and $h' = g$ on $[-r, r]$. Since $r \in \mathbb{R}$ with $0 < r < R$ was arbitrary, the result follows. □

We can now immediately use this result to get the corresponding result about taking an anti-derivative term-by-term.

Corollary 7.4.10. *Let $\langle a_n \rangle_{n=0}^{\infty}$ be a sequence, let R be the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$, and suppose that $R \neq 0$. Define $h \colon (-R, R) \to \mathbb{R}$ by letting*

$$h(x) = \sum_{n=0}^{\infty} a_n x^n$$

for all $x \in (-R, R)$ and define $g \colon (-R, R) \to \mathbb{R}$ by letting

$$\begin{aligned} g(x) &= \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^n \\ &= \sum_{n=1}^{\infty} \frac{a_{n-1}}{n} x^n \end{aligned}$$

for all $x \in (-R, R)$. Then g is differentiable on $(-R, R)$ and $g' = h$ on $(-R, R)$.

7.5 Representing Functions as Power Series

In the previous section, we talked about how any power series defines a function inside its radius of convergence. Moreover, we showed that the resulting function was differentiable, and that its derivative can be expressed by the power series you get by taking a term-by-term derivative (the result of which has the same radius of convergence). This flexibility allows us to construct many interesting examples of new differentiable functions. However, what if we want to go the other way? Suppose we have a function, and we want to express it as a power series. How can we do that?

We start this investigation with the only nontrivial power series that we understand. For any $x \in (-1, 1)$, we know that $\sum_{n=0}^{\infty} x^n$ converges and in fact we have

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \dots$$

We have therefore expressed a function on $(-1, 1)$ in 2 different ways: as a simple quotient $\frac{1}{1-x}$ and as a power series. Once we've expressed a function as a power series, we can often express others as power series

as well. For example, for any $x \in (-1,1)$, we have $-x \in (-1,1)$, hence

$$\begin{aligned}\frac{1}{1+x} &= \frac{1}{1-(-x)} \\ &= \sum_{n=0}^{\infty}(-x)^n \\ &= \sum_{n=0}^{\infty}(-1)^n x^n \\ &= 1 - x + x^2 - x^3 + x^4 - \dots.\end{aligned}$$

For a more interesting example, notice that for any $x \in (-2,2)$, we have $|\frac{x}{2}| < 1$, so $|\frac{x^2}{4}| = |\frac{x}{2}|^2 < 1$, and hence

$$\begin{aligned}\frac{12}{4-x^2} &= 3 \cdot \frac{1}{1-\frac{x^2}{4}} \\ &= 3 \cdot \sum_{n=0}^{\infty}\left(\frac{x^2}{4}\right)^n \\ &= \sum_{n=0}^{\infty}\frac{3x^{2n}}{4^n} \\ &= 3 + \frac{3x^2}{4} + \frac{3x^4}{16} + \frac{3x^6}{64} + \dots.\end{aligned}$$

In other words, we can make a lot of progress in representing functions as power series by simply performing some algebraic tricks on the geometric series.

We can go further by using the results of the previous section. We have

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \dots$$

for all $x \in (-1,1)$, and we know how to take the derivative of $\frac{1}{1-x}$ and how to take the derivative of (a function represented by) a power series. Thus, using Theorem 7.4.9 together with the fact that the derivative of $\frac{1}{1-x}$ is $\frac{1}{(1-x)^2}$, we conclude that

$$\begin{aligned}\frac{1}{(1-x)^2} &= \sum_{n=1}^{\infty} n x^{n-1} \\ &= \sum_{n=0}^{\infty}(n+1)x^n \\ &= 1 + 2x + 3x^2 + 4x^3 + \dots\end{aligned}$$

for all $x \in (-1,1)$.

We can also go in the other direction using Corollary 7.4.10. We know from above that

$$\frac{1}{1+x} = \sum_{n=0}^{\infty}(-1)^n x^n = 1 - x + x^2 - x^3 + \dots$$

for all $x \in (-1,1)$. Thus, by Corollary 7.4.10, the function defined by the power series

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1} &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots\end{aligned}$$

has derivative equal to $\frac{1}{1+x}$ on $(-1,1)$. Since $\ln(1+x)$ also has derivative equal to $\frac{1}{1+x}$ on $(-1,1)$, we can apply Corollary 6.1.14 to fix a $C \in \mathbb{R}$ such that $\log(1+x) = C + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$ for all $x \in (-1,1)$. Plugging in $x = 0$, we see that

$$0 = \ln(1+0) = C + 0,$$

hence $C = 0$. Therefore,

$$\begin{aligned}\ln(1+x) &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots\end{aligned}$$

for all $x \in (-1,1)$. Notice that although

$$\sum_{n=0}^{\infty} (-1)^n x^n = 1 - x + x^2 - x^3 + \dots$$

converges only for $x \in (-1,1)$, the series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

converges for $x \in (-1,1]$ because

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

is the alternating harmonic series (although the series have the same radius of convergence, the endpoints might behave differently). As a result, it is reasonable to expect that the alternating harmonic series converges to $\ln(1+1) = \ln 2$. We have not developed the tools to prove this result, but see Abel's Theorem in the book.

