Real Analysis

Noah Jussila

July 8, 2020

Contents

0 Introduction

This collection of notes is an ongoing project that aims to combine four semesters of real analysis notes into one document. My goal in doing so is to not only stave off the boredom resulting from quarantining due to Covid-19, but also pick and choose from different presentations of standard material in real analysis.

0.1 Prerequisites

These notes will assume familiarity with differential and integral calculus, linear algebra, multivariable calculus, basic set theory, and properties of functions (image, preimage, injectivity, etc.) At times, some material, particularly concepts in linear algebra, will be reviewed. Two very good video series that serve as reviews for calculus and linear algebra can be found at this amazing YouTube channel.

I also will assume familiarity with basic styles of proof (contradiction, contrapositive, etc.). I will try *very hard* to be as detailed as possible with every proof though.

0.2 Organization and Sources

If these notes are ever finished (which is unlikely), they will span about four semesters worth of real analysis: two at an honors undergraduate level, and two at a introductory PhD level.

My selection of sources is motivated not only by the books I really enjoy, but also those that are considered standard. For material presented in an undergraduate course, I absolutely love Tao (2016a) and Tao (2016b). Tao starts at the most basic level of arithmetic and numbers are works all the way up to Lebesgue Integration in his two part series. He motivates nearly every topic, provides examples, and gives the reader a big picture of the subject. That being said, his approach left me confused at times when I first read his books. In the first volume, Tao eschews point-set topology entirely and opts to work exclusively in $\mathbb{R}$. This is great because most people who learn real analysis only ever care about $\mathbb{R}$, but it also makes some concepts overly technical. For instance, I think that seuqences and continuity are easier to present and motivate in general metric spaces. My hunch is that I'm in the minority here, as Terry Tao went the other direction, and he has one more Fields Medal than I do. Tao also doesn't include any figures. Another text that is awesome is the infamous "Baby" Rudin (1976). This book is so widely used for a reason.[1] It's concise, clear, and presents the material with no frills. It makes a great textbook if you have a professor who motivates the material, provides examples, and draws the occasional illustration. If you do not have such a professor, than Rudin (1976) becomes pure torture. This book is very much the outline of the first 7 sections of these notes. Sections 8-11 pull mostly from...

0.3 Presentation

Many standard math textbooks assume the reader can create their own novel examples, fill in the blanks purposefully left in proofs, understand the motivation for the material, and pick up on the subtle "tricks" used in proofs. A good professor will help students do this during a lecture. I want my presentation to do this. In doing this, I'm very much inspired by a professor I had for a second semester honors analysis course I took at Boston College. His presentation of some of the material that will be treated not only will be replicated here, but also motivates how I approach other topics. This means lots of illustrations, footnotes, remarks, and asking hypothetical questions to motivate material. It may seem very pedantic if you're familiar with analysis, but in that case just read Rudin's books.

[1] My uneducated hypothesis about this is that Rudin (1976) really only perfect makes sense if you already know all the material well, so professors who do not remember what it is like to learn analysis just assume everyone understands this book.

1 The Real Numbers

We begin by returning to the most basic concepts in math. What exactly is a number? We begin with the most basic possible set of numbers, and use those to define more complex sets of numbers, with our goal being to define the real numbers. Lastly, we will look at the "size" of these sets, and explore the concept of infinity.

1.1 Natural Numbers, Integers, and Rational Numbers

First, a cursory overview of several sets of numbers is in order. It is given for the sake of exposition, and to illustrate how we define sets of numbers using previously defined sets. For the sake of time, the formal definition of the standard operations (addition, multiplication, etc.) on these sets will be forgone. Rest assured that the operations we are all familiar with are well defined on these sets, and this can be shown rigorously. An excellent reference for this within the context of real analysis can be found in Tao (2016a), who takes nothing as given.

The most basic numbers are those we use to count. We will call these natural numbers.

Definition 1.1. Define the set $\mathbb{N} \coloneqq \{0, 1, 2, \ldots\}$ to be the *natural numbers*.

The operations of addition and multiplication are well defined on $\mathbb{N}$, in that when adding or multiplying natural numbers, the result is a natural number. A more succinct way of putting this is saying that $\mathbb{N}$ is *closed* under addition and multiplication. While the natural numbers are great for things such as counting (a fact we will return to), it fails to be useful for much more. In particular, two basic operations we are familiar with are not well defined on $\mathbb{N}$.

Example 1.1. Suppose we want to find the difference in 2 and 5, both elements in $\mathbb{N}$. The difference in question would be $2 - 5$, but this is not an element of $\mathbb{N}$!

In order to address this shortcoming, we need to broaden our view. In effect, we need to "add more" numbers to $\mathbb{N}$. We want to enlarge the set of natural numbers by the amount necessary for subtraction to be well defined. This can be done by taking the set of all differences of natural numbers. For all intents and purposes, this is how we define the integers.

Definition 1.2. Define the set $\mathbb{Z} \coloneqq \{a - b \mid (a, b) \in \mathbb{N}^2\}$ to be the *integers*. Two integers are equal, $a - b = c - d$, if and only if $a + d = b + c$.[2]

We will take the ordering of $\mathbb{Z}$, the negation of elements of $\mathbb{Z}$, and all arithmetic properties of $\mathbb{Z}$ to be given. There are two things worth noting. The first is that $\mathbb{N} \subseteq \mathbb{Z}$. The identity element $0 \in \mathbb{N}$ gives $a - 0 = a$ for each $a \in \mathbb{N}$, so any natural number can be written as the difference of two natural numbers. Secondly, we defined $\mathbb{Z}$ only by using $\mathbb{N}$. This is crucial, as we will define the rationals only by using $\mathbb{Z}$, and in turn define the real numbers only by using the rationals.

While subtraction is well defined for $\mathbb{Z}$, the same does not hold for division.

Example 1.2. Take the integers -3 and 6, and suppose we are interested in the ratio of the prior to the latter. Obviously,

$$\frac{-3}{6} = -\frac{1}{2},$$

but this is not an element of $\mathbb{Z}$.

We can now "extend" the integers to accommodate for division, in a similar fashion to when we defined the integers using the natural numbers.

Definition 1.3. Define the set $\mathbb{Q} \coloneqq \{a/b \mid (a, b) \in \mathbb{Z}^2,\ b \neq 0\}$ to be the *rational numbers*. Two rational numbers are equal, $a/b = c/d$, if and only if $ad = bc$.[3]

[2] The added specification of when integers are equal can be avoided by defining $\mathbb{Z}$ to be a set of equivalence classes. The equivelence relation $\sim$ would be defined on $\mathbb{N}$ as $(a, b) \sim (c, d)$ when $a + c = b + d$.

[3] Again we could use equivalence classes to define $\mathbb{Q}$. The equivelence relation $\sim$ would be defined on $\mathbb{N}$ as $(a, b) \sim (c, d)$ when $ad = bc$.

1.2 "Holes" in $\mathbb{Q}$

We now begin where the canonical Rudin (1976) opens. Our goal has been, and continues to be, to define the most comprehensive set of numbers possible. It may help to visualize what we have done so far with a number line. We can illustrate any "gaps" or "holes" by using red. This can be seen in Figure 1. Clearly

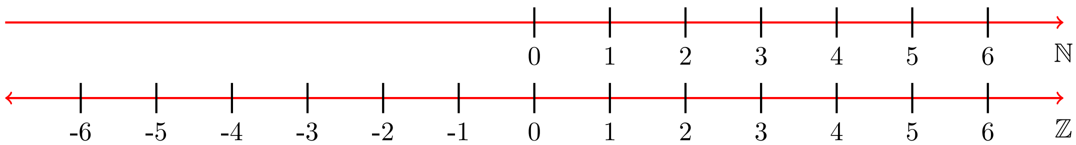

Figure 1: The natural numbers and integers ordered number lines.

the natural numbers and integers are not "comprehensive" in that they have many gaps. This is what led us to define the rational numbers $\mathbb{Q}$. It isn't immediate just how well the rationals do at covering the holes in the integers. We can get a sense of this by introducing a property of the rationals.

Proposition 1.1. (Interspersing of integers by rationals) For any $x, y \in \mathbb{Q}$ where $x < y$, there exists a third rational number $z \in \mathbb{Q}$ such that $x < z < y$.

Proof. Let there be two rationals $x, y \in \mathbb{Q}$ such that $x < y$. We can define the third rational number of interest as $z = (x + y)/2$. We can show that $x < z < y$ by using arithmetic.

$$\begin{aligned} x &< y \\ \frac{x}{2} &< \frac{y}{2} \\ \frac{x}{2} + \frac{y}{2} &< \frac{y}{2} + \frac{y}{2} \\ z &< y \end{aligned}$$

And we can arrive at $x < z$ by adding $x/2$ to each side of the given inequality.

$$\begin{aligned} x &< y \\ \frac{x}{2} &< \frac{y}{2} \\ \frac{x}{2} + \frac{x}{2} &< \frac{y}{2} + \frac{x}{2} \\ x &< z \end{aligned}$$

□

Example 1.3. Take the rational numbers 0 and 1. Using the construction given in the previous proof we have

$$\frac{0+1}{2} = \frac{1}{2}$$

is between 0 and 1. We can now repeat this process using the pairs $(0, 1/2)$ and $(1/2, 1)$.

$$\begin{aligned} \frac{0 + 1/2}{2} &= \frac{1}{4} \\ \frac{1/2 + 1}{2} &= \frac{3}{4} \end{aligned}$$

We could repeat this process an infinite number of times, in effect "filling in" gaps in $\mathbb{Z}$ by successively taking the average of two rational numbers. Figure 2 shows this process on the unit interval in the rationals.
The key question is whether or not this fills *all* the gaps in the integers.

While , and more formally , may lead us to believe that the rational numbers have no gaps, this is unfortunately not the case. There are two classic examples that arise from two of the most basic geometric constructions.

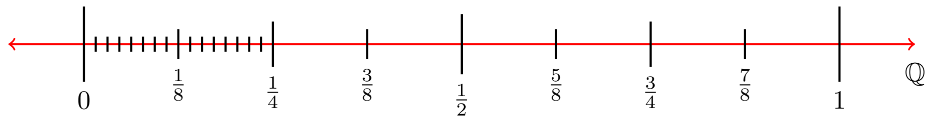

Figure 2:

Example 1.4. Suppose we have a circle with diameter d and circumference c. In this case, the ratio give by c/d is not an element of the rational numbers. This familiar ratio is written as π. For the moment, we can take this as fact. We have not yet developed the tools required to proof that $\pi \notin \mathbb{Q}$, but we will return to this.

Example 1.5. Suppose there is an isosceles right triangle with legs of length 1, as shown in Figure 3. We want to find the length of the hypotenuse x.

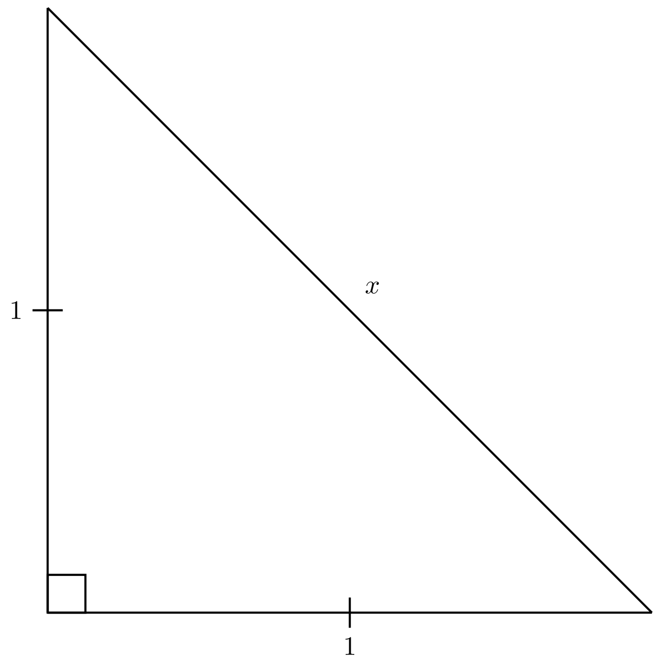

Figure 3:

This is a simple application of the Pythagorean Theorem.

$$\begin{aligned} 1^2 + 1^2 &= x^2 \\ 2 &= x^2 \end{aligned}$$

But this equation has no rational solution, something we can formally prove.

Proposition 1.2. There exists no rational number x which satisfies $x^2 = 2$.

Proof. For the sake of contradiction, suppose that there exists a rational x which satisfies $x^2 = 2$. If this were the case, we could write $x = m/n$ for some $m, n \in \mathbb{Z}$, where m and n are not both even.[4] $\square$

Any x which does satisfy $x^2 = 2$ would be *irrational*, in that it is not an element of $\mathbb{Q}$.

Definition 1.4. A number is *irrational* if it is not an element of $\mathbb{Q}$.

There are *many* irrational numbers, each of which is a gap in the rationals (see Figure 4).

[4]Otherwise we could write x in simpler terms as m and n would have a common factor of 2.

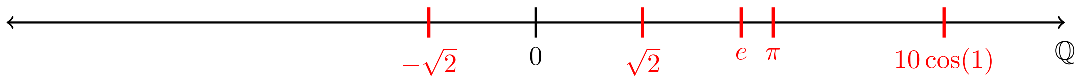

Figure 4:

Our goal now becomes defining a set of numbers that includes not only the rationals, but also all of the irrationals. We began with the natural numbers, and then defined a set $\mathbb{Z}$ which included the additive inverses of the natural numbers. Then we filled more of the gaps in the integers by taking the ratios of integers. We are now faced with the task of defining a set which eliminates the gaps caused by irrational numbers, and doing so eniterly with the set $\mathbb{Q}$.

1.3 sup **and** inf

Before informally constructing the real numbers, it is worth thinking about why $\mathbb{Q}$ has these “holes”, and how it relates to a specific property of sets. It goes without saying that, all the sets of numbers we've discussed up until now have some ordering to them. We can make this formal by defining an ordered set.

Definition 1.5. An *ordered set* is some set S with a binary relation, denoted by $<$,[5] which satisfies the following properties:

1. If $x, y \in S$, then exactly one of the statements

$$x < y,\ x = y,\ y < x$$

   is true.

2. If $x, y, z \in S$, and both $x < y$ and $y < z$, then $x < z$.

The statement “$x < y''$ is read as “x is less than y.”We could also write $y > x$ instead of $x < y$. If we were to negative $y < x$ (“y is not less than x”), we would arrive at “y is either greater than x or equal to x.” This is denoted as $y \geq x$.

Example 1.6. The set $\mathbb{Q}$ is a well ordered set if we define $<$ in the following way for $x, y \in \mathbb{Q}$:

$$x < y \coloneqq y - x \text{ is a postiive rational number.}$$

Note that we can only relate objects that belong to $\mathbb{Q}$. This means that we have no way of comparing rational numbers and the solution to the equation $x^2 = 2$.

We can use the order relation on an ordered set to define bounds on sets.

Definition 1.6. Suppose S is an ordered set, and $E \subseteq S$. If there exists a $\beta \in S$ such that $x \leq \beta$ for all $x \in E$, E is *bounded above*, and β is an *upper bound* of E.

Definition 1.7. Suppose S is an ordered set, and $E \subseteq S$. If there exists a $\beta \in S$ such that $x \geq \beta$ for all $x \in E$, E is *bounded below*, and β is a *lower bound* of E.

A subtlety in both definitions that is extremely important, is that upper and lower bounds must be elements of the ordered set S. The next example highlights this.

Example 1.7. Take $\mathbb{Z}$ to be an ordered set with the natural order. Pick the subset $E = \{-2, -1, 2\} \subseteq \mathbb{Z}$. This set has many upper and lower bounds. For upper bounds we have $2, 3, 4, \ldots$. For lower bounds we have $-2, -3, -4, \ldots$. It may be tempting to say that a fraction such as $5/2$ is an upper bound of E, but it is not. This follows from the fact that $5/2 \notin \mathbb{Z}$, so we have no means of relating it to elements in $\mathbb{Z}$. In this particular case, there are upper and lower bounds of the set are included in the set. This need not be the case, as the next example shows.

[5] In this case, “¡” can mean *any* order. It just so happens that we use the same symbol as the familiar “less than” order, because it is the canonical example of such a relation.

Example 1.8. Let's look at the ordered set $\mathbb{Q}$, and subset $E = \{x \in \mathbb{Q} \mid 0 < x < 1\} \subseteq \mathbb{Q}$.[6] In this case, each element of $\{x \in \mathbb{Q} \mid x \leq 0\}$ is a lower bound of E, and each element of $\{x \in \mathbb{Q} \mid x \geq 1\}$ is an upper bound of E. Even though $0, 1 \notin E$, they are still least and upper bounds of E respectively.

Remark 1.1. It is often obvious what exact order we are talking about when referring to an ordered set, like in the case of $\mathbb{Q}$ and $\mathbb{Z}$. In these cases, we'll just assume we're using the natural order.

We now will introduce two definitions that correspond to a special type of upper and lower bound.

Definition 1.8. Suppose S is an ordered set, $E \subseteq S$, and E is bounded above. We say that α is a *leas-upper-bound* of E if:

1. α is an upper bound of E.
2. If $\gamma < \alpha$, then γ is not an upper bound of E.

Alternatively, we can refer to α as the *supremum* of E, and write $\alpha = \sup E$

Definition 1.9. Suppose S is an ordered set, $E \subseteq S$, and E is bounded below. We say that α is a *greatest-lower -bound* of E if:

1. α is a lower bound of E.
2. If $\gamma > \alpha$, then γ is not a lower bound of E.

Alternatively, we can refer to α as the *infimum* of E, and write $\alpha = \inf E$

Remark 1.2. Both definitions use the definite article *the* before supremum and infimum. This is because they are unique. This is also implied by the use of the superlative *least* and *greatest*. Nevertheless, this is a result of the definition, and can be properly proven.

Example 1.9. If we return to Example 2.7, where $E = \{-2, -1, 2\} \subseteq \mathbb{Z}$, we have $\sup E = 2$ and $\inf E = -2$.

Example 1.10. In example 2.8, $\inf E = 0$ and $\sup E = 1$.

Example 1.11. Sticking with the set $\mathbb{Q}$, consider the subset $E = \{x \in \mathbb{Q} \mid x^2 \leq 2\} \subseteq \mathbb{Q}$. This set has no supremum, because the number satisfying $x^2 = 2$ is not an element of $\mathbb{Q}$ (as shown in Proposition 2.2). We will formally prove this fact shortly.

It is no coincidence that a subset of $\mathbb{Q}$ fails to have a supremum, because of one of the "holes" in $\mathbb{Q}$. The following definition will help us formalize this relationship.

Definition 1.10. Let S be an ordered set. If for all $E \subseteq S$, where E is nonempty and bounded from above, $\sup E$ exists, then S has the *least-upper-bound* property.

The least-upper-bound property ensures that any nontrivial subset of an ordered set has a supremum in that ordered set. We could define an equivelent property known as the greatest-lower-bound property. The next two propositions serves as nice examples of the least-upper-bound property, or lack there of, in action.

Proposition 1.3. The set $\mathbb{Z}$ has the least-upper-bound property.

The idea behind the following proof takes advantage of the fact that $\mathbb{Z}$ is discrete. For some set $E \subseteq \mathbb{Z}$, we can always just look at an upper bound of it, and keep subtracting 1 until the resulting number is in E. Then we will have found our upper bound.

Proof. We will show that an arbitrary nontrivial set $E \subseteq \mathbb{Z}$ has a supremum. Let $x \in E$, and β be an upper bound of E. We know that $\beta \geq x$ for all $x \in E$ We can show that $\sup E$ exists via induction on $\beta - x$ for our arbitrary $x \in E$. Our base case is when $\beta - x = 0$. If this holds, then $\beta \in E$, so $\beta \in \mathbb{Z}$ and $\sup E = \beta$. Now suppose that this statement holds when $\beta - x = k$ for $k \in \mathbb{N}$ (this is our induction hypothesis). It is either the case that $\beta \in E$ or $\beta \notin E$. If $\beta \in E$, then $\sup E = \beta$. If $\beta \notin E$, then let $\beta' = \beta - 1$. Then β' is an upper bound of E, and

$$\beta' - x = \beta - 1 - x = \beta - x - 1 = k + 1 = 1 = k.$$

By the induction hypothesis, $\sup E$ exists. $\square$

[6] The use of the familiar interval notation of $(0, 1)$ will be properly defined and restricted to the real numbers in the following section.

Proposition 1.4. The set $\mathbb{Q}$ does not have the least-upper-bound property.

To prove this, we will first establish that 2 is an upper bound of the set defined in Example 2.11, and then show the set has no supremum via contradiction.

Proof. If suffices to find a single subset of $\mathbb{Q}$ which fails to have a supremum. Let that set be $E = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$.

1. Suppose for contradiction that 2 is not an upper-bound of E. Then there exists an $x \in E$ such that $x > 2$. This would imply that $x^2 > 4$, which contradicts the assumption that $x \in E$.

2. Suppose for contradiction that E has a supremum, and that $\sup E = \alpha$ for $\alpha \in \mathbb{Q}$. Define a new rational number $y \in \mathbb{Q}$ as
$$y = \alpha - \frac{\alpha^2 - 2}{x + 2} = \frac{2(\alpha + 1)}{\alpha + 2}. \tag{1}$$
Squaring this and subtracting 2 gives
$$y^2 - 2 = \frac{4(\alpha + 1)^2}{(\alpha + 2)^2} - \frac{2(\alpha + 2)^2}{(\alpha + 2)^2} = \frac{2(\alpha^2 - 2)}{(\alpha + 2)^2}. \tag{2}$$
We can use y to reach a contradiction in each possible case, those being: $\alpha^2 < 2$, $\alpha^2 = 2$, $\alpha^2 > 2$.

   (a) Suppose that $\alpha^2 < 2$. This means that $\alpha^2 - 2 < 0$, so Equation (1) implies that $y > \alpha$. At the same time, Equation (2) implies that $y^2 - 2 < 0$, which means $y^2 < 2$. This gives that $y \in E$, despite the fact that $\alpha < y$. This contradicts the fact that α is an upper-bound of E

   (b) Suppose $\alpha^2 = 2$. We already know this cannot be the case by Proposition 2.2.

   (c) Finally, assume that $\alpha^2 > 2$, giving $\alpha^2 - 2 = 0$. Equation (1) implies $y < \alpha$ while Equation (2) implies $y^2 - 2 > 0$, meaning $y^2 > 2$. This establishes y as an upper bound for E, but $y < \alpha$, which contradicts $\sup E = \alpha$.

□

The "holes" in $\mathbb{Q}$ are a result of $\mathbb{Q}$ not having the least-upper-bound property. In order to perform calculus, we need a set that has this property, otherwise things like continuity and differentiation would not work. This property is not sufficient in and of itself though. If that were the case then we would have stopped extending out set of numbers at $\mathbb{Z}$. We want a set of numbers as "comprehensive" as $\mathbb{Q}$, but with the least-upper-bound property. It turns out, that this (and a whole lot more) is what we will get from the real numbers.

1.4 The Real Numbers

We will now construct the real numbers using only $\mathbb{Q}$. First, we will define the algebraic structure that the real numbers will take on.

Definition 1.11. A *field* is a set F with two operations, addition and multiplication, which satisfy the following axioms for all $x, y, z \in F$:

1. Axioms for addition:

   (a) $x + y \in F$ (closed under addition)

   (b) $x + y = y + x$ (commutative)

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

   (d) There exists an element $0 \in F$ such that $0 + x = x$ (identity element)

   (e) There exists an element $-x \in F$ such that $x + (-x) = 0$ (inverse element)

2. Axioms for multiplication:

(a) $xy \in F$ (closed under multiplication)

(b) $xy = yx$ (commutative)

(c) $(xy)z = x(yz)$ (associative)

(d) There exists an element $1 \in F$ such that $1x = x$ (identity element)

(e) If $x \neq 0$, there exists an element $1/x \in F$ such that $x(1/x) = 1$ (inverse element)

3. The distributive property:

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

The study of fields is its own entire subject in math, and lives within the discipline if abstract algebra. For more details about fields, see Dummit and Foote (2004). These axioms can be used to reach several familiar conclusions about arithmetic in fields, and can be found as formal propositions in Rudin (1976). A more specific type of field is that which is also an ordered set.

Definition 1.12. An *ordered field* is a field F such that

1. $x + y < x + z$ if $x, y, z \in F$ and $y < z$,
2. $xy > 0$ if $x, y \in F$, $x > 0$, and $y > 0$.

Example 1.12. The set $\mathbb{Q}$ is an ordered field.

Our goal is now to construct an ordered field which not only contains $\mathbb{Q}$, but also has the least-upper-bound property. In order to do this we'll use the fact that $\mathbb{Q}$ has "holes" in it. We'll form a pair of set (A, B) that partition $\mathbb{Q}$ such that each of these partitions corresponds to a real number.

Definition 1.13. A *Dedekind cut* $x = (A, B)$ is a pair of subsets $A, B \subseteq \mathbb{Q}$ satisfying the following:

1. $A \cup B = \mathbb{Q}$, $A \cap B = \emptyset$, $A \neq \emptyset$, and $B \neq \emptyset$.
2. If $a \in A$ and $b \in B$, then $a < b$.
3. A contains no largest element.

Example 1.13. Let $A = \{y \in \mathbb{Q} \mid y < 0\}$ and $B = \{y \in \mathbb{Q} \mid y \geq 0\}$. Our cut is $x = (A, B)$, and can be seen in Figure 5. This cut uniquely represents $0 \in \mathbb{Q}$, as no other cut can be defined in this way "at" 0.

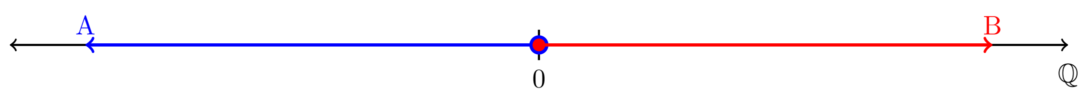

Figure 5: Dedekind cut corresponding to $0 \in \mathbb{Q}$.

Example 1.14. Perhaps a better example is the cut defined by $A = \{q \in \mathbb{Q} \mid q \leq 0 \text{ or } q^2 < 2\}$ and $A = \{q \in \mathbb{Q} \mid q > 0 \text{ or } q^2 > 2\}$. This cut corresponds to the solution of the equation $x^2 = 2$.

Definition 1.14. A *real number* is a Dedekind cut in $\mathbb{Q}$. The set of real numbers is denoted by $\mathbb{R}$.

Definition 1.15. A real number $x = (A, B)$ is a *rational number* if B contains a smallest element (namely x).

Definition 1.16. A real number $x = (A, B)$ is a *irrational number* if B contains no smallest element.

Example 1.15. The cut defined by $A = \{y \in \mathbb{Q} \mid y < 0\}$ and $B = \{y \in \mathbb{Q} \mid y \geq 0\}$ is rational, as B has a smallest element in the form of 0.

Example 1.16. The cut defined by $A = \{q \in \mathbb{Q} \mid q \leq 0 \text{ or } q^2 < 2\}$ and $A = \{q \in \mathbb{Q} \mid q > 0 \text{ or } q^2 > 2\}$ has no smallest element. Therefore it is an irrational number. We will denote this particular number as $\sqrt{2}$.

Now that we have properly defined $\mathbb{R}$, we can *finally* refer to the quantity $\sqrt{2}$! It is no longer a mysterious solution to an equation, and is well defined in $\mathbb{R}$. This is a relatively small payoff, but the real rewards are the two following theorems. These are the main results of this section, and are of the utmost importance. The first will allow us to perform operations on $\mathbb{R}$, and the second will play a key role in proving familiar theorems from calculus. The proof of the first result is rather long, and not very informative, so it is omitted. It is important to understand *how* it would be proved though. Before the big reveal, we will define an order on $\mathbb{R}$

Definition 1.17. Given real numbers $x = (A, B)$ and $y = (C, D)$, we define the following order:

$$x \leq y \coloneqq A \subseteq C.$$

The inequality is strict if $A \subsetneq C$.

Example 1.17. Let $x = 2 = (A, B) = (\{y \in \mathbb{Q} \mid y < 2\}, \{y \in \mathbb{Q} \mid y \geq 2\})$, and $y = 3 = (C, D) = (\{z \in \mathbb{Q} \mid z < 3\}, \{z \in \mathbb{Q} \mid z \geq 3\})$. It should come as no surprise that $2 < 3$, but this is because $A \subsetneq C$.

Theorem 1.1. The set $\mathbb{R}$ is an ordered field containing $\mathbb{Q}$.

Proof. See the appendix of chapter 1 in Rudin (1976). The idea is that addition and multiplication of cuts must be define, and the all the axioms of fields and ordered fields must be verified using the cut definition of a real number. □

Theorem 1.2 (Completeness of the real numbers)**.** The set $\mathbb{R}$ has the least-upper-bound property.

Proof. We will show an arbitrary nonempty subset of $\mathbb{R}$ has a supremum. Let $E \subseteq \mathbb{R}$, where $A \neq \emptyset$, have the upper bound $\beta \in \mathbb{R}$. We may write β as a Dedekind cut, $\beta = (A, B)$. Additionally, we may express each $\alpha \in E$ as a cut $\alpha = (L_\alpha, U_\beta)$. Now we will construct a real number by taking the union of all L_α.

$$\gamma = \left(\bigcup_{\alpha \in E} L_\alpha, \mathbb{Q} \backslash \bigcup_{\alpha \in E} L_\alpha \right) = (L, \mathbb{Q} \backslash L) = (L, U)$$

I claim that $\sup E = \gamma$.

First we must verify that $\gamma \in \mathbb{R}$ by showing that (L, U) is a valid Dedekind cut, and satisfies the requirements of Definition 2.13:

1. The set E is nonempty, so there exists at least one $\alpha = (L_\alpha, U_\alpha) \in E$. Because $L_\alpha \neq \emptyset$ and $U_\alpha \neq \emptyset$ by definition 2.13, we have $L \neq \emptyset$ and $U \neq \emptyset$. By the definition of U as $\mathbb{Q} \backslash L$, we have that $L \cup U = \mathbb{Q}$ and $L \cap U = \emptyset$.
2. FINISH THIS
3. FINISH THIS

□

Therefore $\gamma \in \mathbb{R}$. By construction, $\alpha \leq \gamma$ for all $\alpha \in E$, making γ an upper bound. To show that it is the least-upper-bound, we will now show any number lesser than it cannot be an upper bound. Now suppose $\delta < \gamma$. This means $C \subsetneq L$, where δ is expressed as a cut $\delta = (C, D)$. This means there exists some $s \in L$ such that $s \notin C$. But $s \in L$, so it is in L_α for some $\alpha \in E$. Hence, $C \subsetneq L_\alpha$, giving $\delta < \alpha$. This shows that δ is not an upper bound, meaning $\sup E = \gamma$.

Example 1.18. Consider the set of real numbers $E = \{-1, -1/2, -1/3, -1/4, \ldots\}$. What is the supremum of this set? Intuitively, it should be 0, but we can verify this by constructing it like we did in the previous proof. Each number in E corresponds to a cut (L_n, U_n), for $L_n = \{x \in \mathbb{Q} \mid x < -1/n\}$ and $U_n = \{x \in \mathbb{Q} \mid x \geq -1/n\}$, where $n \in \mathbb{N}$. Our supremum is

$$\gamma = \left(\bigcup_{n \in \mathbb{N}} \{x \in \mathbb{Q} \mid x < -1/n\}, \mathbb{Q} \backslash \bigcup_{n \in \mathbb{N}} \{x \in \mathbb{Q} \mid x < -1/n\} \right) = (\{x \in \mathbb{Q} \mid x < 0\}, \{x \in \mathbb{Q} \mid x \geq 0\}).$$

Therefore, $\gamma = 0$.

You will often here the real numbers referred to as "complete" because they have the least-upper-bound property. This is because the least-upper-bound property ensures there are no "gaps" in the real line like there are in $\mathbb{Q}$. Theorem 1.2 may be the most important result in real analysis. Remember it well, as it will be used often. Most disciplines in math build on themselves over time, and the fact that $\mathbb{R}$ has the least-upper-bound property will be our foundation. One could argue it is $\mathbb{R}$'s defining property.

Finally, we will adopt the familiar notation of intervals in $\mathbb{R}$, and add make an important addition to $\mathbb{R}$.

Definition 1.18. We will use the following notation to refer to *intervals* of $\mathbb{R}$:

$$\begin{aligned}(a,b) &= \{x \in \mathbb{R} \mid a < x < b\}\\ [a,b) &= \{x \in \mathbb{R} \mid a \leq x < b\}\\ (a,b] &= \{x \in \mathbb{R} \mid a < x \leq b\}\\ [a,b] &= \{x \in \mathbb{R} \mid a \leq x \leq b\}\end{aligned}$$

for $a < b$.

Definition 1.19. The *extended real number system* consists of the real field $\mathbb{R}$ and two symbols: ∞, and $-\infty$. The original order of $\mathbb{R}$ is preserved, and we define

$$-\infty < x < \infty$$

for all $x \in \mathbb{R}$. We will denote the extended real numbers as $\overline{\mathbb{R}}$.

The extended real numbers do not form a proper field, but we can adopt some conventions for arithmetic using ∞ and $-\infty$ for $x \in \mathbb{R}$:

1. $x + \infty = \infty$, $x - \infty = -\infty$, $x/\infty = x/-\infty = 0$.
2. For $x > 0$, $x(\infty) = \infty$, and $x(-\infty) = -\infty$.
3. For $x < 0$, $x(\infty) = -\infty$, and $x(-\infty) = \infty$.

The addition of an upper and lower bound on $\mathbb{R}$ make the set $\overline{\mathbb{R}}$ easier to work with in certain situations.[7] The most immediate result of working in $\overline{\mathbb{R}}$ is that *every* subset of $\mathbb{R}$ has a supremum, not just bounded ones (the latter case being the only one stipulated by the least-upper-bound property).

1.5 Properties of $\mathbb{R}$

The importance of Theorem 2.2 can not be understated. It is perhaps *the* defining property of $\mathbb{R}$, and it gives rise to numerous results in analysis. For now, we can use it to prove two additional properties of $\mathbb{R}$.

Theorem 1.3 (Archimedean property of $\mathbb{R}$)**.** For $x, y \in \mathbb{R}$ where $x > 0$, there exists an $n \in \mathbb{N}$ such that $nx > y$.

Proof. Let A be the set of all nx for $x \in \mathbb{R}$ and $n \in \mathbb{N}$, where $x > 0$. For contradiction, suppose that there exists no such $n \in \mathbb{N}$ such that $nx > y$ for $y \in \mathbb{R}$. This makes y an upper bound of A. By the completeness of $\mathbb{R}$, $\sup A = \alpha$ exists. Since $x > 0$, $\alpha - x < \alpha$, and $a - x$ is not an upper bound of A. This means there exists an $m \in \mathbb{N}$ such that $a - x < mx$. But this would imply $\alpha < mx + x = m(x+1)$, where $(m+1)x \in A$. This contradicts the fact that α is an upper bound of A. ☐

Example 1.19. Suppose $x = 10$ and $y = 213$. By the Archimedean property of $\mathbb{R}$, we know there exists a multiple of 10 that is greater than 213.

$$10(22) = 220 > 213$$

Theorem 1.4 ($\mathbb{Q}$ is dense in $\mathbb{R}$)**.** For $x, y \in \mathbb{R}$ where $x < y$, there exists a $p \in \mathbb{Q}$ such that $x < p < y$.

[7] What we are really doing is working with $\overline{\mathbb{R}}$ because it is a complete lattice. A complete lattice is a partially ordered set in which every subset has an infimum or supremum. The real line is not a complete lattice, as any set of the form $(a, \infty) \subset \mathbb{R}$ has no supremum.

Proof. We have $x < y$, giving $y - x > 0$. By the Archimedean property (Theorem 2.3), there exists an $n \in \mathbb{N}$ such that

$$n(y - x) > 1. \tag{3}$$

We can use Theorem 2.3 to find $m_1, m_2 \in \mathbb{N}$ for which:

$$m_1 > nx,$$
$$m_2 > -nx.$$

We can combine these two inequalities to conclude $-m_2 < nx < m_1$. This implies there exists an $m \in \mathbb{N}$ (with $-m_2 \leq m \leq m_1$) such that

$$m - 1 \leq nx < m. \tag{4}$$

If we combine (3) and (4) we get

$$nx < m \leq 1 + nx < ny.$$

Dividing by n (which is positive) gives $x < \frac{m}{n} < y$. $\square$

The density of $\mathbb{Q}$ in $\mathbb{R}$ is both surprising and useful for constructing examples. In practice, it means that every irrational number has a rational number arbitrarily close to it. We can approximate any number in $\mathbb{R}$ arbitrarily well with a rational number.

Example 1.20. We will now mimic the proof of Theorem 2.4 with actual numbers. We will find a rational $p \in \mathbb{Q}$ such that $e < p < \pi$.[8] We have $\pi - e > 0$. We know

$$3(\pi - e) > 1.$$

Next we pick whole numbers $m_1 = 9$ and $m_2 = 8$, and get the inequality $8 < 3e < 9$. Now take m to be 9 and reach our final inequality of

$$3e < 9 < 1 + 3e < 3\pi.$$

Dividing by $n = 3$ gives our desired rational number is $9/3 = 3$.

Example 1.21. Let $\sqrt{2} \in \mathbb{R}$, and let $\varepsilon > 0$.[9] By the density of $\mathbb{Q}$ in $\mathbb{R}$, there exists $p \in \mathbb{Q}$ such that $\sqrt{2} - \varepsilon < p < \sqrt{2}$. This will hold *for all* $\varepsilon > 0$ so as we let ε become smaller and smaller, we will have an increasingly accurate rational approximation of $\sqrt{2}$.

1.6 Cardinality

So far, I've been intentional in avoiding any discussion of the size of the sets we have been working with. When constructing $\mathbb{Z}$ from $\mathbb{N}$, it was never stated that $\mathbb{Z}$ was somehow "bigger" than $\mathbb{N}$. All we know is that $\mathbb{Z}$ has elements that $\mathbb{N}$ does not. The same can be said for $\mathbb{Z}$ and $\mathbb{Q}$, or $\mathbb{Q}$ and $\mathbb{R}$. It is now time that we turn our attention to this matter, and more generally the size, or "cardinality" of sets.

Determining the size of a set amounts to counting the number of elements in that set. But how do we make the notion of counting formal? We will do this with functions. Before formally defining anything, consider how you may count something. If you are tasked with counting the number of elements in the set $X = \{a, b, c\}$, you're answer will surely by 3. How did you get that number? You said assigned the number 1 to a, 2 to b, and 3 to c. We should note three different things about this process:

1. Each number we use is from $\mathbb{N}$.
2. Each element of X is assigned a number. We wouldn't have counted properly if we skipped some element.
3. No number in $\mathbb{N}$ is assigned to multiple elements in X. We do not want to count multiple elements as a single element.

[8] You shouldn't even need to perform the construction to arrive at an answer, as e and π are not particularly "close" to each other. We could for instance take $p = 3$.

[9] This is the first time we're using the infamous ε. It just stands in for any arbitrarily small positive number.

This process of assigning elements in $\mathbb{N}$ to those in X is shockingly similar to the notion of a function, as we are mapping elements from one set to those in another set. Furthermore, the properties we must obey while counting have their own analogous forms with functions: surjectivity, and injectivity. For this reason, we will use functions to formalize the size of a set.

First, we will address the abstract case of when two sets have the same number of elements, and then we will transition to the size of sets.

Definition 1.20. The *cardinality* of a set X, denoted $|X|$, is the number of elements that belong to the set.

Definition 1.21. Two sets X and Y have *the same cardinal number* if there exists a bijection $f : X \to Y$ from X to Y.

Proposition 1.5. Define the relation $X \sim Y$ *if and only if* X and Y have the same cardinal number ($|X| = |Y|$). The relation $\sim$ is an equivalence relation.

Proof. We have that $X \sim X$ by letting $f : X \to X$ be $f(x) = x$, so $\sim$ is reflexive. If $X \sim Y$, there exists a bijection $f : X \to Y$. Since f is a bijection, it has an inverse $f^{-1} : Y \to X$. This inverse is itself a bijection, so $Y \sim X$, making $\sim$ symmetric. Lastly, assume $X \sim Y$ and $Y \sim Z$. We have bijections $f : X \to Y$ and $g : Y \to Z$. The composition of two bijections is a bijection, so $h : X \to Z$ is a bijection where $h : X \to Z$. This makes $\sim$ transitive. $\square$

Example 1.22. Let $X = \{1, 2, 3\}$ and $Y = \{\sqrt{2}, e, \pi\}$. We can define $f : X \to Y$ as

$$f(x) = \begin{cases} \sqrt{2} \text{ if } x = 1 \\ e \text{ if } x = 2 \\ \pi \text{ if } x = 3 \end{cases} .$$

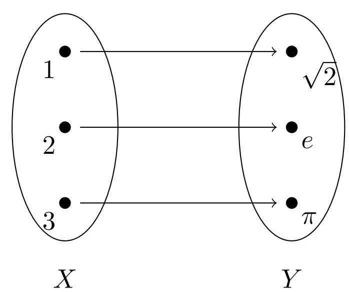

Figure 6: Bijection $f : X \to Y$.

This function is clearly a bijection, and we have that $|X| = |Y|$.

Example 1.23. Let $\mathbb{Z}^- = \{x \in \mathbb{Z} \mid x < 0\}$ and $\mathbb{Z}^+ = \{x \in \mathbb{Z} \mid x > 0\}$. There exists a very natural bijection between these sets, namely that which maps each element in $\mathbb{Z}^+$ to it's negative counterpart in $\mathbb{Z}^-$. Formally, $f : \mathbb{Z}^+ \to \mathbb{Z}^-$ is define as $f(x) = -x$. This function is clearly a bijection, and its existence shows that $|Z^+| = |Z^-|$. There are the same number of positive integers as negative integers

Remark 1.3. Because $f : X \to Y$ is a bijection, it doesn't matter which set is the domain and which set is the codomain. A function is invertible if and only if it is a bijection, and a functions inverse is a bijection, so we would just have $f^{-1} : Y \to X$ if we picked the sets in the other order. In example 2.23, we could have instead assigned each negative integer to it's positive counterpart and had $f : \mathbb{Z}^- \to \mathbb{Z}^+$. In this case we would still have $f(x) = -x$, as this particular function is its own inverse!

As discussed earlier, counting is intrinsically linked to the set of natural numbers $\mathbb{N}$. We will now make this formal by defining three types of sets: finite sets, countably infinite sets, and uncountably infinite sets.

Definition 1.22. A set X is *finite* if there exists a subset of the whole numbers $N \subseteq \mathbb{N}$ for which X and N have the same cardinal number.

Definition 1.23. A set X is *countably infinite* if X has the same cardinal number as $\mathbb{N}$. Alternatively, X is countably infinite if there exists a bijection $f : X \to \mathbb{N}$. This is sometimes denoted as $|X| = \aleph_0$.[10]

Definition 1.24. A set X is *countable* if it is countably infinite or finite.

Remark 1.4. From here on out, I'm going to use countable and countably infinite interchangeably, as nearly all the sets we are interested in are infinite.

Definition 1.25. A set X is *uncountably infinite* (or uncountable) if it is neither finite nor countably infinite.

Before jumping into examples, let's unpack some of this. Finiteness and countable infiniteness depend on whether a set has the same cardinal number as a subset of $\mathbb{N}$ or $\mathbb{N}$ itself. This means we can find a bijection between the set and a subset of $\mathbb{N}$ or $\mathbb{N}$ itself. Definition 2.22 and 2.23 are often presented in terms of this hypothetical bijection. Secondly, two of these definitions involve infinity. A set can be either countably infinite or uncountably infinite. In a sense, some infinite sets have so many elements that they cannot even be counted, and are "bigger" than other uncountable sets! These two concepts of infinity will show up constantly in real analysis. Hopefully examples will make this clear. Some of the following examples are so important that they will be presented as formal results.