Although we have had some real success manipulating one particular power series to express other functions as power series, such an approach can only go so far. For example, it's not clear how we would make progress on representing some trigonometric or exponential functions as a power series. Now at first glance, in order to check that a given function can be represented as a power series, we need to check all possible power series arising from all possible sequences $\langle a_n \rangle_{n=0}^{\infty}$. Fortunately, the sequence of coefficients $\langle a_n \rangle_{n=0}^{\infty}$ can be "read off" from properties of the function defined by $\sum_{n=0}^{\infty} a_n x^n$. This fact will allow us to focus on just one particular power series to see whether it represents the function we have in mind.

Proposition 7.5.1. *Suppose that $\langle a_n \rangle_{n=0}^{\infty}$ is a sequence. Let R be the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$ and suppose $R \neq 0$. Suppose that f is function such that there exists $\delta \in \mathbb{R}$ with $0 < \delta \leq R$ such that*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

for all $x \in (-\delta, \delta)$. We then have that f is infinitely differentiable on $(-\delta, \delta)$ and that $a_n = \frac{f^{(n)}(0)}{n!}$ for all $n \in \mathbb{N}$, where $f^{(n)}$ is the n^{th} derivative of f.

Proof. By Theorem 7.4.9, we know that the function $x \mapsto \sum_{n=0}^{\infty} a_n x^n$ is differentiable on $(-R, R)$, that $\sum_{n=0}^{\infty}(n+1)a_{n+1}x^n$ has radius of convergence R, and that

$$f'(x) = \sum_{n=0}^{\infty}(n+1)a_{n+1}x^n$$

for all $x \in (-\delta, \delta)$. Applying Theorem 7.4.9 again, it follows that $\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n$ has radius of convergence R, and that

$$f''(x) = \sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n$$

for all $x \in (-\delta, \delta)$. By a simple induction, we conclude that for all $k \in \mathbb{N}$, the power series

$$\sum_{n=0}^{\infty}(n+k)(n+k-1)(n+k-2)\cdots(n+1)a_{n+k}x^n$$

has radius of convergence R, and that

$$f^{(k)}(x) = \sum_{n=0}^{\infty}(n+k)(n+k-1)(n+k-2)\cdots(n+1)a_{n+k}(x-c)^n$$

for all $x \in (-\delta, \delta)$. Therefore, for any $k \in \mathbb{N}$, we have

$$\begin{aligned} f^{(k)}(0) &= \sum_{n=0}^{\infty}(n+k)(n+k-1)(n+k-2)\cdots(n+1)a_{n+k}0^n \\ &= k(k-1)(k-2)\cdots 1 \cdot a_k \\ &= k! \cdot a_k, \end{aligned}$$

and hence $a_k = \frac{f^{(k)}(0)}{k!}$. □

Thus, if we have a function f and we know that f is infinitely differentiable in some neighborhood of 0, then there is only one possible power series that can represent f in a neighborhood of 0. We give this power series a special name.

Definition 7.5.2. *Let f be a function. Suppose there there exists $\delta > 0$ such that f is infinitely differentiable on $(-\delta, \delta)$, i.e. such that $f^{(n)}(x)$ exists for all $x \in (-\delta, \delta)$ and all $n \in \mathbb{N}$. The power series*

$$\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n$$

is called the Taylor series of f centered at 0*, or the* Maclaurin series of f*.*

For example, let $f(x) = e^x$ for all $x \in \mathbb{R}$. Since $f^{(n)}(x) = e^x$ for all $n \in \mathbb{N}^+$, it follows that $f^{(n)}(0) = e^0 = 1$ for all $n \in \mathbb{N}$. Therefore, the Taylor series of f centered at 0 is the series

$$\sum_{n=0}^{\infty}\frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots,$$

which we know has radius of convergence ∞ (see Proposition 3.3.3).