Example 1.24. We will modify Example 2.22. Let Let $X = \{\sqrt{2}, e, \pi\}$ and $N = \{1, 2, 3\} \subseteq \mathbb{N}$. Take our bijection to be the inverse of the function defined previously in Example 2.22. This shows that X is finite, and $|X| = 3$.

Proposition 1.6. The set $\mathbb{N}$ is countably infinite.

Proof. Define $f : \mathbb{N} \to \mathbb{N}$ as $f(n) = n$. This function is clearly a bijection. □

Proposition 1.7. The set $\mathbb{Z}$ is countably infinite.

Before proving this result, it's worth acknowledging that it seem paradoxical. How can it be that $\mathbb{Z}$ is the same size as $\mathbb{N}$, despite $\mathbb{Z}$ being defined as $\mathbb{N}$ plus more elements? It would make more sense for $\mathbb{Z}$ to have twice the cardinality as $\mathbb{N}$, but this is not the case. The result may sit better if you consider just how we would count $\mathbb{Z}$. If you started at $1 \in \mathbb{Z}$, followed by $2 \in \mathbb{Z}$, etc. then you would miss all the negative numbers! Instead, we need to be clever in the order in which we count $\mathbb{Z}$. We will instead count in the following order: $0, 1, -1, 2, -2, \ldots$. We could count like this forever and never run out of $\mathbb{N}$, and never miss any elements of $\mathbb{Z}$. This is why we have $|\mathbb{N}| = |\mathbb{Z}|$.

Proof. Recall that we can select either $\mathbb{Z}$ or $\mathbb{N}$ to be the domain of our bijection. We will define $f : \mathbb{N} \to \mathbb{Z}$ as

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

This function counts $\mathbb{Z}$ in the aforementioned manner of alternating between positive and negative integers. We now will verify that f is a bijection, by showing it is injective and surjective.[11] Let $y \in \mathbb{Z}$, and pick $x \in \mathbb{N}$ such that $x = 2y$ if y is even, and $x = -2y + 1$ if y is odd. This choice of $x \in \mathbb{N}$ gives $f(x) = y$, so f is surjective. Now suppose that $f(x_1) = f(x_2)$. If $x_1/2 = x_2/2$, then $x_1 = x_2$. If $-(x_1 - 1)/2 = -(x_2 - 1)/2$, then $x_1 = x_2$. Therefore, f is injective. □

An even more surprising result is that not only $|\mathbb{N}| = |\mathbb{Q}|$, but also $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$! Even if we add every possible fraction to the integers, the size of our set remains the same.

Theorem 1.5. The set $\mathbb{Q}$ is countably infinite.

Proof. We can enumerate the rational numbers in the following way:

$$\frac{0}{1}, \frac{1}{1}, \frac{-1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{1}, \frac{-2}{1}, \frac{-1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{-1}{4}, \frac{2}{3}, \ldots$$

This particular ordering can be seen in Figure 7. The red arrows in Figure 7 show the order in which we count, and it becomes evident that we will eventually count every possible fraction. Note that we only count fractions which are expressed in simplest terms, with others in gray. □

[10] This symbol is an "aleph", and is the first letter of the Hebrew alphabet.
[11] It would be quicker to show f has an inverse, but that approach is not as instructive.

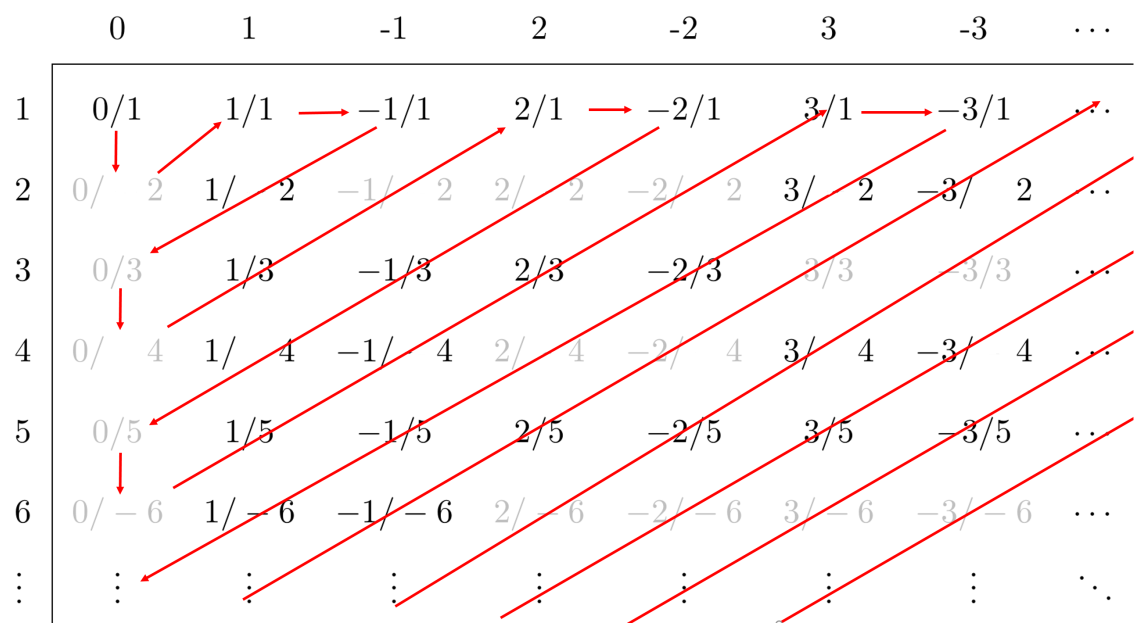

Figure 7:

It may not be a surprise that it is not possible to count $\mathbb{R}$. This puts $\mathbb{R}$ in our second category of infinite sets, uncountably infinite.

Theorem 1.6. The set $\mathbb{R}$ is uncountably infinite.

The proof of this theorem is a classic, and is due to Cantor.

Proof. Suppose for contradiction that $\mathbb{R}$ were countably infinite. There exists a bijection $f : \mathbb{N} \to \mathbb{R}$, and we make a table of values the function takes on (see Table 1). Table 1 is just an example of what such a bijection may look like, and the exact values are moot. Because f is a bijection, this table should go on

$n \in \mathbb{N}$	$f(n) \in \mathbb{R}$
1	4.3214875...
2	1.4918401...
3	3.0194510...
4	9.0194510...
5	0.3917293...
6	5.9184017...
7	1.9284010...
⋮	⋮

forever, and count every element of $\mathbb{R}$. To reach a contradiction, we will simply show there exists a real number that was not counted.[12] "Construct" this uncounted real number in the following way: let the n^{th} digit (decimal places included) take on the value of the n^{th} digit of $f(n)$ minus one (if it is 0, set it to 9). For Table 1, the first digit would be the first digit of $f(1)$ minus 1, which is $4 - 1 = 3$. The second digit would be the second digit of $f(2)$ minus 1, which is $4 - 1 = 3$. We repeat this process for all $n \in \mathbb{N}$, and in our case we get

$$3.308690\ldots$$

By construction, the n^{th} digit of this number is different from at least one of the n^{th} digits of $f(n)$. This holds for every $n \in \mathbb{N}$, so this number is different from every value of $f(n)$, and was therefore not counted. □

[12] There in fact exist *many* that would go uncounted, but it suffices to find just one.

Corollary 1.1. Every infinite subset of $\mathbb{R}$ is uncountable.

Example 1.25. Every interval $[a, b] \subseteq \mathbb{R}$ is uncountable.

Now that we've seen which familiar sets are and are not countable, there are several key results involving the cardinality of sets that deserve attention. These will establish what happens to the cardinality of sets when different set operations are performed.

Proposition 1.8. Let $\{E_n\}$, $n \in \mathbb{N}$, be a sequence of countably infinite sets, and let $E = \cup_{n \in N} E_n$ be a countable union. The set E is countably infinite.

Proposition 1.9. We will prove via induction. Our base case is $n = 2$. The sets E_1 and E_2 are countably infinite, so there exists bijections $f : \mathbb{N} \to E_1$ and $g : \mathbb{N} \to E_2$. Without loss of generality, assume $E_1 \cap E_2 = \emptyset$.[13] Define $h : \mathbb{N} \to E_1 \cup E_2$ as

$$h(k) = \begin{cases} f(k/2) \text{ if } k \text{ is even} \\ g((k+1)/2) \text{ if } k \text{ is odd} \end{cases} .$$

This function counts the elements in the set by alternating between those in E_1 and E_2 (like in the proof of Proposition 2.7). The function h is a bijection, so $E_1 \cup E_2$ is countably infinite. Now suppose this holds for $E_1, \ldots E_{n-1}$. We can write E as a union of two countably infinite sets by taking the union over $E_1, \ldots E_{n-1}$, which is countably infinite by the induction hypothesis.

$$\begin{aligned} E &= \bigcup_{n \in \mathbb{N}} E_n \\ &= E_1 \cup E_2 \cup \cdots E_{n-1} \cup E_n \\ &= (E_1 \cup E_2 \cup \cdots E_{n-1}) \cup E_n \end{aligned}$$

Therefore E is countably infinite.

Corollary 1.2. If X is uncountable, and $E \subseteq X$ is countably infinite, then $X \backslash E$ is uncountably infinite.

Example 1.26. Let $E_n = \{m/n \mid m \in \mathbb{Z}\}$ for $n \in \mathbb{N}$. Each E_n is countable as there is a bijection $f_n : E_n \to \mathbb{Z}$ defined as $f_n(x) = nx$, and $\mathbb{Z}$ is countably infinite.[14] Note that

$$\bigcup_{n \in \mathbb{N}} E_n = \mathbb{Q},$$

which is indeed countably infinite.

Example 1.27. The set $\mathbb{R}$ is uncountable. We have that $\mathbb{Q} \subseteq \mathbb{R}$ is countable. By Corollary 2.1, $\mathbb{R} \backslash \mathbb{Q}$ (the set of irrational numbers) is uncountable. This means that in a certain sense, there are more gaps in $\mathbb{Q}$ than there aren't! We are not even capable of counting all the gaps, whereas we can count $\mathbb{Q}$.

Proposition 1.10. Let X be a countable set. Any subset $Y \subseteq X$ is countable

Proof. There exists a bijection $f : \mathbb{N} \to X$. If we restrict the codomain of f to be Y, f is still a bijection. $\square$

Example 1.28. Every subset of $\mathbb{Q}$ is countable, because $\mathbb{Q}$ is countable. We already know two such examples: $\mathbb{N}$ and $\mathbb{Z}$.

Proposition 1.11. Let $\{E_n\}$, $n = 1, \ldots, m$, be a finite sequence of countably sets, and let $E = \times_{n=1}^{m} E_n$ be a countable Cartesian product. The set E is countably infinite.

[13] Otherwise, we could replace E_1 with $E_1 \backslash E_2$.
[14] This allows us to use the transitivity of sets having the same cardinality.

Proof. It suffices to show the result for two sets E_1 and E_2, and then apply induction using the same argument used in the proof of Proposition 2.9. We have bijections $f : E_1 \to \mathbb{N}$ and $g : E_2 \to \mathbb{N}$. Define $h : E_1 \times E_2$ as

$$h((a, b)) = 2^{f(a)}3^{g(a)},$$

where $(a, b) \in E_1 \times E_2$. Each element in $h(E_1 \times E_2)$ is a whole number with a prime factorization comprised of only 2 and/or 3. Because each element of $\mathbb{N}$ is uniquely determined by its prime factorization, h is injective. Unfortunately, h is no surjective, as there exist many elements of $\mathbb{N}$ with prime factorizations that include more than 2 and/or 3. If we restrict the codomain of h to just its image, we have a bijection $h' : E_1 \times E_2 \to h(E_1 \times E_2)$ We do have that $h(E_1 \times E_2) \subseteq \mathbb{N}$, so by Proposition 2.10, $|h(E_1 \times E_2)| = \aleph_0$. By transitivity, $|E_1 \times E_2| = \aleph_0$. The aforementioned induction can be applied to conclude $|E| = \aleph_0$. $\square$

Example 1.29. The set of all pairs of rational numbers $\mathbb{Q}^2$ is countable.

1.7 Exercises

Exercise 1.1. Show that $\sqrt{3}$ is irrational.

Exercise 1.2. Let $(S, <)$ be an ordered set, and T be a nonempty subset of S. Verify that T has at most one supremum.

Exercise 1.3. Let $(S, <)$ be an ordered set, and A and B be nonempty subsets of T. Show that if $A \subset B$, then $\sup A \leq B$ and $\inf B \leq \inf A$.

Exercise 1.4. Let E be a nonempty subset of an ordered set; suppose α is a lower bound of E and β is an upper bound of E. Prove that $\alpha \leq \beta$.

Exercise 1.5. Let E be a nonempty subset of $\mathbb{R}$ which is bounded below. Let $-E = \{-x \mid x \in E\}$. Show that $\inf E = -\sup(-E)$.

Exercise 1.6. A set has the least-upper-bound property *if and only if* it has the greatest-lower-bound property.

2 Point-Set Topology in Metric Spaces

One of the main goals of calculus is to study rates of change and limiting behavior. Both of these concepts require some notion of distance, and to that end we will study *metric spaces*, sets equipped with a distance function. Any such space has induced "topological" properties. This is a fancy way of saying that we can use distance to categorize different types of sets. Of particular interest, will be the different types of sets in $\mathbb{R}$ and $\mathbb{R}^n$, as the properties these sets have will have major implications down the road.

2.1 Metric Spaces

Our first definition will outline how we endow a set with some notion of distance.

Definition 2.1. A *metric space* is an ordered pair (M, d) where M is a set and $d : M \times M \to [0, \infty]$ is a function which satisfies:

1. $d(x, y) = 0 \iff x = y$.
2. $d(x, y) = d(y, x)$ for all $x, y \in M$ where $x \neq y$.
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in M$. (Triangle Inequality)

The function d is often called the metric, and most of its properties are compatible with our everyday understanding of distance. Firstly, distance cannot be negative. There is no distance between a point and itself. The distance from x to y is the same from y to x. The final property may not be as immediate, but it is extremely important.

Suppose you are traveling from point x to z. If you decide to take a detour to point y before heading to z, then the triangle inequality ensures that you travel a weakly greater distance. An illustration shown in Figure 8 of this gives rise to the inequalities name. Geometrically, this is equivalent to saying that the length

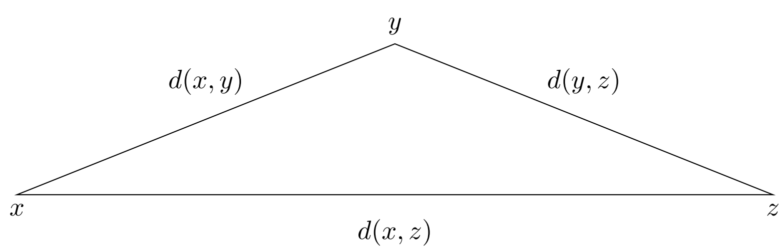

Figure 8: The triangle inequality.

of any side of a triangle cannot be greater than the sum of the other two lengths. Whenever presented with a weak inequality, it is often helpful to ask "when does this hold with equality"? In this case the answer is when y is on the line segment formed by x and z. In this case going to y isn't a detour at all, but just a trivial stop on the way from x to z!

Example 2.1 (Euclidean Metric)**.** The real line $\mathbb{R}$ is a metric space when equipped with the metric $d(x, y) = |x - y|$.

Example 2.2 (Euclidean Metric)**.** Euclidean space $\mathbb{R}^n$ is a metric space when equipped with the metric

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

The Euclidean metric is sometimes refered to as the ℓ^2-metric. The Euclidean metric is intimately linked to the concept of a *norm*. Recall from linear algebra that Euclidean space is a vector space where vectors are elements of $\mathbb{R}^n$, and scalers are elements of $\mathbb{R}$. This space is equipped with function $\|\cdot\|_2 : \mathbb{R}^n \to \mathbb{R}$ that measures the length of vectors.

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$$

We can write the ℓ^2-metric as $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$. Right now, don't worry too much about norms.[15]

Example 2.3 (Taxi-Cab Metric)**.** If our set is $\mathbb{R}^2$ we can let $d(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2|$. This is often referred to as the taxi-cab metric, as it is how you would measure distance if driving a car on a grid. We can extend to $\mathbb{R}^n$ and let $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|$. This metric is also called the ℓ^1-metric.

Example 2.4 (p-Adic Metric)**.** The previous examples are easy to verify, but this may not always be the case. Suppose are set is $\mathbb{Z}$ and $d : \mathbb{Z} \times \mathbb{Z} \to [0, \infty]$ is defined as

$$d(x, y) = \begin{cases} 0 \text{ if } x = y \\ p^{-\max\{m \in \mathbb{N} |\ p^m | (x-y)\}} \text{ otherwise} \end{cases}$$

for some prime number p. Before we verify this is a metric, it's worth getting a feel for how the metric actually works. If $x \neq y$, then the distance between two points is p raised to some negative power. That negative power is defined to be the maximum whole number m such that $(x - y)$ is divisble by p^m. This gives us the vague idea that distance between points x and y is somehow related to how many times p shows up in the prime factorization of $(x - y)$ (where m is the number of times). Let's take $p = 3$, and pick several points in $\mathbb{Z}$ to measure the distance between.

x	y	$x - y$	prime factorization of $x - y$	m	p^{-m}
100	19	81	3^4	4	1/81
368	8	360	$2^3 \cdot 5 \cdot 9$	0	1
35	5	30	$2 \cdot 3 \cdot 5$	1	1/3

It turns out that the more factors of p that go into the prime factorization of $(x - y)$, the closer x and y are. Furthermore, the maximum distance between any two points is 1, as $p^0 = 1$ for all p. We will now verify that this is indeed a metric.

1. The function $d(x, y)$ is defined such that $d(x, y) = 0$ if and only if $x = y$.
2. We have $(x - y) = -(y - x)$. Therefore, the prime factorization of each number differ only in sign, and give the same value m. This implies that $d(x, y) = d(y, x)$.
3. Note that to show $d(x, z) \leq d(x, y) + d(y, z)$ for all points in $\mathbb{Z}$, it suffices to show that $d(x, z) \leq \max\{d(x, y), d(y, z)\}$. This inequality happens to be a stronger condition that implies the triangle inequality. Suppose $p^m \mid (x - y)$ and $p^n \mid (y - z)$. For some $s, r \in \mathbb{Z}$, we having

   $$\begin{aligned} x - y &= p^m r \\ y - z &= p^n s. \end{aligned}$$

   We can combine these equations to conclude

   $$x - z = (x - y) + (y - z) = p^m r + p^n s.$$

   If $m > n$, then $x - z = p^n(p^{m-n}r + s)$ and $d(x, z) = d(y, z)$. Similarly, if $n > m$, $d(x, z) = d(x, y)$. Finally if $n = m$, then

   $$x - z = p^n(r + s) = p^m(r + s),$$

   and $d(x, z) = d(x, y) = d(y, z)$. These three cases gives the desired inequality.

Definition 2.2. Let X be a metric space. A set $E \subseteq X$ is *bounded* if there is a positive number $M \in \mathbb{R}$ and a point $x \in X$ such that $d(x, y) < M$ for all $x \in E$. If a set is no bounded, we say it is *unbounded*.

Boundedness insures that a set doesn't "go off to infinity".

Example 2.5. The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are all unbounded.

[15] I admittedly am not certain of when it is best to introduce the concept of a norm. I don't like how Rudin (1976) talks about it in passing in when reviewing Euclidean space. Introducing it latter on when covering functional analysis is also problematic, because we're going to use the sup-norm before that to measure the distance between two functions.

Example 2.6. Both the intervals $[a, b]$ and (a, b) are bounded in $\mathbb{R}$. For any $x, y \in [a, b]$, $d(x, y) < d(a, b)+1$. The same holds for (a, b).

The metric space we are most interested in is of course $\mathbb{R}^n$ equipped with the familiar Euclidean metric. We can use this metric to define the notion of an open or closed ball in $\mathbb{R}^n$.

Definition 2.3. If $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$, the *open ball* with center $\mathbf{x}$ and radius r is defined as

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \mid |\mathbf{y} - \mathbf{x}| < r\}.$$

Definition 2.4. If $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$, the *closed ball* with center $\mathbf{x}$ and radius r is defined as

$$\bar{B}_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \mid |\mathbf{y} - \mathbf{x}| \leq r\}.$$

Open and closed balls in $\mathbb{R}^n$ are a generalization of the open and closed intervals you were first introduced to in high school, and Figure 9 provides an illustration in $\mathbb{R}^2$.

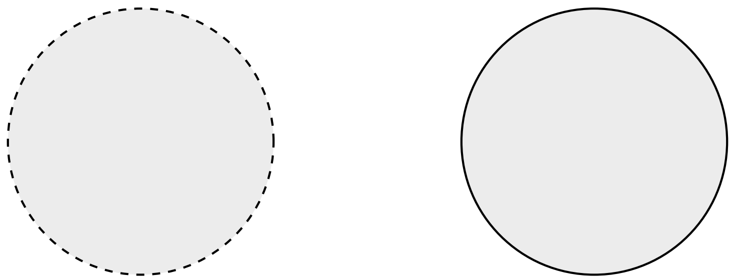

Figure 9: Open and closed balls in $\mathbb{R}^2$.

2.2 Open Sets, Closed Sets, and Boundaries

We want to generalize the notion of open and closed balls in $\mathbb{R}^n$ to any metric space. In order to do these, we'll need to outline a couple preliminary definitions that classify the elements of a metric space.

Definition 2.5. A *neighborhood* of x in a metric space X is defined as $N_r(x) = \{y \in X \mid d(x, y) < r\}$ for a radius $r > 0$.

As the name implies, a neighborhood centered at x is simply all the points "around" x. A neighborhood is its own set, and we will use them constantly. They are sort of like "sets of utility", because we will use them as tools to analyze the properties of other sets. If there is a set E in a metric space X, we can use neighborhoods in X to learn about the points in E. Are some neighborhoods subsets of E? Do some neighborhoods intersect E? Will the answers to these questions change if we make r really big or really small? It is worth thinking about these questions taking E to be one of the balls defined in Definition 3.3. Do the answers change if E is an open ball versus a closed ball?

Much like the open ball of Definition 3.2, neighborhoods do not include the points that are exactly a distance of r away from the point x. In fact, if our metric space is $\mathbb{R}^n$ with the standard metric, a neighborhood and open ball are precisely the same.

Example 2.7. Let our metric space be $\mathbb{Z}^2$ equipped with the taxi-cab metric. Figure 10 shows the neighborhood centered at the origin of radius 3.

Example 2.8. Let $d_{\ell^2} : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty]$ be the euclidean metric, and $d_{\ell^1} : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty]$ be the taxi-cab metric. A neighborhood in $(\mathbb{R}^2, d_{\ell^1})$ may have a different "shape" then it would in $(\mathbb{R}^2, d_{\ell^2})$. We will denote $N_3(\mathbf{0}) \in \mathbb{R}^2$ as E and F, in (X, d_{ℓ^2}) and (X, d_{ℓ^1}) respectively. These neighborhoods are shown in Figure 11.

Definition 2.6. Let X be a metric space. A point $x \in X$ is a *limit point* of the set $E \subseteq X$ if *every* neighborhood of x contains a point $y \in E$, where $y \neq x$. We will denote the *set of all limit points* of E as $E' = \{x \in X \mid x \text{ is a limit point of } E\} = \{x \in X \mid\in (N_r(x) \cap E) \backslash \{x\} \neq \emptyset \;\forall r > 0\}$.

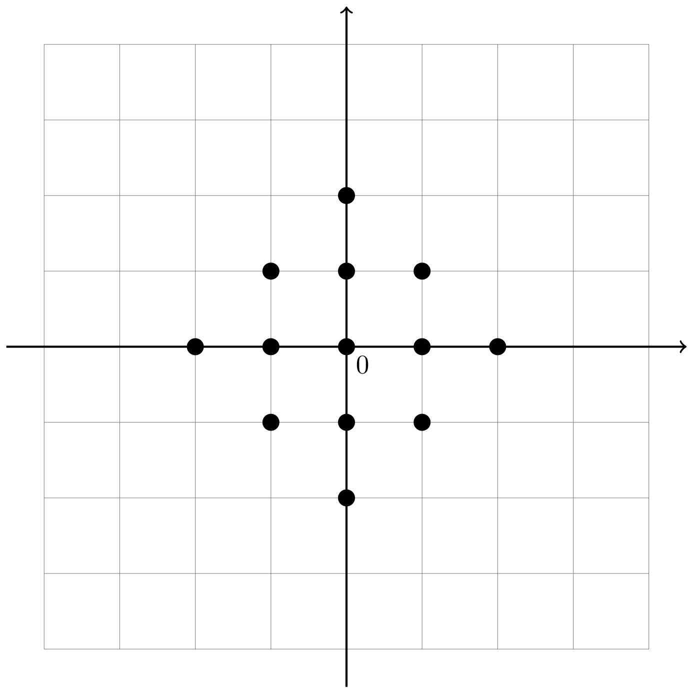

Figure 10: The set $N_3(\mathbf{0}) = \{\mathbf{y} \in \mathbb{Z}^2 \mid |y_1| + |y_2| < 3\}$.

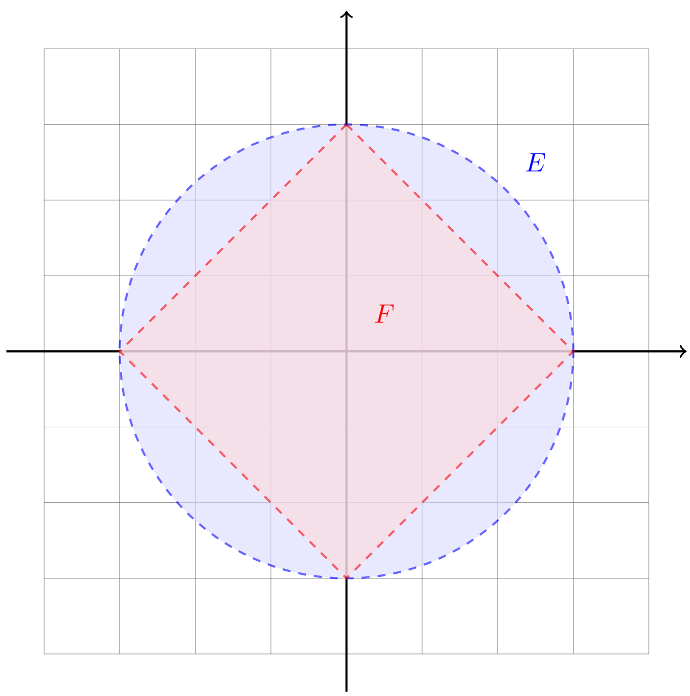

Figure 11: $N_3(\mathbf{0}) \in \mathbb{R}^2$ in $(\mathbb{R}^2, d_{\ell^2})$ and $(\mathbb{R}^2, d_{\ell^1})$.

Notation 2.1. For the remainder of Section 2, we will use X to denote a metric space, and E as some subset of X.

A limit point of a set is in some sense always "close" to points of the set. If x is a limit point of $E \subseteq X$, then $N_r(x)$ will always include points other than x, no matter what we take r to be! We could make r smaller and smaller, but the set $N_r(x)$ will never just be x. In this sense, a limit point can always be "approximated" by elements in E.

Remark 2.1. Definition 3.5 never specifically said that a limit point of some set belongs to the set. As the next example shows, being a limit point has nothing to do with whether or not a point is included in the set in question.

Example 2.9. $\mathbb{R}^2$ is a good starting place. Suppose we have a set $E \subseteq \mathbb{R}$ that for the most part forms a rectangle. The "border" of the rectangle is no included in E. Also note that E includes an "isolated" point z (see Figure 12). Let's consider three points in $\mathbb{R}^2$: x, y, and z.

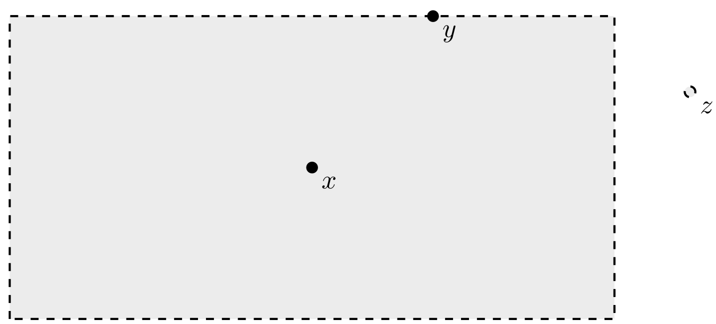

Figure 12: The set $E \subseteq \mathbb{R}^2$.

The point x belongs to E. Furthermore, no matter what we take r to be, $N_r(x)$ will never become a singleton of just $\{x\}$. For the sake or argument, suppose $x = (2, 2)$. If $r = 0.5$, then $(2, 2.49) \in N_{0.5}(x)$. Of $r = 0.01$, then we still have $(2, 2.001) \in N_{0.01}(x)$. In fact, for every r, we have $(2, 2 + r/2) \in N_r(x)$. This means that x is a limit point of E.

Now consider y. This point does not belong to E, but it is still a limit point! We could repeat the same argument we made for x without running into trouble, because every neighborhood of y will include points "just below" y, all of which are in E! What matters with limit points is not what set the point belongs to, but what set the points nearby it belong to.

Lastly, the point z is not a limit point. If we took r to be sufficiently large, then $N_r(z)$ would include points in E that form the rectangle. Unfortunately, we could easily take r to be so small that $N_r(z) = \{z\}$. It only takes one such r to rule out the chance of z being a limit point. We can provide a definition that corresponds to points like z.

Definition 2.7. Let X be a metric space. For a set $E \subseteq X$, $x \in E$ is an *isolated point* if it is not a limit point. That is, there exists an $r > 0$ such that $N_r(x) = \{x\}$.

By the definition of an isolated point, it is the opposite of a limit point, rendering the two definitions mutually exclusive. This definition also means any point $x \in X$ is *either* a limit point *or* an isolated point. An isolated point of any set is also included in the set, which is not the case for limit points.

Definition 2.8. Let X be a metric space. A point $x \in X$ is an *interior point* of $E \subseteq X$ if there exists *a single* $r > 0$ such that $N_r(x) \subseteq E$.

Example 2.10. Again, let's look at an example in $\mathbb{R}^2$. Let $E \subseteq \mathbb{R}^2$ be a closed ball that is "punctured" at $z \in \mathbb{R}^2$ such that $z \notin E$. This can be seen in figure 13. The point x is in an interior point, as we could find some small r for which $N_r(x) \subseteq E$. The point y is not an interior point of E, because every single $N_r(y)$ will contain some point outside of E, meaning $N_r(y) \nsubseteq E$. Finally, the point z is not an interior point, as each neighborhood of $N_r(z)$ contains z, and $z \notin E$. Even though we can make r small enough to guarantee the only point in $N_r(z)$ which is not in E is z ($N_r(z) \backslash E = \{z\}$), this point is all it takes to guarantee $N_r(z) \nsubseteq E$ for all r.

Remark 2.2. While a limit point of $E \subseteq X$ need not be a point in E, an interior point of E must be an element of E. If x is an interior point, then $x \in N_r(x) \subseteq E$ for some r, so $x \in E$.

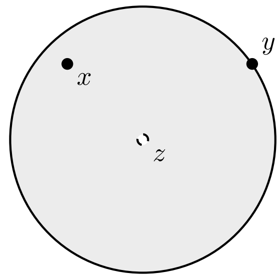

Figure 13: The set $E \subseteq \mathbb{R}^2$.

Example 2.11. It may be tempting to conclude that an interior point must be a limit point, after all, if we can find an $N_r(x) \subseteq E$, then it is likely each neighborhood would contain infinite points of X. This logic makes the dangerous assumption that X is in infinite, and $d(x, y)$ "behaves like" the Euclidean metric. Consider a metric space X with the discrete metric

$$d(x,y) = \begin{cases} 1 \text{ if } x \neq y \\ 0 \text{ if } x = y \end{cases}.$$

For any $x \in X$, $N_{1/2}(x) = \{x\} \subseteq X$. We have that x is an interior point, but not a limit point.

We now briefly introduce the idea of an exterior point. The only difference between the definition of an interior point and exterior point, will be that the neighborhood of a point will be in the complement of E for an exterior point. This small change in language will make a big difference in meaning.

Definition 2.9. Let X be a metric space. A point $x \in X$ is an *exterior point* of $E \subseteq X$ if there exists *a single* $r > 0$ such that $N_r(x) \subseteq E^c$.

Remark 2.3. Any point $x \in X$ is *either* an interior point *or* an exterior point.

Example 2.12. Let $[0, 1] \in \mathbb{R}$. The point $2 \in \mathbb{R}$ is an exterior point of $[0, 1]$.

Remark 2.4 (VERY IMPORTANT THEME)**.** Nearly every definition in this section specifies a metric space X. This means the metric space we work in could affect how we classify a point (and later sets). If we have two metric spaces X and Y where $X \subseteq Y$, a point $x \in E \subseteq X$ may be a limit point/interior point/exterior point in X but not in Y.

We will see this come up again, and again. How a set/point behaves or is classified is contingent on what space we are in. A small change, whether it be the inclusion of some additional points, or changing the metric, can make a big difference. This means it is important to specify what space we're in if it is ever unclear. On the bright side, this all makes for great examples!

Example 2.13. The set $\mathbb{R}$ has no limit points when equipped with the discrete metric

$$d(x,y) = \begin{cases} 1 \text{ if } x \neq y \\ 0 \text{ if } x = y \end{cases}.$$

We already generalized the idea of some open or closed interval in $\mathbb{R}$ to the concept of an open or closed ball in $\mathbb{R}^n$. Now we will go one step further, by bringing these concepts to any metric space.

Definition 2.10. Let X be a metric space. A set $E \subseteq X$ is *open* if every point of E is an interior point.

Definition 2.11. Let X be a metric space. A set $E \subseteq X$ is *closed* if it contains all its limit points. That is, $E' \subseteq E$.

Example 2.14. Let $(a, b) \subseteq \mathbb{R}$. This set is open, as for all $x \in (a, b)$, we can find an r such that $N_r(x) \subseteq (a, b)$. If $d(a, x) \geq d(x, b)$, let $r = d(x, b)/2$. If $d(x, b) > d(a, x)$ let $r = d(x, a)/2$. Figure 14 shows this neighborhood for x where $d(a, x) \geq d(b, x)$. By construction, our neighborhood will always be a proper subset of (a, b), so each point of x is an interior point. Therefore (a, b) is open. On the other hand, (a, b) is not closed, because a and b are limit points, but neither are in the set (a, b).

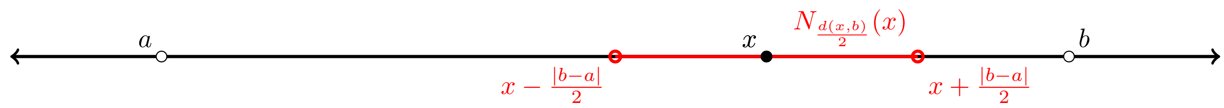

Figure 14: The open interval $(a, b) \subseteq \mathbb{R}$.

Example 2.15. Let $[a, b] \subseteq \mathbb{R}$. This set is not open, as a and b are not interior points. For instance, $a - r/2 \in N_r(a)$ for all $r > 0$. The number $a - r/2 \notin [a, b]$, so $N_r(a) \not\subseteq [a, b]$ for all r. While $[a, b]$ is not open, it is closed. Every point in $[a, b]$ is a limit point because $\mathbb{R}$ is complete. The interval $[a, b]$ trivially contains itself, so it contains all its limit points and is closed.

Example 2.16. Let X be a metric space, and $E \subseteq X$. Any set with no limit points, $E' = \emptyset$, is closed, because $\emptyset \subseteq E$. This means any finite set is closed, as no finite set has limit points. Let $X = \{x_1, \dots, x_n\}$, and $E \subseteq X$. If we let $y \in E$, and set $r = \min_{x \in X} d(x, y)$, then $N_r(y) = \{y\}$. This means y fails to be a limit point for all $y \in E$.

Remark 2.5. The definitions of open and closed sets never imply that a set is either closed or open. It is possible for a set to be both closed and open, or be neither closed nor open.

Example 2.17. The set $\emptyset$ in any metric space X is closed and open. This set has no limit points ($\emptyset' = \emptyset$), and $\emptyset \subseteq \emptyset$, so it is closed. The set also has no points, so every point is an interior point, making $\emptyset$ open.

Example 2.18. The set of rationals $\mathbb{Q}$ is neither open nor closed in $\mathbb{R}$. For all $x \in \mathbb{Q}$, $N_r(x)$ will contain irrational numbers for all r, meaning $N_r(x) \not\subseteq \mathbb{Q}$. Therefore no elements of $\mathbb{Q}$ are interior points. The set $\mathbb{Q}$ also does not contain all of its limit points, as any irrational number is a limit point as a result of Theorem 2.4. For example, any neighborhood of $\sqrt{2}$ will contain elements of $\mathbb{Q}$ (which are not $\sqrt{2}$), making it a limit point.

Remark 2.6. The point made in Remark 3.3 is especially relevant for open and closed sets. A set could be open in one metric space, but closed in another. In most cases it's clear what metric space we are working in, but sometimes it is not. In cases where it is vague, it's always best to say a set is *open in* X or *closed in* X. For this reason it is a good practice to either specify $E \subseteq X$, or include "in X." Many topics in analysis concern the metric space $\mathbb{R}^n$ or $\mathbb{R}$, so if you say a set is closed or open in conversation, it is usually assumed the metric space is one of these spaces. For example, if you were to ask someone "are the integers closed or open?", they would most likely assume you mean "are the integers closed or open in $\mathbb{R}$?"

Example 2.19. Suppose we want to determine if $\mathbb{Z}$ is open or closed in $\mathbb{Z}$. Every point is an interior point as $N_{1/2}(x) = x \subseteq \mathbb{Z}$ for all $x \in \mathbb{Z}$, so $\mathbb{Z}$ is open in $\mathbb{Z}$. The set $\mathbb{Z}$ has no limit points in $\mathbb{Z}$, as $N_{1/2}(x)$ does not include any points $y \in \mathbb{Z}$ where $y \neq x$. This gives that $\mathbb{Z}' = \emptyset$,[16] so $\mathbb{Z} \subseteq \mathbb{Z}'$, and $\mathbb{Z}$ is closed in $\mathbb{Z}$.

Now let our metric space be $\mathbb{R}$. Is $\mathbb{Z}$ open in $\mathbb{R}$? Let $x \in \mathbb{Z}$. For any $N_r(x)$ such that $r < 1$, $x - r/2 \in N_r(x)$, where $x - r/2 \notin \mathbb{Z}$. If $r \geq 1$, then $x - 1/2 \in N_r(x)$, where $x - 1/2 \notin \mathbb{Z}$.[17] Therefore, there exists no r such that $N_r(x) \subseteq \mathbb{Z}$, so $\mathbb{Z}$ is not open in $\mathbb{R}$. We have that $\mathbb{Z}$ is closed in $\mathbb{R}$, as each point of $\mathbb{Z}$ is still isolated.

Before proving some useful properties of open and closed sets, there is one more definition that can prove helpful at times. It formalizes the notion of points in a set that are just on the border of a set, like the endpoints of $[a, b] \subseteq \mathbb{R}$.

Definition 2.12. Let X be a metric space, and $E \subseteq X$. The *boundary* of E, denoted ∂E, is the set of points in X such that every neighborhood of p contains at least one point of E and at least one point not of E.

$$\partial E = \{x \in X \mid N_r(x) \cap E \neq \emptyset \text{ and } N_r(x) \cap E^c \neq \emptyset \ \forall r > 0\}$$

Any element of ∂E is a *boundary point*.

[16] This also means that every point of $\mathbb{Z}$ is isolated.

[17] The case where $r \geq 1$ handles the situation where $r/2 \in \mathbb{Z}$. If this were the case, then $x - r/2 \in \mathbb{Z}$. This is not a problem, as $N_r(x)$ would still contain an uncountably infinite number of reals numbers, but it makes explicitly finding one of those reals a little tricky. It's easier to just add or subtract $1/2$ from x and call it a day.

There are several equivalent definitions of ∂E, many of which are more popular than this specific one. These other definitions use terms that we will cover in Section 3.4, so we will circle back then and discuss the boundary of a again. The first example one's mind should jump to are open and closed balls in $\mathbb{R}^n$.

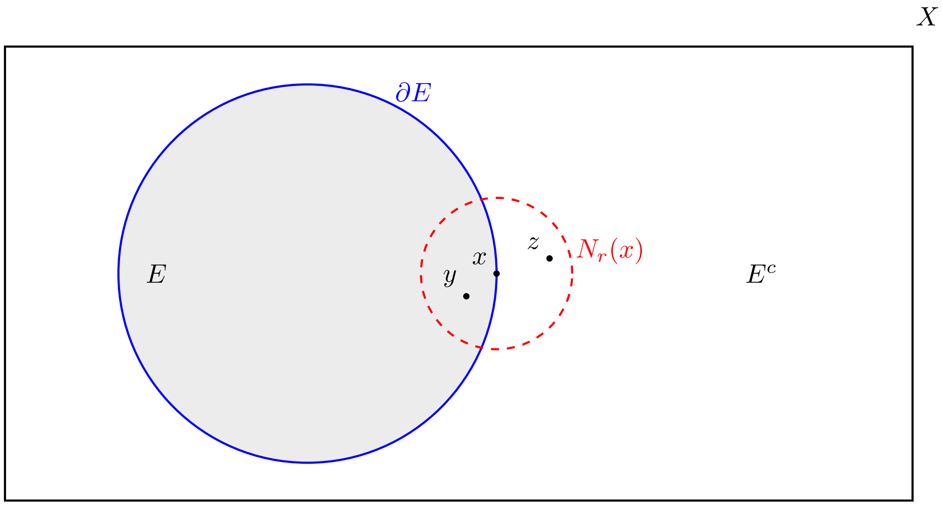

Figure 15: The boundary of a set E is shown in blue. Let $\mathbf{x} \in X$. No matter how small we make r, $N_r(x)$ will always contain some $y \in E$ and some $z \in E^c$, so $x \in \partial E$.

Example 2.20. Let $B_r(\mathbf{x})$ be the open ball of radius r centered at $x \in \mathbb{R}^n$ (we could also denote this as $N_r(\mathbf{x})$). As the name suggests, the boundary is just all the points that are exactly a distance of r away from $\mathbf{x}$, meaning $\partial B_r(\mathbf{x}) = \{\mathbf{y} \in X \mid |\mathbf{x}-\mathbf{y}| = r\}$. In this particular case, no points in the boundary are in $B_r(\mathbf{x})$, so we have $B_r(\mathbf{x}) \cap \partial B_r(\mathbf{x}) = \emptyset$. If we take $\bar{B}_r(\mathbf{x})$ to be the closed ball, then we have the same boundary.

$$\partial \bar{B}_r(\mathbf{x}) = \partial B_r(\mathbf{x}) = \{\mathbf{y} \in X \mid |\mathbf{x} - \mathbf{y}| = r\}$$

We also have $\partial \bar{B}_r(\mathbf{x}) \cap \bar{B}_r(\mathbf{x}) = \bar{B}_r(\mathbf{x})$, and $B_r(\mathbf{x}) \cup \partial B_r(\mathbf{x}) = \bar{B}_r(\mathbf{x})$.

Remark 2.7. It is very tempting to think all boundary points are limit points. At first glance, the definition of a boundary point seems to imply a point $x \in \partial E$ is not only a limit point of E, but also a limit point of E^c. This is not true! Suppose $x \in E$ is a limit point. The definition of a limit point not only requires that $N_r(x) \cap E$ for all r, but also requires that there are points *other than* x in $N_r(x)$. A boundary point needn't satisfy this second requirement, so even if $N_r(x) \cap E = \{x\}$ for all r, x can still be a boundary point!