For another example, let $f(x) = \sin x$ for all $x \in \mathbb{R}$. We have

$$\begin{aligned} f(x) &= \sin x \\ f^{(1)}(x) &= \cos x \\ f^{(2)}(x) &= -\sin x \\ f^{(3)}(x) &= -\cos x \\ f^{(4)}(x) &= \sin x \\ \ldots &= \ldots \end{aligned}$$

Therefore, by a simple induction, it follows that

$$\begin{array}{llllll} 0 & = f^{(0)}(0) & = f^{(2)}(0) & = f^{(4)}(0) & = f^{(6)}(0) & = \ldots \\ 1 & = f^{(1)}(0) & = f^{(5)}(0) & = f^{(9)}(0) & = f^{(13)}(0) & = \ldots \\ -1 & = f^{(3)}(0) & = f^{(7)}(0) & = f^{(11)}(0) & = f^{(15)}(0) & = \ldots \end{array}$$

It follows that the Taylor series of f centered at 0 is the series

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots.$$

It is now a simple exercise to check that this series has radius of convergence ∞.

Notice that at this point, if we are given a function f, then we know the only potential example of a power series that will represent f in a neighborhood of 0. In other words, for the two functions above, the corresponding power series are the only possible series that could give e^x and $\sin x$. However, we do *not* yet know that these power series do indeed work! They are simply the only options we need to consider.

In fact, in general, the Taylor series of f centered at 0 might only agree with f at 0 itself. For an example of this phenomenon, consider the function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Notice that

$$\begin{aligned} \lim_{x \to 0} e^{-1/x^2} &= \lim_{x \to 0} \frac{1}{e^{1/x^2}} \\ &= 0, \end{aligned}$$

so f is continuous at 0, and hence is continuous everywhere. To go further, we need to use the fact that for all $k \in \mathbb{N}$, we have

$$\lim_{x \to 0} \frac{e^{-1/x^2}}{x^k} = 0,$$

which can be proven by some careful inequality work, or by L'Hopital's Rule.

Now notice that for any $x \neq 0$, we have that f is differentiable at x and that

$$f'(x) = 2x^{-3} e^{-1/x^2}.$$

We also have

$$\begin{aligned} f'(0) &= \lim_{x \to 0} \frac{e^{-1/x^2}}{x} \\ &= 0 \end{aligned}$$

by the above limit, so

$$f'(x) = \begin{cases} 2x^{-3}e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

When we look at the second derivative, things get interesting. For any $x \neq 0$, we have

$$\begin{aligned} f''(x) &= 2x^{-3} \cdot 2x^{-3}e^{-1/x^2} + (-6x^{-4}) \cdot e^{-1/x^2} \\ &= (4x^{-6} - 6x^{-4}) \cdot e^{-1/x^2} \\ &= (4 - 6x^2) \cdot x^{-6} \cdot e^{-1/x^2}. \end{aligned}$$

We also have

$$\begin{aligned} f''(0) &= \lim_{x \to 0} \frac{2x^{-3}e^{-1/x^2}}{x} \\ &= \lim_{x \to 0} \frac{2e^{-1/x^2}}{x^4} \\ &= 2 \cdot \lim_{x \to 0} \frac{e^{-1/x^2}}{x^4} \\ &= 0 \end{aligned}$$

by the above limit. Following these ideas, it is possible to show by induction that for each $n \in \mathbb{N}$, there exists a polynomial $p(x)$ and an $m \in \mathbb{N}$ such that

$$f^{(n)}(x) = \begin{cases} p(x)x^{-m}e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

It follows that $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$. Hence, the Taylor series of f centered at 0 is $\sum_{n=0}^{\infty} 0x^n$, which is the zero function. Thus, the Taylor series of f centered at 0 has radius of convergence ∞, but $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n$ if and only if $x = 0$.

With such a setback, we need tools to determine when the Taylor series of f centered at 0 actually does represent f.

Proposition 7.5.3. *Let f be a function defined on $(-r, r)$, and assume that f is infinitely differentiable in some neighborhood of 0. For each $n \in \mathbb{N}$, let $a_n = \frac{1}{n!} \cdot f^{(n)}(0)$. For each $m \in \mathbb{N}$, let*

$$S_m(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m,$$

and let

$$E_m(x) = f(x) - S_m(x).$$

If $x \in (-r, r)$ and $\lim_{m \to \infty} E_m(x) = 0$, then $\sum_{n=0}^{\infty} a_nx^n$ converges and equals $f(x)$.