Example 2.21. Let $E \subseteq \mathbb{R}^2$ be the union of a disk punctured at z and an isolated point y (Figure 16).

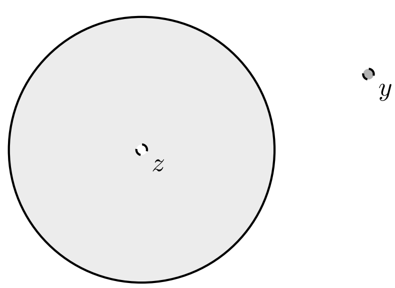

Figure 16: The set $E \subseteq \mathbb{R}^2$.

The points y and z are both boundary points. Despite this, y is not a limit point of E, and z is not a limit point of E^c. This follows from the reasoning in the previous remark.

Remark 2.8. We have now introduced five different definitions that classify points: limit points, isolated points, interior points, boundary points, and exterior points. This is *a lot* to take in all at once. By far the most important concepts introduced here were open and closed sets. Being able to determine if a set is open and/or closed is one of the most important skills to have for this section, and those that follow.

2.3 Properties of Open and Closed Sets

Open set and closed sets will play a roll in many of the results and theorems to come, so it is important to be able to identify which sets are open and which are closed. We will know introduce several tools that make this easier.

Proposition 2.1. Every neighborhood is an open set.

Proof. Suppose we have a metric space X, and some point $x \in X$. We will show that any point $y \in N_r(x)$ is an interior point. There exists some h such that

$$d(x, y) = r - h.$$

I claim that $N_{r'}(y) \subseteq E$ for $r' < h$. For all points z such that $d(y, z) = r' < h$, the triangle inequality gives

$$d(x, z) \leq d(x, y) + d(y, z) < r - h + h = r,$$

so $z \in N_r(x) = E$ for all z by the definition of $N_r(x)$. This implies that $N_{r'}(y) \subseteq E$, making y an interior point. (Figure 17) $\square$

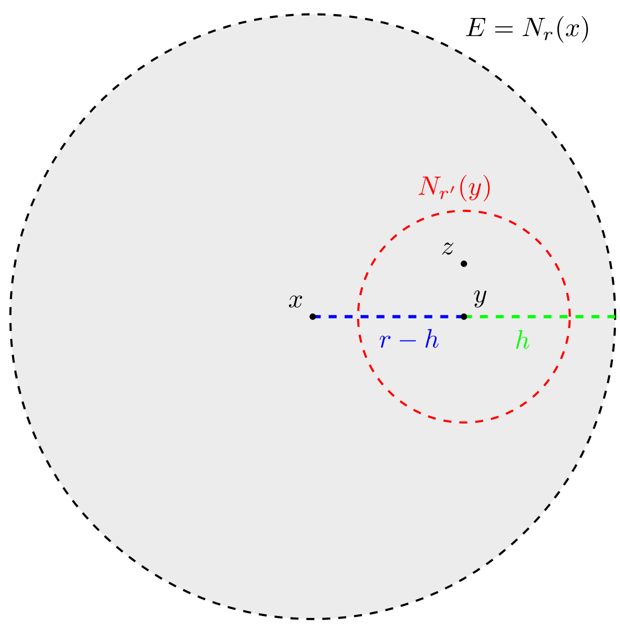

Figure 17: The set $E \subseteq X$. We constructed a neighborhood $N_{r'}(y)$ for an arbitrary $y \in N_r(x)$ such that $N_{r'}(y) \subseteq E$.

Proposition 2.2. If $x \in X$ is a limit point of E ($x \in E'$), then every neighborhood of x contains infinitely many points of E.

Proof. Let $x \in X$. Suppose for contradiction, there exists some $N_r(x)$ which contains only a finite number of points of E. Let this finite set of points be $\{y_1, \ldots, y_n\} \subseteq N_r(x) \cap E$. Pick the radius of $N_r(x)$ to be the distance between x and the point to which it is closest in the finite set $\{y_1, \ldots, y_n\}$:

$$r = \min_{1 \leq m \leq n} d(x, y_m).$$

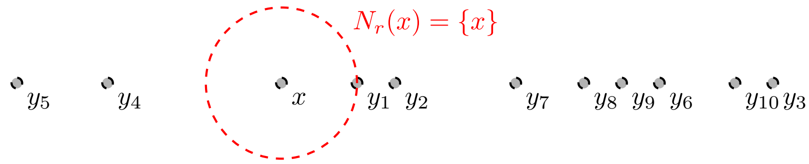

Figure 18: If our finite set of points of E is $\{y_1, \ldots, y_{10}\}$, then we can reach a contradiction by constructing a neighborhood of x with $r = \min_{1 \leq m \leq n} d(x, y_m) = d(x, y_1)$. This will hold no matter where the points $\{y_1, \ldots, y_{10}\}$ happen to be in E. In this case $x \in E$, but remember that this isn't a requirement.

By construction, $N_r(x)$ contains no point $y \in E$ such that $y \neq x$, so x is not a limit point of E. This is a contradiction. (Figure 18) $\square$

Corollary 2.1. A finite set has no limit points. (see Example 3.13)

The next theorem and its corollary allows us to determine if a set is open or closed based on its complement. At first, this may not seem helpful, but if it is not clear if E is open or closed, one can just use E^c! We will provide examples where using E^c is easier.

Theorem 2.1. A set E is open *if and only if* its complement is closed.

Proof.

($\Longrightarrow$) Suppose E is open. Let $x \in X$ be a limit point of E^c. Every neighborhood $N_r(x)$ contains a point of E^c, so $N_r(x) \nsubseteq E$, meaning x is *not* and interior point of E. But we have assumed every point of E is an interior point, so $x \in E^c$. Therefore E^c includes all its limit points and is closed.

($\Longleftarrow$) Suppose E^c is closed. Let $x \in E$. We have $x \notin E^c$, so x is not a limit point of E^c (otherwise it would be in E^c, as E^c is closed). If x is not a limit point of, then there exists an $N_r(x) \cap E^c = \emptyset$, giving $N_r(x) \subseteq E$. Thus $x \in E$ is an interior point, and E is open.

$\square$

Corollary 2.2. A set E is closed *if and only if* its complement is open.

One practical consequence of these results, is that if you find it more difficult to check if a set is open or closed (or vice versa), you can always just work with the complement.

Example 2.22. Let X be any metric space. Recall from Example 3.14 that $\emptyset$ is closed and open. This means that $\emptyset^c = X$ is closed and open as well. This allows us to conclude that $\mathbb{R}$ in $\mathbb{R}$ is open and closed.

Example 2.23. The set $[a, b] \subseteq \mathbb{R}$ is closed. This implies that $[a, b]^c = (-\infty, a) \cup (b, \infty)$ is open.

We often define some set of interest as a union or intersection of a collection of sets. For instance, the proof of Theorem 1.2, the supremum of a set in $\mathbb{R}$ was defined as the union of the Dedekind cuts that comprise the set. In Example 1.26, $\mathbb{Q}$ was written as the countably infinite union of intervals. In situations like this, it is possible to know if a set is open or closed if the sets over which we take the union/intersection are open or closed.

Theorem 2.2. Let $\{G_\alpha\}$ and $\{F_\alpha\}$ be an arbitrary collection of open sets and closed sets respectively. Let $G_1, \ldots, G_n$ and $F_1, \ldots, F_n$ be a finite collection of open sets and closed sets respectively. In this case we have:

1. $\bigcup_\alpha G_\alpha$ is open.
2. $\bigcap_\alpha F_\alpha$ is closed.
3. $\bigcap_{i=1}^n G_i$ is open.
4. $\bigcup_{i=1}^n F_i$ is closed.

Proof.

1. Suppose G_α is open for all α. Let $x \in \bigcup_\alpha G_\alpha$. For some α, $x \in G_\alpha$. There exists a neighborhood $N_r(x)$ such that $N_r(x) \subseteq G_\alpha$, because G_α is open. Therefore $\bigcup_\alpha G_\alpha$ is open, as $N_r(x) \subseteq G_\alpha \subseteq \bigcup_\alpha G_\alpha$.

2. Suppose F_α is closed for all α. It suffices to show that $(\bigcap_\alpha F_\alpha)^c$ is open using Theorem 2.1. By the aforementioned theorem, F_α^c is closed for all α. Therefore the union of F_α^c is open by part (1). This completes our proof, as De Morgan's Law gives

$$\left(\bigcap_\alpha F_\alpha\right)^c = \bigcap_\alpha F_\alpha^c.$$

3. Suppose the sets $G_1, \ldots, G_n$ are open. Let $x \in \bigcap_{i=1}^n G_i$. For all $x \in \bigcap_{i=1}^n G_i$, there exists neighborhoods $N_{r_i}(x)$ with radii r_i, such that $N_{r_i}(x) \subseteq G_i$ for all i. Let $r = \min\{r_1, \ldots, r_n\}$. This radius gives us $N_r(x) \subseteq G_i$ for all i, meaning $N_r(x) \subseteq \bigcap_{i=1}^n G_i$. Therefore x is an interior point, and $\bigcap_{i=1}^n G_i$ is open. (Figure 19)

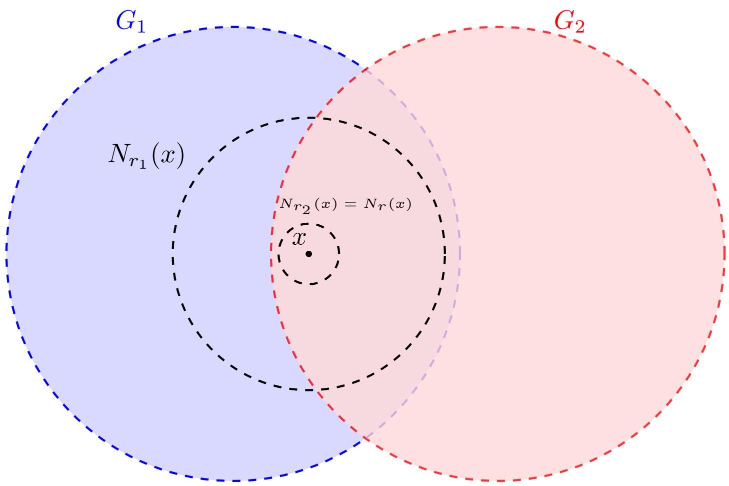

Figure 19: In this simplified setting, we have two open sets: G_1, and G_2. By letting $r = \min\{r_1, r_2\} = r_2$, we find a neighborhood $N_r(x)$ such that $N_r(x) \subseteq G_1 \cap G_2$.

4. Suppose the sets $F_1, \ldots, F_n$ are closed. It suffices to show that $(\bigcap_{i=1}^n F_i)^c$ is open using Theorem 2.1. By the aforementioned theorem, F_i^c is open for all i. By part (2) the the intersection of all F_i^c is open This complete out proof, as De Morgan's Law gives

$$\left(\bigcap_{i=1}^n F_i\right)^c = \bigcap_{i=1}^n F_n^c.$$

□

Part (2) and (4) of Theorem 2.2 require that the collection of sets is finite, otherwise the minimum over the finite number of radii of neighborhoods many not be well defined. The following two examples show that Theorem 2.2 does not hold if we take these collections to be infinite.

Example 2.24. Let $G_n = (1/n, 1 + 1/n)$ for all $n \in \mathbb{N}$. The set G_n is open in $\mathbb{R}$ for all n. Taking the intersection gives

$$\bigcap_n G_n = [0, 1].$$

The interval $[0, 1]$ is closed, despite each G_n being open.

Example 2.25. Let $F_n = [1/n, \infty)$ for all $n \in \mathbb{N}$. The set F_n is closed in $\mathbb{R}$ for all n. Taking the union gives

$$\bigcup_n F_n = (0, \infty).$$

This interval is open in $\mathbb{R}$, despited each F_n being closed.

2.4 Closures, Interiors, Dense Sets, and Perfect Sets

There are a handful of other definitions related to open set and closed sets that deserve a bit of attention.

Definition 2.13. Let X be a metric space. The *interior* of a set $E \subseteq X$, denoted E°, is the set of all interior points of E.

$$E^\circ = \{x \in X \mid x \text{ is an interior point of } E\}$$

The set E° is clearly open, as by definition it is comprised only of interior points. Because interior points of E must be in E, we have $E^\circ \subseteq E$. Informally, we can think of E° as the smallest open set contained within E. This interpretation leads to the conclusion that if E is open, then $E = E^\circ$.

Example 2.26. Let $[a, b] \subseteq \mathbb{R}$. The interior of this set is (a, b).

Definition 2.14. Let X be a metric space. The *closure* of a set $E \subseteq X$ is $\bar{E} = E \cup E'$.

The closure $\bar{E}$ is the opposite of the interior in a certain sense. The closure can be though of as the smallest closed set that E is contained in. If E is closed, then $E' \subseteq E$, and $E = \bar{E}$.

Example 2.27. Let $(a, b) \subseteq \mathbb{R}$. The closure of this set is $[a, b]$.

Definition 2.15. Let X be a metric space. The set $E \subseteq X$ is *dense in X* if every point of X is a limit point of E, or a point of E. ($\bar{E} = X$)

Informally, if E is dense in X, then we can approximate any point of X with a point in E arbitrarily well. This follow from the fact that any point in X is either in E, or a limit point of E', or both. We have already seen one example of this with Theorem 1.4.

Example 2.28. The set $\mathbb{Q}$ is dense in $\mathbb{R}$, that is $\bar{\mathbb{Q}} = \mathbb{R}$. One implication of this fact is that the irrational numbers $\mathbb{R} \backslash \mathbb{Q}$ are limit points of $\mathbb{Q}$. We also have that the irrationals $\mathbb{R} \backslash \mathbb{Q}$ are dense in $\mathbb{Q}$!

Many results involving approximation can be stated in terms of dense sets. One of these is the Weierstrass Approximation Theorem. This will be formally treated and proved in Section 7, but for now we will give the result as an example of a dense set.

Example 2.29 (Weierstrass Approximation Theorem)**.** Let $\mathscr{C}([a, b]) = \{f \mid f : [a, b] \to \mathbb{R} \text{ and } f \text{ continuous}\}$ be the set of real valued continuous functions with domain $[a, b]$. Now let $\mathscr{P}([a, b])$ be the set of all real valued polynomials with domain $[a, b]$.[18] The set $\mathscr{P}([a, b])$ is dense in $\mathscr{C}([a, b])$. We will always be able to approximate a continuous function on a bounded interval arbitrarily well with polynomials. This result, and spaces of functions, will be discussed again in Section 7.

Definition 2.16. Let X be a metric space. The set $E \subseteq X$ is *perfect* if every point of E is a limit point of E ($E = E'$).

If E is perfect, every point in E can be approximated arbitrarily well by other points in E.

Example 2.30. The real line $\mathbb{R}$ is a perfect set.

[18]This set is traditionally denoted as $\mathbb{R}[x]$ in abstract algebra.

2.5 Compact Sets

We have encountered infinity several times now. Sets can have an infinite number of element, in which case they are either countable or uncountable. A set can "take up an infinite amount of space" if it is unbounded. Each limit point of a set contains an infinite number of points in the set. These factors can result in sets that are difficult to work with. For example, suppose a set is unbounded. It can be hard to determine how functions or sequences behave on sets like this, because the distance between points can become arbitrarily large. What kind of headaches to limit points cause? Any neighborhood of a limit point will contain an infinite number of points in that set (Proposition 2.2). If the limit point is not in the set, this can also pose problems. As we will see later on, it could be possibly to get arbitrarily close to that limit point while never leaving the set. In a sense, we would be getting closer to a point in a set, where the destination is not even included in the set. To prevent this from happening, all limit points should be included in a set, i.e the set should be closed. Later on when working with sequences and continuous functions, two concepts intrinsically linked by the idea of getting arbitrarily close to a point, sets that are "nice", and will not illicit the two mentioned complications, will lead to nice results. Our goal now is to characterize these sets, and attempt to motivate their characterization.[19]

One somewhat trivial way to guarantee a set is both closed and bounded is by restricting out attention to finite sets. If E is finite, it has no limit points and is trivially closed. A finite set must also be bounded. While finite may have the nice properties we are looking for, they are not that interesting. Real analysis almost always involves the real numbers, an uncountably infinite set. So what criteria would guarentee an infinite set, whether it be countable or uncountable, will behave like a finite set?

The way we will go about defining these sets is by looking how we can "cover" them with a collection of open sets. The idea is that a set may be infinite, but perhaps we can cover it with a finite collection of open sets.

Definition 2.17. Let X be a metric space. An *open cover* of a set $E \subseteq X$ is a collection of open sets $\{G_\alpha\} \subseteq X$ such that $E \subseteq \cup_\alpha G_\alpha$.

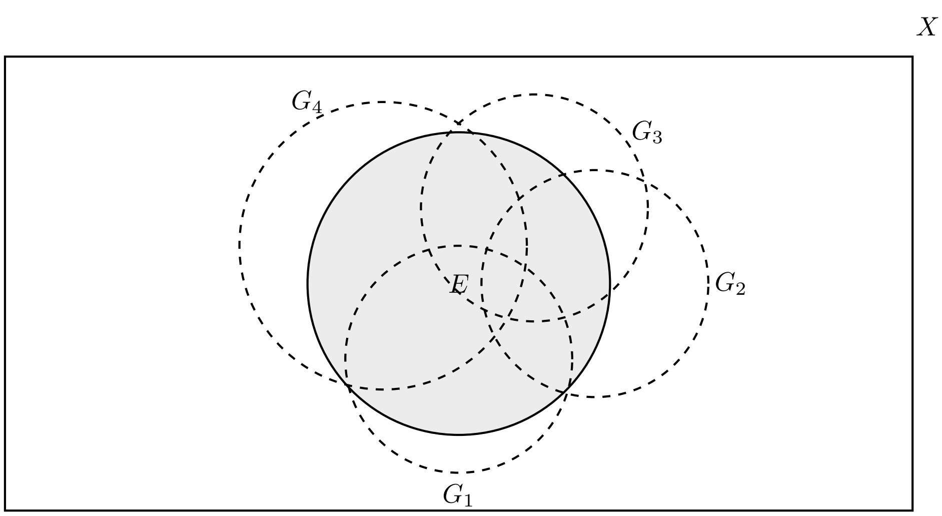

Figure 20: The collection of set $\{G_1, \ldots G_4\}$ forms an open cover of E.

An open cover is a *collection* of sets. Each element of an open cover is a single open set. This means that the cardinality of an open set has *nothing* to do with the cardinality of the sets it is comprised of. The collection $\{\mathbb{R}, (0,1), (0,2)\}$ is 3. We do not care whatsoever about the fact that each of the sets in the collection are uncountably infinite. The emphasis here is due to the fact that the cardinality of these covers (and a second type we will define soon) will be the bases of our criteria of what makes a set "nice".

[19]Motivating the main definition of this subsection is infamously difficult, as it is not clear how it will be used in the future. It would be like explaining what a hammer is to someone who has no idea what a nail is.

Example 2.31. Let $E = N_r(x)$ be a subset of a metric space X. One open cover of the set E is the single set $N_{r+1}(x)$. Another would be $N_{r+2}(x)$.

Example 2.32. Let $\mathbb{R}$ be the entire real line. We can cover this with the union of all sets of the form $G_n = (-n, n)$ for $n \in \mathbb{N}$.

$$\mathbb{R} \subseteq \bigcup_{n\in\mathbb{N}} (-n, n)$$

It is not enough to require that a set has a finite open cover. Any set E can trivially be covered by itself, forming an open cover consisting of one element. We could require that every open cover is finite, but this could never be satisfied. For example, take the closed interval $(a, b) \subseteq \mathbb{R}$. We could cover this with a finite open cover $\{(a, b)\}$. We could also cover it with the infinite open cover $\{(a, b), (-1, 1), (-2, 2), (-3, 3), \ldots\}$. We can just take the cover $\{(a, b)\}$ and throw in an infinite number of random intervals of $\mathbb{R}$ and still end up with an open cover. This is complete "overkill" when it comes to covering (a, b)! To address the fact that we will always have infinite open covers, we will introduce a new type of open cover that is both finite, and limits any redundant additions to the cover.

Definition 2.18. Let X be a metric space, and $E \subseteq X$. A *finite subcover* of an open cover $\{G_\alpha\}$ of E is a collection of open sets $\{G_{\alpha_1}, \ldots, G_{\alpha_n}\}$ such that

$$E \subseteq \bigcup_{i=1}^{n} G_{\alpha_i} \subseteq \bigcup_{\alpha} G_\alpha.$$

Example 2.33. Let X be a metric space and $E \subseteq X$. The collection $\{E\}$ is a trivial open cover. We also have a trivial finite open subcover in $\{E\}$.

Example 2.34. Let $\{(a, b), (-1, 1), (-2, 2), (-3, 3), \ldots\}$ be an open cover of $(a, b) \subseteq \mathbb{R}$. One finite subcover of this open cover is $\{(a, b)\}$. Another finite subcover is $\{(a, b), (-1, 1)\}$. In fact, any set $\{(a, b), (-1, 1), \ldots (-n, n)\}$ is a finite subcover.

This simple example shows that some open covers actually have an infinite number of finite subcovers. The introduction of finite subcovers may be a bit confusing, so it is worth recapping what we have done before presenting the main definition of this subsection:

- We have some metric space X and some set $E \subseteq X$. We can cover this set with a collection of sets $\{G_\alpha\}$ called an open cover, the cardinality of which is determined by the number of sets in the collection $\{G_\alpha\}$. We like the idea of finite open covers.
- There exists an infinite number of open covers of a set. There also always exists a finite open cover of a set, so we need to do better than just having a finite open cover.
- A finite subcover $\{G_{\alpha_1}, \ldots, G_{\alpha_n}\}$ of the open cover $\{G_\alpha\}$ is a subset of $\{G_\alpha\}$, which also covers E. The finite subcover of one open cover $\{G_\alpha\}$ *is not necessarily* a finite subcover of another open cover $\{G'_\alpha\}$, so whenever we talk about finite subcovers, it is with respect to some fixed open cover. Some open covers of sets have an infinite number of finite subcovers (Example 2.29).

We are now ready give a proper definition and name to the "nice" sets we want to characterize. In doing so, we will answer an important question about finite subcovers that may have arisen by now – some open covers of sets have an infinite number of finite subcovers, but do *all* open covers of a set have *at least one* finite subcover?

Definition 2.19. Let X be a metric space, and K be a subset of X. The set K is *compact* if *every* open cover of K contains *at least one* finite subcover.

The answer to our question turns out to be no. If it were yes, then there would be no need to define compactness, because every set would be compact. Compact sets turn out to be the nice sets we were looking for. As we'll see, they can be infinite, but they do not cause the complications with infinity that we discussed at the open of this subsection. In some sense, compact sets are the next best thing to finite sets.

Example 2.35. The set $(a, b) \subseteq \mathbb{R}$ is not compact. In order verify this, we just need to find a single open cover that has no finite subcover. Let $\{G_n\}$ be an open cover where $G_n = (a + 1/n, b)$. We have that

$$(a, b) \subseteq \bigcup_{n \in \mathbb{N}} G_n = \bigcup_{n \in \mathbb{N}} (a + 1/n, b) = (a, b).$$

No finite subset of $\{G_n\}$ will be a finite subcover of (a, b). If we had a finite subset of $\{G_n\}$ then there would exist some $N \in \mathbb{N}$ such that $(a, a + 1/N)$ is "uncovered". Therefore any open interval in $\mathbb{R}$ is not compact.

Example 2.36. The real line $\mathbb{R}$ is not compact. The open cover $\{G_n\}$ where $G_n = (-n, n)$ has no finite subcover. If we had a finite subset of $\{G_n\}$, then there would exist some $N \in \mathbb{N}$ such that $(-\infty, -N) \cup (N, \infty)$ is "uncovered". Therefore the real line $\mathbb{R}$ is not compact.

These two examples should not be entirely surprising. When compactness was motivated, complications involving two types of sets were cited: unbounded sets, and sets that were not closed. (a, b) and $\mathbb{R}$ both fall into exactly one of these categories.

Example 2.37. Suppose $X = \{x_1, \ldots, x_n\}$ is a finite metric space. The entire space X is compact. Let $\{G_\alpha\}$ be an open cover of X. For each $x_i \in X$, there exists an G_{α_i} such that $x_i \in G_{\alpha_i}$. Therefore, the open cover has a finite subcover in $\{G_{\alpha_1}, \ldots, G_{\alpha_n}\}$.

Using the definition of compactness to verify that a set is not compact takes some creativity, but only requires one to find a single counterexample. This is opposed to verifying a set is compact. This requires we somehow verify that every single open cover has a finite subcover. This is prohibitively difficult to do in most cases. This is why we want to find conditions that are easy to verify and that imply compactness. Of chief concern, is doing this for subsets of $\mathbb{R}^n$.

2.6 Properties of Compact Sets

Before restricting our attention to $\mathbb{R}^n$, we need to establish some properties of compact sets that will allow come in handy when working with them. Unfortunately, we still do not have any nontrivial examples of compact sets, so some of these results will not have examples presented alongside them. Once we are able to identify compact sets in $\mathbb{R}^n$ by means other then the definition of compactness, these results can be verified.

Lemma 2.1. Suppose $Y \subseteq X$. A subset $E \subseteq Y$ is open in Y *if and only if* $E = Y \cap G$ for some open $G \subseteq X$.

Proof.

($\Longrightarrow$) Suppose $E \subseteq Y$ is open in Y. For each $x \in E$, there exists a r_x such that $N_{r_x}(x) \subseteq E$. By the definition of a neighborhood, for all $y \in Y$ satisfying $d(x, y) < r_x$, we have $y \in E$. Denote $N_{r_x}(x) = V_x$ for all $x \in E$, and define

$$G = \bigcup_{x \in E} V_x.$$

The set G is a union of open sets, so it is open (Figure 21). I now claim that $E = G \cap Y$, which is our desired result. For $x \in E$, we have $x \in V_x$ for all $x \in X$, so $x \in E$ and $x \in Y$. This gives $E \subseteq G \cap Y$. Now let $x \in G \cap Y$. For the corresponding V_x, $V_x \cap Y \subseteq E$. This implies $G \cap Y \subseteq E$.

($\Longleftarrow$) Suppose $E = Y \cap G$ for some open G in X. The set G is open, so for all $x \in E$ there exists a neighborhood $N_r(x) \subseteq E$. This gives $N_r(x) \subseteq G$, as $E = Y \cap G \subseteq G$. Intersecting $N_r(x)$ with Y yields $N_r(x) \cap Y \subseteq E$, so E is open in Y.

□

Proposition 2.3. Suppose $K \subseteq Y \subseteq X$, where Y and X are metric spaces. The subset K is compact in X *if and only if* K is compact in Y.

Proof.

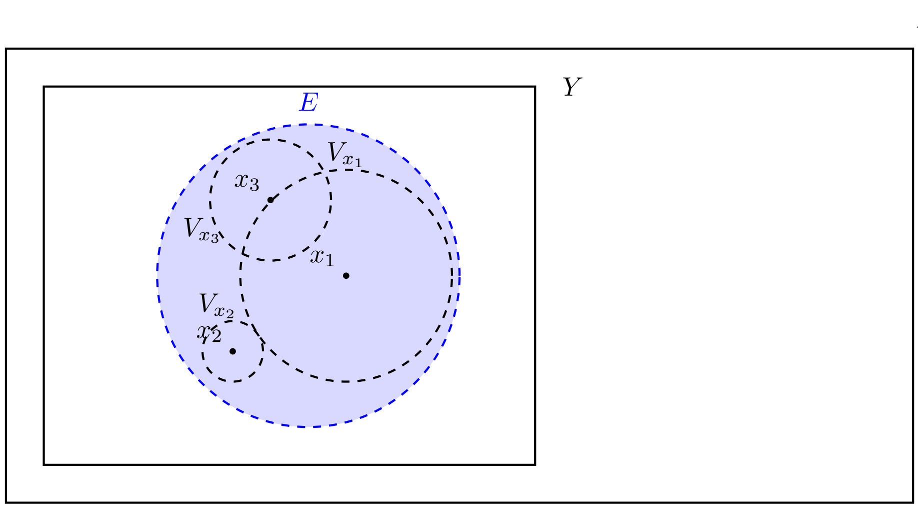

Figure 21: We take V_x to be the neighborhood of x contained in E. We will always be able to find such a neighborhood because E is open in Y. In this case, we have shown only three such neighborhoods. The (possibly infinite) union of all such neighborhoods is G, which itself is open.

($\Longrightarrow$) Suppose K is compact in X. Let $\{V_\alpha\}$ be an arbitrary collection of open sets in Y which cover K. By Lemma 2.1, there exist sets G_α, open in X, such that $V_\alpha = Y \cap G_\alpha$ for all α. By the compactness of K, there exists a finite subcover $\{G_{\alpha_1}, \dots, G_{\alpha_1}\}$. We have

$$K \subseteq \bigcup_{i=1}^{n} G_{\alpha_i},$$

but $K \subseteq Y$, so

$$K \subseteq \bigcup_{i=1}^{n} V_{\alpha_i}.$$

This makes $\{V_{\alpha_1}, \dots, V_{\alpha_1}\}$ a finite subcover, so K is compact in X.

($\Longleftarrow$) Suppose K is compact in Y. Let $\{G_\alpha\}$ be an open cover of K in X. If we let $V_\alpha = G_\alpha \cap Y$, then $\{V_\alpha\}$ is an open cover of K in Y. By the compactness of K we have a finite subcover $\{V_{\alpha_1}, \dots, V_{\alpha_n}\}$ in Y. Since $V_\alpha \subseteq G_\alpha$ for all α, then $\{G_\alpha\}$ has a finite subcover in $\{G_{\alpha_1}, \dots, G_{\alpha_n}\}$, so K is compact in X.

□

This result does not seem particularly interesting, but it is novel if you consider open and closed sets. We gave several examples of sets that may be open or closed in some intermediate space, but not a larger space. For instance, $\mathbb{Z}$ is open in $\mathbb{Z}$, but it is not open in $\mathbb{R}$. This result tells us that results like this are not possible with compactness!

Example 2.38. Suppose $E \subseteq \mathbb{Z}$ is compact in $\mathbb{Z}$. This implies that E is compact in $\mathbb{Q}$ and $\mathbb{R}$. If we had another compact set $F \subseteq \mathbb{R}$ which is compact in $\mathbb{R}$, then it is compact in $\mathbb{Z}$ and $\mathbb{Q}$ as well.

The next two theorems establish that all compact sets are both closed and bounded. This should feel somewhat natural, as compactness can be interpreted as a generalization of closed and bounded sets.

Theorem 2.3. Let X be a metric space, and $K \subseteq X$ be compact. The set K is closed.

Proof. Let K be a compact subset of a metric space X. It suffices to show that K^c is open in X. Suppose $x \in K^c$, and $y \in K$. For $r < \frac{1}{2}d(x, y)$, let $V_y = N_r(x)$ and $W_y = N_r(y)$ (Figure 24). For our fixed $x \in K^c$, we will repeat this process for multiple points in K. By compactness, we know there exists a finite

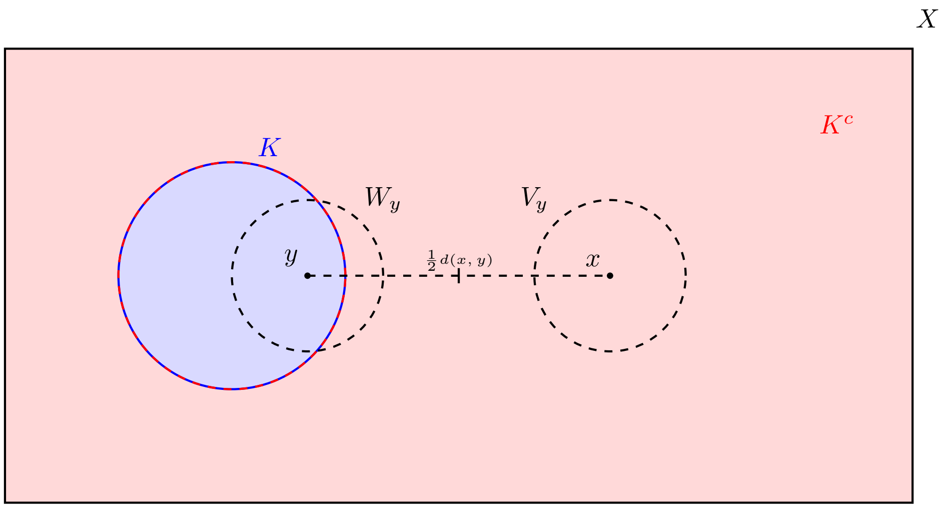

Figure 22: For some points $x \in K^c$ and $y \in K$, we construct neighborhoods $W_y = N_r(y)$ and $V_y = N_r(x)$ such that $r < \frac{1}{2}d(x,y)$. This choice of radius ensures $V_y \cap W_y = \emptyset$.

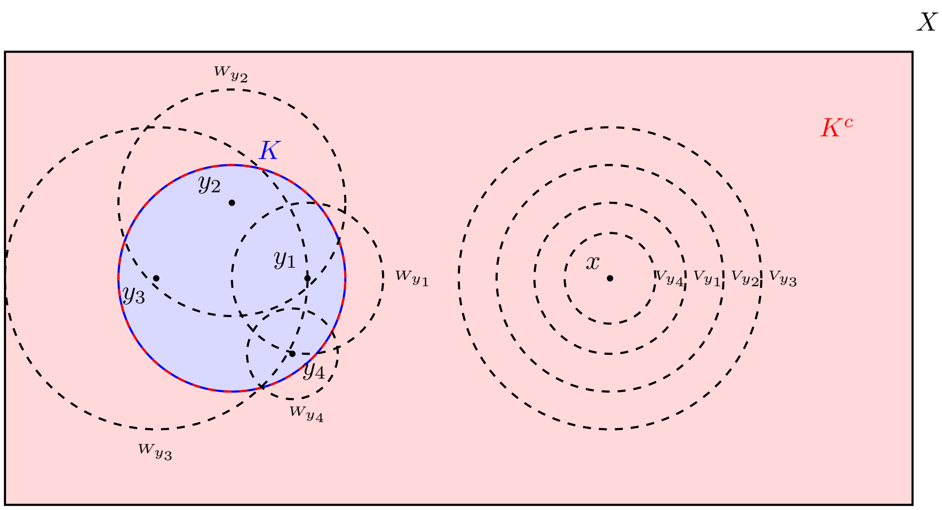

Figure 23: For the fixed value $x \in K^c$, repeat the process illustrated in Figure 21 until we have a finite subcover of K, $\{W_{y_1}, \ldots, W_{y_4}\}$. If we let V be the intersection of all V_{y_i}, then $x \in V \subseteq K^c$, rendering x an interior point of K^c.

set of $\{y_1, \ldots, y_n\}$ such that $\{W_{y_1}, \ldots, W_{y_n}\}$ is a finite subcover of K. In constructing this subcover, we also constructed the corresponding sets $\{V_{y_1}, \ldots, V_{y_n}\}$ (Figure 23). If we let $V = \cap_{i=1}^{n} V_{y_i}$, then we have $x \in V \subseteq K^c$, making x an interior point of K^c. Therefore K^c is open. $\square$

Example 2.39. The set $\mathbb{Q}$ in $\mathbb{R}$ is not closed (Example 2.18), so it is not compact in $\mathbb{R}$.

Theorem 2.4. Let X be a metric space, and $K \subseteq X$ be compact. The set K is bounded.

Proof. Let $x \in K$. The collection of neighborhoods $\{N_r(x)\}$ for $r \in \mathbb{N}$ forms an open cover of K. By compactness, this open cover has a finite subcover $\{N_{r_1}(x), \ldots, N_{r_n}\}$. If we take $r^* = \max\{r_1, \ldots, r_n\}$, then $K \subseteq N_{r^*}(x)$, and $d(x,y) < r^*$ for all $y \in K$. (Figure 24) $\square$

Example 2.40. The set $(0, \infty)$ in $\mathbb{R}$ is not bounded so it is not compact in $\mathbb{R}$.

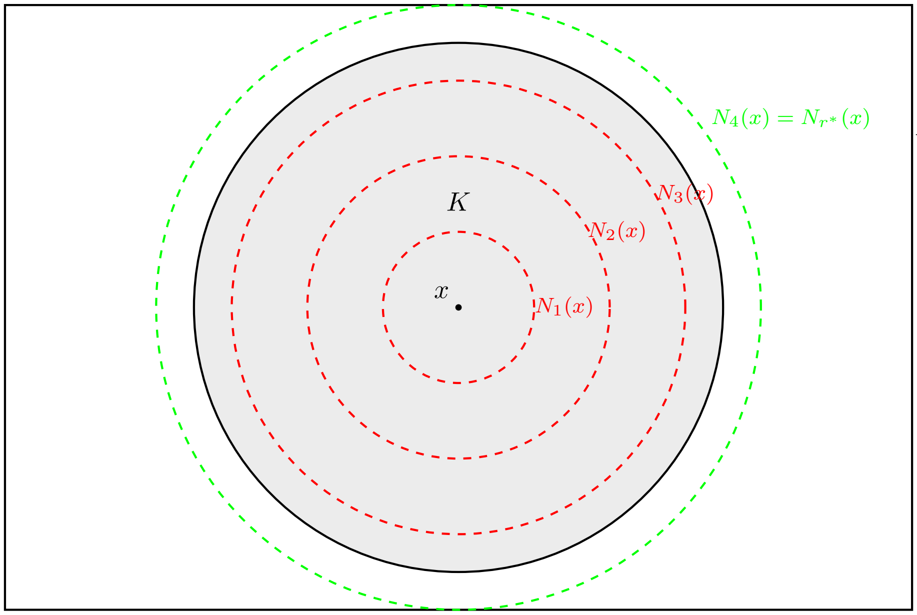

Figure 24: We cover the set K with an infinite open cover comprised of neighborhoods of radii in $\mathbb{N}$. By compactness there is an open subcover, such as $\{N_1(x), \ldots, N_4(x)\}$. The set is bounded by the maximum radii $r^* = 4$ in this finite collection.

Remark 2.9. While compactness implies closed and bounded, the converse *is not necessarily true*. Soon, we will see that the converse will hold in $\mathbb{R}^n$, but in general, this is not the case. The next example illustrates this.

Example 2.41. Let $X = \{1/n \mid n \in \mathbb{N}\}$ be a metric space equipped with

$$d(x, y) = \begin{cases} 0 \text{ if } x = y \\ 1 \text{ if } x \neq y \end{cases} .$$

The set X is closed, as it is the whole space. It is also bounded, as $d(x, y) \leq 1$ for all $x, y \in X$. Let $G_n = \{1/n\}$. Each G_n is open, as $N_{1/2}(1/n) \subseteq G_n$. This makes G_n an open cover of X, as $X \subseteq \bigcup_{n\in\mathbb{N}} G_n = X$. The set X fails to be compact, because this open cover has no finite subcover. Any finite cover $\{G_1, \ldots, G_N\}$ would not "cover" $\{1/(N+1), 1/(N+2), \ldots\} \subseteq X$.

We often restrict our attention to subsets of compact sets, so it would be nice to know if subsets of compact sets are compact. Unfortunately, this is not true in general, but it becomes true if we require the subset satisfy one condition.

Proposition 2.4. Closed subsets of compact sets are compact.

Proof. Suppose K in X is compact, and $F \subseteq K$ is closed in X. Let $\{V_\alpha\}$ be an arbitrary open cover of F. If we add F^c to this collection of open sets, we have an open cover $\Omega = \{V_{\alpha_1}, V_{\alpha_2}, \ldots, F^c\}$ of K, because $K = F \cup F^c$.

$$K \subseteq \left(\bigcup_\alpha V_\alpha \right) \cup F^c.$$

The set K is compact, so there exists a finite open cover of Ω, $\Phi = \{V_{\alpha_1}, \ldots, V_{\alpha_n}, F^c\}$. We can now remove F^c from Φ, resulting in a finite subcover for $\{V_\alpha\}$. □

Corollary 2.3. If F is closed and K is compact, $F \cap K$ is compact.

An interesting property of compact sets is that if we have a decrease sequence of nested compact intervals, than their intersection is nonempty. This result follows as a corollary of a more general result.

Proposition 2.5. If $\{K_\alpha\}$ is a collection of compact subsets of a metric space X such that the intersection of every finite subcollection of $\{K_\alpha\}$ is nonempty, then $\cap_\alpha K_\alpha \neq \emptyset$.

Proof. For the sake of contradiction, assume that $\cap_\alpha K_\alpha = \emptyset$. This means there is some fixed $K_1 \in \{K_\alpha\}$ such that no point of K_1 belongs to every K_α. Let $G_\alpha = K_\alpha^c$. The collection $\{G_\alpha\}$ forms an open cover of K_1. Since K_1 is compact, there exists a finite subcover $\{G_{\alpha_1}, \ldots, G_{\alpha_n}\}$.

$$K \subseteq \bigcup_{i=1}^{n} G_{\alpha_i}$$

This inclusion, along with the definition of G_α implies that $K_1 \cap \left(\bigcap_{i=1}^n K_{\alpha_i}\right) = \emptyset$. This contradicts our assumption that every finite subcollection of $\{K_\alpha\}$ is nonempty. □

Corollary 2.4 (Cantor's Intersection Theorem)**.** If $\{K_\alpha\}$ is a sequence of nonempty compact sets such $K_n \supset K_{n+1}$ for $n \in \mathbb{N}$, then $\cap_{i=1}^\infty K_n$ is not empty.

Example 2.42. Recall that open intervals in $\mathbb{R}$ are not compact (Example 2.35), and do not satisfy the requirement of Cantor's Intersection Theorem. Let $G_n = (0, 1/n)$. We have $G_{n+1} \subseteq G_n$ for all $n \in \mathbb{N}$, so $\{G_n\}$ is a decreasing nested sequence of intervals.

$$\cdots \subseteq \left(0, \frac{1}{4}\right) \subseteq \left(0, \frac{1}{3}\right) \subseteq \left(0, \frac{1}{4}\right) \subseteq \left(0, \frac{1}{2}\right) \subseteq (0, 1)$$

The intersection of these sets is empty.

Proposition 2.6 (Bolzano-Weierstrass Property)**.** If E is an infinite subset of a compact set K, then E has a limit point in K.

Proof. For the sake of contradiction, assume that no point of K is a limit point of E. For each $x \in K$, there exists some r such that $V_x = N_r(x) = \{x\}$ if $x \in E$, or $V_x = N_r(x) = \emptyset$ if $x \notin E$. The collection $\{V_x\}$ is an open cover of E. The set E is infinite, so we cannot find a finite subcover for $\{V_x\}$, as each set is at most a singleton. Because $E \subseteq K$, the same is true with respect to K, which contradicts the assumption that K is compact. □

Informally, the Bolzano-Weierstrass Property tells us that we can approximate some point of K with points in an infinite subset E. Right now, this may not seem significant, but it will become important when we work in $\mathbb{R}^n$.

Remark 2.10. The Bolzano-Weierstrass Property is often referred to as *limit point compactness*. In the context of metric spaces, limit point compactness and compactness are equivalent[20], so in most real analysis texts the prior is never even given a name. As we'll see *much* later on, the two are not equivalent when we explore point-set topology in general (Section 15).

2.7 Compact Sets in $\mathbb{R}^n$

Now we can start working towards sufficient conditions for compactness in $\mathbb{R}^n$. This will culminate in the famed Heine-Borel Theorem. This theorem establishes sufficient conditions for compactness in $\mathbb{R}^n$. Specifically, the converses of Theorem 2.3 and Theorem 2.4 will hold in $\mathbb{R}^n$.

In order to make the proof of this result a bit more clear, we will show a series of results beforehand which build to the Heine-Borel Theorem.