Proof. We have $S_m(x) = f(x) - E_m(x)$. Since $E_m(x)$ goes to 0 and $f(x)$ goes to $f(x)$, it follows that $S_m(x)$ converges and equals $f(x)$. □

Theorem 7.5.4 (Lagrange's Remainder Theorem)**.** *Let f be a function defined on $(-r, r)$, and assume that the first $m + 1$ derivatives of f all exist throughout $(-r, r)$. For each $n \leq m$, let $a_n = \frac{1}{n!} \cdot f^{(n)}(0)$. Let*

$$S_m(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m,$$

and let

$$E_m(x) = f(x) - S_m(x).$$

For every $x \in (-r, r)$ with $x \neq 0$, there exists c between 0 and x such that

$$E_m(x) = \frac{f^{(m+1)}(c)}{(m+1)!} \cdot x^{m+1}$$

Proof. Notice that for all $n \leq m$, we have $f^{(n)}(0) = S_m^{(n)}(0)$, and hence $E_m^{(n)}(0) = 0$. Let $x \in (-r, r)$ be arbitrary. Assume that $x > 0$ (the case where $x < 0$ is similar). We define a sequence, starting by letting $z_0 = x$. Applying the Cauchy Mean Value Theorem to E_m and $y \mapsto y^{m+1}$, we can fix $z_1 \in (0, z_0)$ with

$$\frac{E_m'(z_1)}{(m+1)z_1^m} = \frac{E_m(z_0) - E_m(0)}{z_0^{m+1} - 0^{m+1}} = \frac{E_m(z_0)}{z_0^{m+1}}.$$

Applying the Cauchy Mean Value Theorem to E_m' and $y \mapsto (m+1)y^m$, we can fix $z_2 \in (0, z_1)$ with

$$\frac{E_m''(z_2)}{(m+1)mz_2^{m-1}} = \frac{E_m'(z_1) - E_m'(0)}{(m+1)z_1^m - (m+1)0^{m+1}} = \frac{E_m'(z_1)}{(m+1)z_1^m} = \frac{E_m(z_0)}{z_0^{m+1}}.$$

Applying the Cauchy Mean Value Theorem to E_m'' and $y \mapsto (m+1)my^{m-1}$, we can fix $z_3 \in (0, z_2)$ with

$$\frac{E_m^{(3)}(z_3)}{(m+1)m(m-1)z_3^{m-2}} = \frac{E_m''(z_2) - E_m''(0)}{(m+1)mz_2^{m-1} - (m+1)m0^{m+1}} = \frac{E_m''(z_2)}{(m+1)mz_2^{m-1}} = \frac{E_m(z_0)}{z_0^{m+1}}.$$

Continue in this way until we get to z_{m+1} with

$$\frac{E_m^{(m+1)}(z_{m+1})}{(m+1)!} = \frac{E_m(z_0)}{z_0^{m+1}}$$

Now notice that $E^{(m+1)} = f^{(m+1)}$ and $z_0 = x$, so

$$\frac{f^{(m+1)}(z_{m+1})}{(m+1)!} = \frac{E_m(x)}{x^{m+1}}$$

□

We're now in a position to prove that the Taylor series of various natural functions do indeed converge to the function throughout their radius of convergence.

Proposition 7.5.5. *For all $x \in \mathbb{R}$, we have*

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

Proof. For $x = 0$, this is trivial. Let $x > 0$ be arbitrary. Let $m \in \mathbb{N}$. Using the Lagrange Remainder Theorem, we can fix $c \in [0, x]$ such that

$$E_m(x) = \frac{f^{(m+1)}(c)}{(m+1)!} x^{m+1}.$$

Notice that $f^{(m+1)}(c)$ is one of $\sin c$, $\cos c$, $-\sin c$ or $-\cos c$, hence $|f^{(m+1)}(c)| \leq 1$. Therefore,

$$\begin{aligned} |E_m(x)| &= \left| \frac{f^{(m+1)}(c)}{(m+1)!} x^{m+1} \right| \\ &\leq \frac{1}{(m+1)!} \cdot x^{m+1} \\ &= \frac{x^{m+1}}{(m+1)!}. \end{aligned}$$

Since $\lim\limits_{m\to\infty} \frac{x^{m+1}}{(m+1)!} = 0$ by Proposition 3.3.3 and Proposition 3.1.3, it follows from the Squeeze Theorem that $\lim\limits_{m\to\infty} |E_m(x)| = 0$, and hence $\lim\limits_{m\to\infty} E_m(x) = 0$. Using Proposition 7.5.3, we conclude that

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

The case when $x < 0$ is similar. □