Lemma 2.2. If $\{I_n\}$ is a sequence of closed intervals in $\mathbb{R}$, such that $I_n \supset I_{n+1}$ for $n \in \mathbb{N}$, then $\cap_{i=1}^n I_n \neq \emptyset$.

[20]Proving this requires some concepts that are beyond this treatment of metric spaces.

Figure 25: The first three intervals in the type of sequence $\{I_n\}$ described in Proposition 2.7.

Proof. Suppose $I_n = [a_n, b_n]$. Define E to be the set of all the a_n. The set E is nonempty and bounded above by b_1, because

$$\cdots a_3 \leq a_2 \leq a_1 \leq b_1 \leq b_2 \leq b_3 \leq \cdots.$$

We have $E \subseteq \mathbb{R}$ and is bounded, so it has a supremum in $\mathbb{R}$. Let $x = \sup E$. For all $m, n \in \mathbb{N}$,

$$a_n \leq a_{m+n} \leq b_{m+n} \leq b_m,$$

so $x \leq b_m$ for all m. By the definition of $\sup E$, $a_m \leq x$. If $x \leq b_m$ and $a_m \leq x$, then $x \in [a_m, b_m] = I_m$ for all $m \in \mathbb{N}$. Therefore $x \in \cap_{n\in\mathbb{N}} I_n$. $\square$

Example 2.43. If we modify Example 2.42 so the open intervals are closed, then we have a sequence $\{I_n\}$ where $I_n = [0, 1/n]$. This gives $\cap_{n\in\mathbb{N}} I_n = \{0\} \neq \emptyset$. If we worked through the proof of Lemma 2.2 with this particular example, we would find that 0 is the supremum of the set of lower bounds of I_n.

This result should look familiar. Cantor's Intersection Theorem states a similar result for compact sets. This in and of itself does not show that closed intervals are compact, but it should catch our attention, as closed intervals and compact sets share a noteworthy property. It should also be noted that this property of closed intervals follows from the least-upper-bound property of $\mathbb{R}$. This is one of the magical results we get because $\mathbb{R}$ is complete. We can generalize Proposition 2.7 to *k-cells* in $\mathbb{R}^k$. An k-cell is the set of all points $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ which satisfy $a_i \leq x_i \leq b_i$ for $i = 1, \ldots, k$, where $a_i, b_i \in \mathbb{R}$, and $a_i \leq b_i$.

Lemma 2.3. Let $k \in \mathbb{N}$. If $\{I_n\}$ is a sequence of $k-$cells such that $I_n \supset I_{n+1}$ for all $n \in \mathbb{N}$, then $\cap_{n\in\mathbb{N}} I_n \neq \emptyset$.

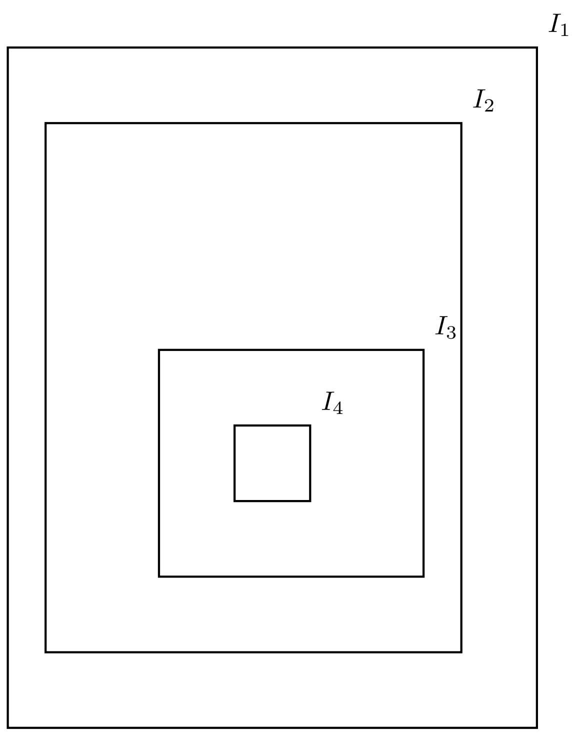

Figure 26: The first four $2-$cells in the type of sequence $\{I_n\}$ described in Proposition 2.8.

Proof. Let I_n be the set of points $\mathbf{x} = (x_1, \ldots, x_k)$ such that $a_{n,j} \leq x_j \leq b_{n,j}$ for $j = 1, \ldots, k$ and $n \in \mathbb{N}$, and write $I_{n,j} = [a_{n,j}, a_{b,j}]$. By Lemma 2.2, for each $I_{n,j}$, $\cap_{n\in\mathbb{N}} I_{n,j} \neq \emptyset$, so there is some $x_j^* \in \cap_{n\in\mathbb{N}} I_{n,j}$ which satisfies

$$a_{n,j} \leq x_j^* \leq b_{n,j}$$

for $j = 1, \ldots, k$ and $n \in \mathbb{N}$. If we set $\mathbf{x}^* = (x_1^*, \ldots, x_k^*)$, then $\mathbf{x}^* \in I_n$ for all $n \in \mathbb{N}$. Therefore $\cap_{n\in\mathbb{N}} I_n \neq \emptyset$. $\square$

We will now prove that each $k-$cell is compact. The definition of a $k-$cell is equivalent to a closed and bounded set in $\mathbb{R}$, so this result will give us our sufficient conditions for compactness in $\mathbb{R}^n$. This result will give rise to the Heine-Borel Theorem which is an equivalence result, which will follow immediately from the compactness of k-cells, Theorem 2.3, and Theorem 2.4. That being said, the proof that each $k-$cell is compact is not immediate, and is on the more difficult side for a proof in an introductory analysis course.

Lemma 2.4. Every $k-$cell is compact.

Proof. Let $I = \{\mathbf{x} \in \mathbb{R}^k \mid a_j \leq x_j \leq b_j \ j = 1, \ldots, k\}$ be a $k-$cell. Let δ be the maximum distance between any two points in I.

$$\max_{\mathbf{x},\mathbf{y}\in I} d(\mathbf{x}, \mathbf{y}) = \delta = \left(\sum_{j=1}^{k} (b_j - a_j)^2\right)^{1/2}$$

For all $\mathbf{x}, \mathbf{y} \in I$, $|\mathbf{x} - \mathbf{y}| \leq \delta$ (Figure 27 shows this for $k = 2$).

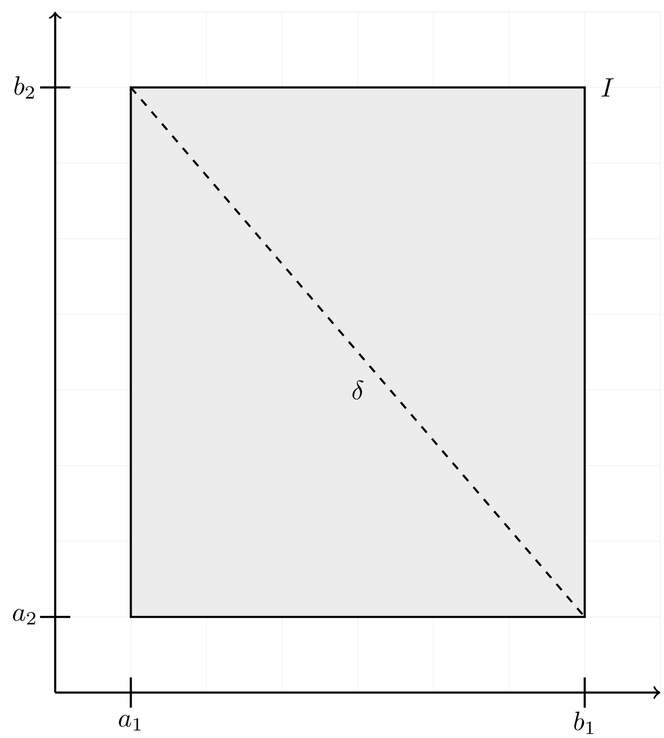

Figure 27: If I is a 2-cell.

For the sake of contradiction, assume that there exists some arbitrary open cover $\{G_\alpha\}$ of I which contains no finite subcover of I. Let $c_j = (a_j + b_j)/2$. The intervals $[a_j, c_j]$ and $[c_j, b_j]$ give rise to 2^k k-cells Q_i whose union is I. If each of these cells Q_i could be covered by a finite subcollection of $\{G_\alpha\}$, then I could be covered by the union of all these finite subcollections, which is finite. Since we've assumed K is not compact, then it must be that there exists at least one Q_i, call it I_1, that cannot be covered by any finite subcollection of $\{G_\alpha\}$. Now we divide I_1 into 2^k $k-$cells and repeat this process indefinitely, giving us a sequence $\{I_n\}$ (Figure 29). This sequence of k-cells was constructed to have three properties:

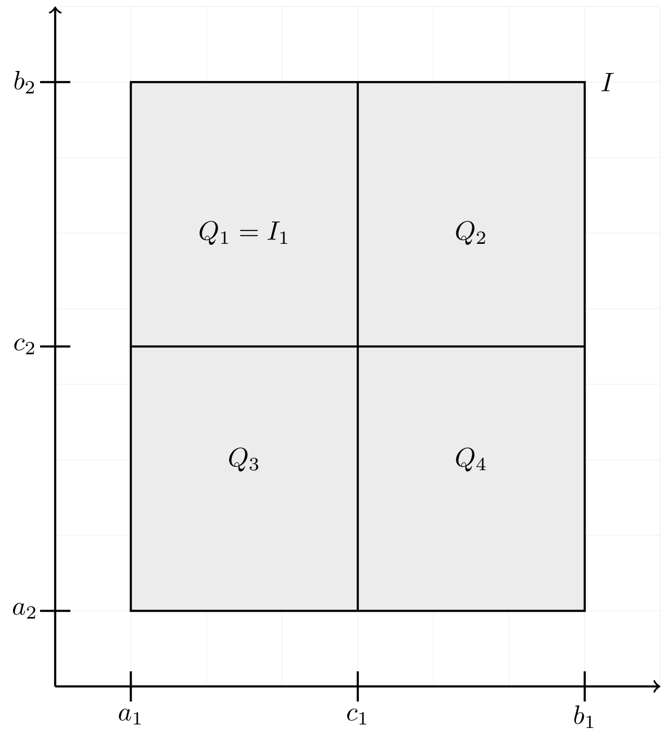

Figure 28: We partition I into 2^k cells Q_i. If I is not compact, then there is some Q_i which is not compact. Suppose in this case Q_1 is not compact, and call it I_1.

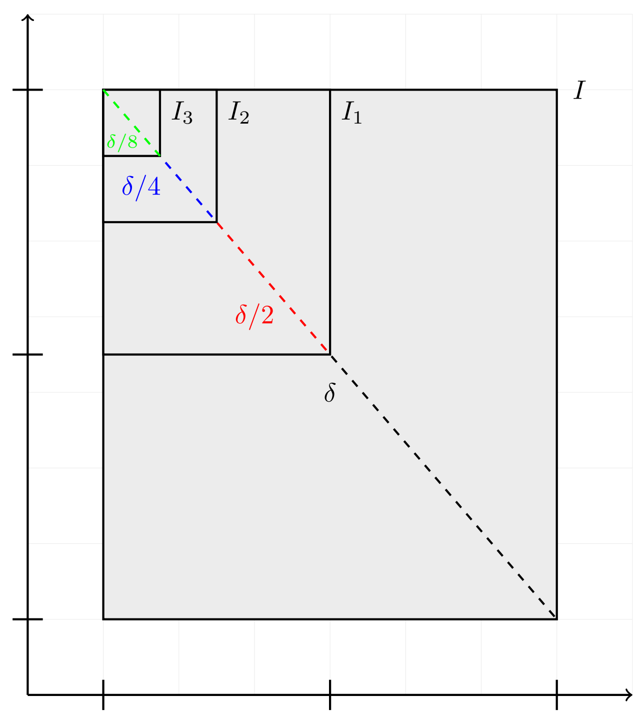

Figure 29: The first three 2-cells in the sequence $\{I_n\}$.

1. $I \supset I_1 \supset I_2 \supset I_3 \supset \cdots$

2. I_n cannot be covered by any finite subcollection of $\{G_\alpha\}$, otherwise K would not be compact.[21]

3. If $\mathbf{x}, \mathbf{y} \in I_n$, then $|\mathbf{x} - \mathbf{y}| \leq \delta/2^n$.[22]

By the first property of this sequence, we can invoke Lemma 2.3 to determine $\cap_{n\in\mathbb{N}} I_n \neq \emptyset$. This means there exists some $\mathbf{x}^* \in \mathbb{R}^k$ such that $\mathbf{x}^* \in I_n$ *for all* $n \in \mathbb{N}$. There must be some α such that $\mathbf{x}^* \in G_\alpha$, otherwise $\{G_\alpha\}$ would not be an open cover of I. Since G_α is open, there exists some $r > 0$ such that $N_r(x) \subseteq G_\alpha$. Alternatively, we could say that $|\mathbf{y} - \mathbf{x}^*| < r$ implies $y \in G_\alpha$ by the definition of $N_r(x)$. If we take n to be so large that $\delta/2^n < r$,[23] then $I_n \subseteq G_\alpha$, but this would mean I_n has a finite subcover. This contradicts property 2 of our sequence $\{I_n\}$, thereby contradicting the assumption that K is not compact. $\square$

Example 2.44. In order to make this proof a bit more concrete we'll walk through it with the 2-cell define by $I = [0, 1] \times [0, 1]$, and a specific open cover.[24] Let our open cover $G_\alpha = N_{0.01}(\alpha)$ for $\alpha \in I$.

$$I \subseteq \bigcup_{\alpha\in I} G_\alpha = \bigcup_{\alpha\in I} N_{0.01}(\alpha)$$

Assume that this open cover has no finite subcover. We have

$$\delta = ((1-0)^2 + (1-0)^2)^{1/2} = \sqrt{2}.$$

We divide I into four 2-cells: $Q_1 = [0, 1/2] \times [0, 1/2]$, $Q_2 = [0, 1/2] \times [1/2, 1]$, $Q_3 = [0.5, 1] \times [0, 1/2]$, and $Q_4 = [1/2, 1] \times [1/2, 1]$. If $\{G_\alpha\}$ has no finite subcover for I, then the same can be said for one of these Q_i. Suppose this is the case for Q_1, and let $I_1 = Q_1 = [0, 1/2] \times [0, 1/2]$. Repeat this process seven times until we arrive at $I_7 = [0, 1/256] \times [0, 1/256]$. The maximum distance between any two points in I_7 is

$$\left((1/256 - 0)^2 + (1/256 - 0)^2\right)^{1/2} \approx 0.0055 < 0.01.$$

Therefore, we can cover I_7 with a single $N_{0.01}(\alpha) \in \{G_\alpha\}$ for any $\alpha \in I_7$. But this means we can cover I_6 with four elements in $\{G_\alpha\}$,[25] and cover I_5 with 4^2 elements in $\{G_\alpha\}$, etc. We can cover I with 4^7 elements in $\{G_\alpha\}$. This contradicts the assumption that $\{G_\alpha\}$ has no finite subcover.

One of the subtler, but nevertheless important, parts of this process is that we could repeatedly divide the cells until we found a cell that fit in a neighborhood of radius 0.01. We will always be able to do this because of the Archimedean property as discussed in Footnote 22. It is important, that you do not associate this particular "trick" with the fact that I is a subset of $\mathbb{R}^2$. We are discussing the Archimedean property in the context of distances and radii, so we're actually using the fact that $\mathbb{R}^2$ is equipped with a distance function $d : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty]$. We're using the fact that the codomain of d has the Archimedean property!

We have now shown that every k-cell is compact. The proof is made more manageable with illustrations, but is still rather technical. Our contradiction came from the fact that as n becomes large, I_n becomes small. Eventually I_n will be so small, that it must have a finite open subcover. This contradicts the assumption that I is not compact.

Theorem 2.5 (Heine-Borel Theorem)**.** A set E in $\mathbb{R}^n$ is compact *if and only if* it is closed and bounded.

Proof.

($\Longrightarrow$) All compact sets are closed and bounded by Theorems 2.3 and 2.4.

[21] This follows from the same reasoning applied to the 2^k k-cells Q_i we initially chose I_1 from.

[22] When we divided I, the length of the diagonal for I_1 became half of δ. When we divide I_1, the length of the diagonal of I_2 became half of that of I_1. This means we can always write the diagonal of I_n in terms of powers of $1/2$ and δ.

[23] We can always find such an n. If not, $2^n \leq \delta/r$ for all $n \in \mathbb{N}$. This can't be though because $\mathbb{R}$ has the Archimedean property (Theorem 1.3).

[24] In order to give such an example, we need to specify an open cover to work with. In doing so, we're sort of shooting ourselves in the foot. The whole point of compactness is that *every* open cover has a finite subcover. What we're really proving in this example, is this specific open cover has no finite subcover.

[25] I'm playing a little fast and loose here with which exact elements, because I'm not specifying where the neighborhoods are centered.

($\Longleftarrow$) If E is closed and bounded then $E \subseteq I$ for some $n-$cell. Any closed subset of a compact set is compact by Proposition 2.4, so E is compact.

$\square$

Example 2.45. Any closed interval $[a, b] \subseteq \mathbb{R}$ is compact.

Theorem 2.6 (Bolzano–Weierstrass Theorem)**.** Every bounded infinite subset of $\mathbb{R}^n$ has a limit point in R^k.

Proof. Let $E \subseteq \mathbb{R}^n$ be bounded and infinite. There is some n-cell $I \subseteq \mathbb{R}^k$, such that $E \subseteq I$. By Lemma 2.4 I is compact. We know apply Proposition 2.6 to conclude that E has a limit point in I, which is also in $\mathbb{R}^k$. $\square$

Example 2.46. The set $(a, b) \subseteq \mathbb{R}^n$ is infinite and bounded. It has an infinite number of limit points in $\mathbb{R}^n$,[26] including a and b.

2.8 Exercises

Exercise 2.1. Show that the set of all algebraic numbers is countably infinite.

Exercise 2.2. Show that the set of all binary numbers with infinite digits is uncountably infinite.

Exercise 2.3. Verify that the taxi-cab metric on $\mathbb{R}^n$ is a valid metric space.

Exercise 2.4. Let X be an infinite set with the metric,

$$d(x, y) = \begin{cases} 1 \text{ if } x \neq y \\ 0 \text{ if } x = y \end{cases} .$$

Prove that (X, d) is a metric space. Which subsets of X are open? Which are closed?

Exercise 2.5. Prove that E° is open.

Exercise 2.6. Prove that E is open *if and only if* $E = E^\circ$.

Exercise 2.7. If $G \subseteq E$ and G is open, prove that $G \subseteq E^\circ$

Exercise 2.8. Prove that $(E^\circ)^c = \overline{E^c}$.

Exercise 2.9. Find an example of a set E in a metric space such that $E^\circ \neq (\bar{E})^\circ$.

Exercise 2.10. Find an example of a set E in a metric space such that $\bar{E} \neq \overline{E^\circ}$.

Exercise 2.11. Prove that ∂E is closed.

Exercise 2.12. Prove that $\partial(E^\circ) \subseteq \partial E$, and $\partial(\bar{E}) \subseteq \partial E$

Exercise 2.13. Prove that $\partial E = \partial(E^c)$.

Exercise 2.14. Suppose E is closed. Show that $(\partial E)^\circ = \emptyset$.

Exercise 2.15. Prove that $\partial(\partial E) \subseteq \partial E$. When will the sets be equal?

Exercise 2.16. Prove that E is closed *if and only if* $E \cap \partial E = \emptyset$.

Exercise 2.17. Prove that $\partial E = \emptyset$ *if and only if* E is closed and open.

Exercise 2.18. Prove that $\bar{E} = E \cup \partial E$.

Exercise 2.19. Prove that $(\partial \bar{E})^\circ = \emptyset$.

[26] In fact, the set of limit points is $[a, b] \subseteq \mathbb{R}^n$

3 Sequences and Series

Now that we are intimately familiar with the behavior of metric spaces, we can discuss a topic that may be familiar from calculus – sequences and series. Metric spaces will allow us to rigorously define convergence, and the properties related to the convergence os sequences and series. While we will derive some results and examples in general metric spaces, we will also start introducing results specific to $\mathbb{R}$.

3.1 Convergence

Definition 3.1. Let X be a metric space. A *sequence $\{x_n\}$* is an function from $f : \mathbb{N} \to X$. We will sometimes refer to an entire sequence as $x_n = f(n)$.

Using an arbitrary metric space X in this definition means that we, once again, always need to pay attention to what metric space we are in. We saw this with open sets, closed sets, and compact sets, and we will see it again. It will become especially relevant when determining if sequences converge.

Example 3.1. Let $x_n = 1/n$ be a sequence in $\mathbb{R}$. The first several terms in this sequence are

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots.$$

This set is also a sequence in $\mathbb{Q}$. This sequence is not defined in $\mathbb{Z}$ or $\mathbb{N}$, as neither of these sets has fractions.

Example 3.2. Let $x_n = 2$ be a sequence in $\mathbb{R}$. This constant sequence always takes on the value 2. This sequence is also a sequence in $\mathbb{N}, \mathbb{Z}$, and $\mathbb{Q}$.

Example 3.3. Let $x_n = (-1)^n$ be a sequence in $\mathbb{R}$. This sequence alternates between -1 and 1 for all values in $\mathbb{N}$.

Now we are ready to formalize what it means for a sequence to converge. When the idea of convergence is first introduced, you often hear phrases like "arbitrarily close". If a sequence converges to some point $x \in X$, we can *always* get closer to x. For any value in $\{x_n\}$, we can find other points "later" in the sequence $\{x_n\}$ that is even closer. If convergence is a recipe, then these are the ingredients:

1. No matter how "close" we get, we can always get closer with another point in $\{x_n\}$. Fortunately, we're in a metric space (X, d), so we can use d to determine how close we are.
2. Well actually, it cannot be *any* other points "later" in $\{x_n\}$. For instance, suppose we have the following sequence:
$$1, \frac{1}{2}, 2, \frac{1}{3}, 3, \frac{1}{4}, 4, \ldots.$$
The even terms of this series are getting closer to 0, while the odd terms are growing. The latter fact means this sequence doesn't converge. This happens because not *all* the points "later" in $\{x_n\}$ are closer.

These two ingredients will correspond to the ε and N in our definition.

Definition 3.2. A sequence $\{x_n\}$ in a metric space X *converges (in X)* if there exists an $x \in X$ such that *for all* $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ *for all* $n \geq N$. We will call x the *limit* of $\{x_n\}$, and write either $x_n \to x$, or

$$\lim_{n \to \infty} x_n = x.$$

We can think of the convergence of a sequence in the context of a hypothetical game. Suppose you and a friend have some convergent sequence $\{x_n\}$ in X. Your friend says some small number ε, and challenges you to find an N such that $d(x_n, x) < \varepsilon$ for all $n \geq N$. By the definition of convergence, you can do this. This frustrates your friend, so he demands you do it for an even smaller value of ε. Unfortunately for him, you will always be able to find such an N. No matter how small ε is, you will be able to do this.

Remark 3.1. We can formulate an equivalent definition of convergence using neighborhoods. If for all $\varepsilon > 0$, $d(x_n, x) < \varepsilon$ whenever $n \geq N$, then we could also say $x_n \in N_\varepsilon(x)$ for all $n \geq N$.

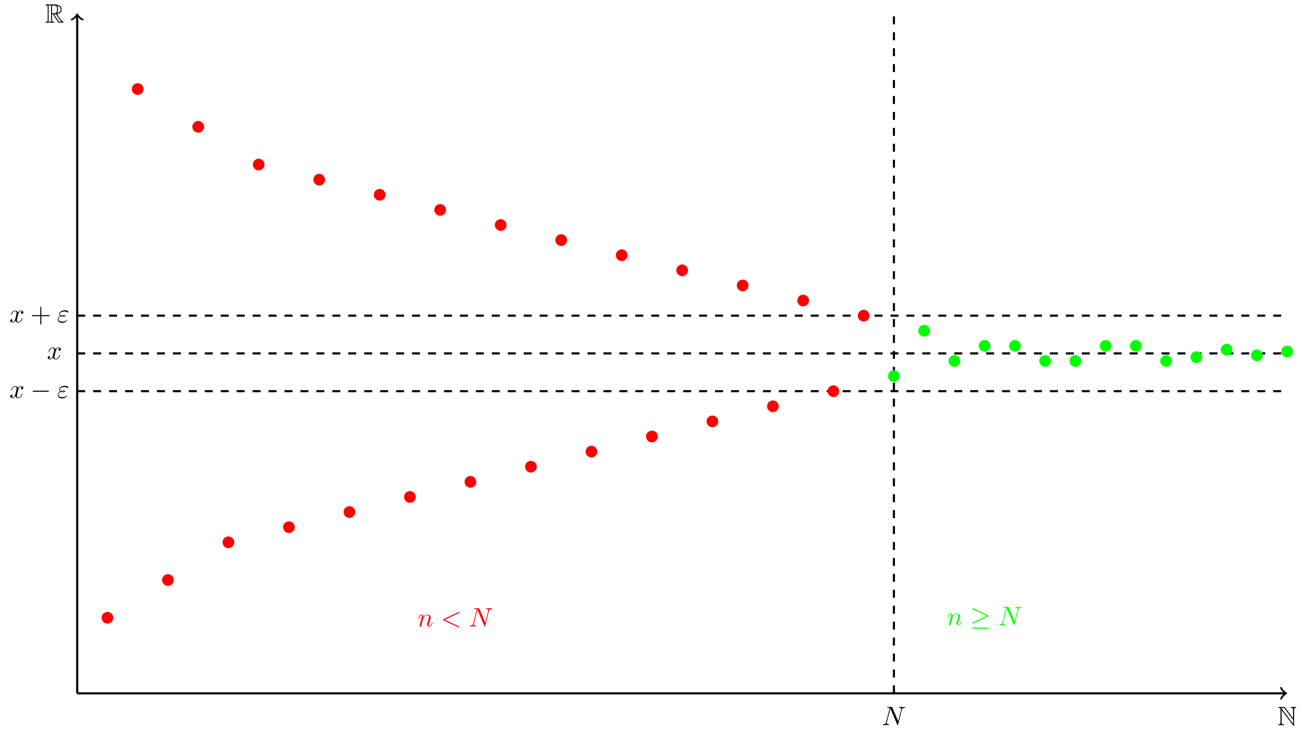

Figure 30: A convergent sequence $\{x_n\}$ in $\mathbb{R}$. No matter how small we take ε to be, we can always find some N such that all $d(x_n, x) = |x_n - x| < \varepsilon$ for all $n \geq N$. We could also write $x_n \in N_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$.

Example 3.4. The sequence $x_n = 1/n$ converges to 0 in $\mathbb{R}$. Suppose your friend lets $\varepsilon = 0.01$, and asks you to find an $N \in \mathbb{N}$ such that

$$d(1/n, 0) = |1/n - 0| = 1/n < \varepsilon = 0.01$$

for all $n \geq N$. If you let $N = 101$, then you have done this.

$$\begin{gathered} 1/101 < 0.01 \\ 1/102 < 0.01 \\ 1/103 < 0.01 \\ \vdots \end{gathered}$$

Your friend then gives you $\varepsilon = 0.001$. In this case, let $N = 1001$.

$$\begin{gathered} 1/1001 < 0.01 \\ 1/1002 < 0.01 \\ 1/1003 < 0.01 \\ \vdots \end{gathered}$$

You're already bored of this game, so you get an idea. Maybe you can find some function of ε that will give you your value of N. In order to do this, you just manipulate the inequality you must satisfy.

$$\begin{aligned} d(1/n, 0) &< \varepsilon \\ |1/n - 0| &< \varepsilon \\ 1/n &< \varepsilon \\ n &> \varepsilon^{-1} \end{aligned}$$

We just let $N = \varepsilon^{-1} + 1$. This way, for all $n \geq N$, we have $n > \varepsilon^{-1}$, which implies $d(1/n, 0) < \varepsilon$.

Remark 3.2. The definition of convergence has two inequalities. The inequality $d(x_n, x) < \varepsilon$ is strict, while $n \geq N$ is not. In the grand scheme of things, it doesn't matter if these are strict or not. If we instead had $d(x_n, x) \leq \varepsilon$, then we could just have taken $N = \varepsilon^{-1}$ in the previous example. Alternatively, if we had $d(x_n, x) < \varepsilon$ and $n > N$, then $N = \varepsilon^{-1}$ would work as well. I'm going to try very hard to stick with the inequalities in the definition, but I may make a mistake. Just know that it doesn't change the results of proofs at all. It does mean we need to be a little careful when using neighborhoods though, because they are open set.

Example 3.5. We can verify that the sequence $x_n = (n+1)/(n-1)$ converges to 1 in $\mathbb{R}$. Let $\varepsilon > 0$. We want to find the value of N in terms of ε that satisfies $d(x_n, x) < \varepsilon$ for all $n \geq N$. First notice that

$$d(x_n, x) = \left|\frac{n-1}{n+1} - 1\right| = \left|\frac{-2}{n+1}\right| = \frac{2}{n+1}.$$

We can satisfy $\frac{2}{n+1} < \varepsilon$ with $N = 2/\varepsilon - 1$.[27] We have that $d(x_n, x) < \varepsilon$ for all $n \geq 2/\varepsilon - 1$.

Example 3.6. Let $\mathbb{Z}$ be a metric space equipped with the p-adic metric (see Example 2.4). The sequence $x_n = p^n$ converges to 0 for any prime p. We have

$$d(x_n, 0) = p^{-\max\{m \in \mathbb{N} \mid p^m | (p^n - 0)\}} = p^{-n}.$$

If we let $N = \ln(2/\varepsilon)/\ln(p)$, then for all $\varepsilon > 0$ we have

$$d(x_n, 0) \leq d(x_N, 0) = p^{-\frac{\ln(2/\varepsilon)}{\ln(p)}} = p^{\frac{\ln(2/\varepsilon)}{\ln(p)}} = p^{\log_p(2/\varepsilon)} = \frac{2}{\varepsilon} < \varepsilon$$

for all $n \geq N$. Let's let $p = 2$. The sequence $x_n = 2^n$ will converge in $\mathbb{Z}$ with the 2-adic metric.

$$\{2^n\} = \{2, 4, 8, 16, 32, 64, 128, \ldots\}$$

n	x_n	$d(x_n, 0)$
1	2	1/2
2	4	1/4
3	8	1/8
4	16	1/16
5	32	1/32
6	64	1/64

Example 3.7. The series $x_n = (1 + 1/n)^n$ converges to e in $\mathbb{R}$. This series does not converge in $\mathbb{Q}$, because $e \notin \mathbb{Q}$.

Example 3.8. Let $X = (0, 1]$ be equipped with the Euclidean metric. The sequence $x_n = 1/n$ does not converge in X, because $0 \notin X$.

Example 3.9. Let $x_n = c$ for some constant $c \in \mathbb{R}$. This type of sequence is often called a *constant sequence*. All constant sequences converge to c in $\mathbb{R}$, as

$$|x_n - c| = |c - c| = 0 < \varepsilon$$

for all ε. Constant sequences are the only sequences that converge in $\mathbb{Z}$ equipped with the Euclidean metric.

Remark 3.3 (Where's My Limit?!)**.** These last two examples really emphasize the fact that the limit must be in the same metric space as our sequence. If our sequence is defined by some $f : \mathbb{N} \to X$, we need $x \in X$! Note that this is different from requiring that the value x is actually realized by our sequence. That is, it needn't be the case that f is in the image of X. We just need it to be in the codomain of X. In Example 3.4, $0 \notin f(\mathbb{N})$ (the image/range of $\mathbb{N}$), but it is in $\mathbb{R}$, and that's all that matters.

[27] You may be thinking that it isn't always true that this N will be in $\mathbb{N}$. That's fine. We could just round the answer to get a whole number that satisfies the inequality. Generally, we're not too worried about this.

3.2 Properties Related to Convergence

Now we'll cover some basic properties of sequences and convergence. We will also prove several results that are specific to sequences in $\mathbb{R}$, many of which should be familiar from calculus.

Proposition 3.1. Let $\{x_n\}$ be a sequence in a metric space X. The sequence $\{x_n\}$ converges to $x \in X$ *if and only if* every neighborhood of p contains x_n for all but finitely many n.

Proof.

($\Longrightarrow$) Suppose $x_n \to\ x$, and let $N_r(x)$ be some neighborhood of x. For some $\varepsilon = r > 0$, $x_n \in N_r(x)$ for all $n \geq N$ (see Remark 3.1). Therefore $N_r(x)$ contains x_n for all but finitely n, those being $\{x_1, \cdots x_{N-1}\}$.

($\Longleftarrow$) Suppose every neighborhood of x contains all but finitely many x_n. Fix $\varepsilon > 0$, and observe $N_\varepsilon(x)$. By our assumption, there exists an N such that $x_n \in N_\varepsilon(x)$ for $n \geq N$. Therefore we have $d(x_n, x) < \varepsilon$ if $n \geq N$, so $x_n \to x$.

$\square$

Remark 3.4 (Limit vs. Limit Point)**.** This result implies that if $x_n \to x$, then x is a limit point of the range of x_n. The converse is not necessarily true. Take the sequence $x_n = (-1)^n(1 + 1/n)$ as an example.

$$-2, \frac{3}{2}, -\frac{4}{3}, \frac{5}{4}, -\frac{6}{5}, \ldots$$

The range of this sequence has two limit points, $\{-1, 1\}$. Neither of these are limits of the sequence, as this sequences fails to converge because it alternates.

Example 3.10. Let $x_n = 1/n$ be a sequence in $\mathbb{R}$. Example 3.4 showed that $x_n \to 0$. For the neighborhood $N_{0.01}(0)$ we have:

$$N_{0.01}(0) = \left\{\frac{1}{101}, \frac{1}{102}, \frac{1}{103}, \ldots\right\},$$

where our finite set of points not in $N_{0.01}(0)$ is

$$\{x_n\} \backslash N_{0.01}(0) = \left\{1, \frac{1}{2}, \ldots, \frac{1}{100}\right\}.$$

Proposition 3.2 (Uniqueness of Limits)**.** Let $\{x_n\}$ be a sequence in a metric space X. If $x, x' \in X$, and if $\{x_n\}$ converges to x and x', then $x = x'$.

Remark 3.5 (Playing with ε)**.** Before we prove this, we should highlight a "trick" we will use. It is the most common technique used in proofs involving ε. If we know for all $\varepsilon > 0$, $d(x_n, x) < \varepsilon$ for all $n \geq N$, then we can use *any* $\varepsilon \in (0, \infty]$. This means we have $d(x_n, x) < f(\varepsilon)$ for all $n \geq N$, where $f : (0, \infty] \to (0, \infty]$. Return to the hypothetical game you are playing with your friend. At first, your friend wants to let $\varepsilon = 0.01$, but then he thinks "no that's too big, let me cut it in half", and he uses $\varepsilon/2 = 0.005$. The number $\varepsilon/2 > 0$ so it still is valid. He could even say "let me square it, divide it by 4, and then add π". In this case $f : (0, \infty] \to (0, \infty]$ is define as $f(\varepsilon) = \varepsilon^2/4 + \pi$, and is still valid because $f(\varepsilon) > 0$.

Why would we want to do this? We may want to show something converges, and do so using inequalities we already know that involve ε. We want our end result to show that $d(x_n, x) < \varepsilon$, so we need to choose our initial inequalities involving ε such that they yield a single ε. This probably isn't too clear right now, but this next proof, and that for Theorem 3.2 will hopefully make this more clear.

Proof. Assume that $x_n \to x$, and $x_n \to x'$. Let $\varepsilon > 0$. There exists $N', N'' \in \mathbb{N}$ such that

$$\begin{aligned} n \geq N &\implies d(x_n, x) < \varepsilon/2, \\ n \geq N' &\implies d(x_n, x') < \varepsilon/2. \end{aligned}$$

If we let $N = \max\{N', N''\}$, then using the triangle inequality gives

$$d(x, x') \leq d(x, x_n) + d(x_n, x') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,^{28}$$

for all $n \geq N$. If this holds for all $\varepsilon > 0$, then it must be that $d(x, x') = 0$, so $x = x'$. □

Definition 3.3. A set $\{x_n\}$ in a metric space X is *bounded* if its range is bounded. That is, if $f : \mathbb{N} \to \mathbb{R}$ is the function corresponding to x_n, the set $f(\mathbb{N})$ is bounded in X.

Example 3.11. The sequence $x_n = 1/n$ in $\mathbb{R}$ is bounded. We have a range of $f(\mathbb{N}) \subseteq (0, 1]$, so the range is clearly bounded.

Example 3.12. The sequence 2^n in the 2-adic metric (see Example 3.6) is bounded. The range of this sequence is $\{2, 4, 8, 16, \ldots\}$, but we have $d(x, y) \leq 1/2$ for all $x, y \in \{2, 4, 8, 16, \ldots\}$.

Proposition 3.3. Let $\{x_n\}$ be a sequence in a metric space X. If $\{x_n\}$ converges, then $\{x_n\}$ is bounded

Proof. Suppose $x_n \to x$. For all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n > N$. We have $1 > 0$, so we can let $\varepsilon = 1$. There is some N such that $d(x_n, x) < 1$ for all $n > N$. Let

$$r = \max\{1, d(p_1, p), \ldots, d(p_N, p)\}.$$

We have $d(p_n, p) \leq r$ for all $n \in \mathbb{N}$, so the sequence is bounded. □

Example 3.13. Let's work through this proof with an actual example. Let $x_n = 1/n$ in $\mathbb{R}$.[29] We know $x_n \to 0$, so there exists an N such that $d(x_n, 0) < 1$ for all $n \geq N$. In this case $N = 2$. We let $r = \max\{1, d(p_1, p)\} = \max\{1, 1\} = 1$. We have $d(x_n, 0) \leq 1$ for all $n \in \mathbb{N}$.

This proof is short, but interesting. We know every neighborhood of our limit 0 contains all but finitely many n (Proposition 3.1). The maximum distance becomes that between the limit and the finite points excluded from some arbitrary $N_\varepsilon(0)$.[30]

Example 3.14. The converse of Proposition 3.3 *is not true*. Take $x_n = (-1)^n$ in $\mathbb{R}$. The sequence is bounded, as the range of the sequence is $\{-1, 1\}$. Despite being bounded, the sequence does not converge as it alternates between 1 and -1.

Recall that in Section 2, we sometimes discussed approximating a limit point of some set E with elements in E. This was not the most formal of discussions, but our next theorem will make this fact explicit. The theorem is a statement about existence, and does not provide an actual construction of the sequence in claims exists. It's *very* important to be able to distinguish when a result does one of these, but not the other. It can often have practical implications for problem solving.

Theorem 3.1. Let $\{x_n\}$ be a sequence in a metric space X. If $E \subseteq X$ and if x is a limit point of E, then there is a sequence $\{x_n\}$ in E such that $x_n \to x$ in X.

Proof. The point $x \in X$ is a limit point of E, so all $N_r(x)$ contain some points of E that are not x. If we let $r = 1/n$, then we have that there exists some point $x_n \in N_{1/n}(x)$. Equivelently, $d(x_n, x) < 1/n$. For all $\varepsilon > 0$, choose N such that $N\varepsilon > 1$.[31] This way, we have

$$d(x_n, x) < \frac{1}{1/\varepsilon} = \varepsilon$$

for all $n > N$. Therefor $x_n \to x$. □

[28] Becuase we picked $\varepsilon/2$, they added to the desired ε. If we hadn't done this, then we would have $d(x, x') < 2\varepsilon$. Sometimes, people are fine with this would just argue "well $2\varepsilon \to 0$ as $\varepsilon \to 0$" . This isn't wrong, but it's not exactly kosher. It's best to just satisfy the definition without having to use a limiting process with ε.

[29] Yes, I will keep using this very trite example.

[30] In this case we just took $\varepsilon = 1$, but any other number would work just fine.

[31] This gives $N > 1\varepsilon$.

Example 3.15. Let $E = (0,1] \subseteq \mathbb{R}$ and, $x_n = 1/n$ in E. The point $0 \in \mathbb{R}$ is a limit point of E. The sequence $\{x_n\}$ converges to 0 in X. Note that $\{x_n\}$ does not converge in E (Example 3.7).

Corollary 3.1. Let E be a subset of a metric space X. If E is dense in X, then there for all $x \in X$, there exists some sequence $\{x_n\}$ in E such that $x_n \to x$.

Corollary 3.1 is an amazingly useful result once we become more comfortable with limits (and continuity). We may want to prove that some set X has a certain property, which could require we verify some condition for each $x \in X$. If X has some dense subset Y, then we could just prove that the property exists for a limit of a sequence in E, because each point in X is a limit of such a sequence! This sounds like it would be more a more complicated method of proof, but that is because we are just starting to build the toolkit required to work with limits. If points in E are easier to work with than those in X,[32] then it may just be easier to take limits of them.

Example 3.16. The set $\mathbb{Q}$ is dense in $\mathbb{R}$. This means that every number in $\mathbb{R}$ is the limit of a sequence in $\mathbb{Q}$, including irrational numbers. We saw this already in Example 3.6. In this specific case, we actually can write down the sequence. Even if we do not know the explicit form of the sequence, we still know it at least exists.[33] You probably don't know any sequence in $\mathbb{Q}$ that converges to π, but you do know that such a sequence exists because of Corollary 3.1.[34]

The next two results pertain only to sequences in $\mathbb{R}^n$. The first is a set of familiar results from calculus, and the second gives us the means to determine if sequences of real vectors converge.

Theorem 3.2. Suppose $\{x_n\}$ and $\{y_n\}$ are sequences in $\mathbb{R}$, $\lim_{n\to\infty} x_n = x$, and $\lim_{n\to\infty} y_n = y$. Then

1. $\lim\limits_{n\to\infty} (x_n + y_n) = x + y$;
2. $\lim\limits_{n\to\infty} cx_n = cx$, for all $c \in \mathbb{R}$;
3. $\lim\limits_{n\to\infty} (x_n y_n) = xy$, for all $c \in \mathbb{R}$;
4. $\lim\limits_{n\to\infty} 1/x_n = 1/x$, provided $s_n \neq 0$ for all $n \in \mathbb{N}$, and $s \neq 0$.

Proof.

1. There exists $N_1, N_2 \in \mathbb{N}$ such that

$$\begin{aligned} n \geq N_1 &\implies |x_n - x| < \frac{\varepsilon}{2}, \\ n \geq N_2 &\implies |y_n - y| < \frac{\varepsilon}{2}. \end{aligned}$$

   If we set $N = \max\{N_1, N_2\}$, then for all $n \geq N$ we have

$$|(s_n + t_n) - (s + t)| = |(s_n - s) + (t_n - t)| \leq |(s_n - s)| + |(t_n - t)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

   Therefore $x_n + y_n \to x + y$.

2. Given $\varepsilon > 0$, there exists a $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon/c$ for all $n \geq N$. This means for all $n \geq N$, we have

$$|cx_n - cx| = c|x_n - x| < c \cdot \frac{\varepsilon}{c} = \varepsilon,$$

   so $cx_n \to cx$.

[32] Nealy every time it's actually the points in $X \backslash E$ that are the ones that are harder to work with.

[33] This could be considered a drawback of this proof. It is not a proof via construction, so we don't have some blueprint that tells us how to find the sequence.

[34] I cannot think of a sequence that does this off the top of my head, but the series $\sum_{n=0}^{\infty} \frac{4(-1)^k}{2k+1}$ will. We'll prove this in Section 7.

3. There exists $N_1, N_2 \in \mathbb{N}$ such that
$$n \geq N_1 \implies |x_n - x| < \frac{\sqrt{2}}{\varepsilon},$$
$$n \geq N_2 \implies |y_n - y| < \frac{\sqrt{2}}{\varepsilon}.$$
If we let $N = \max\{N_1, N_2\}$, then
$$\begin{aligned}
|x_n y_n - xy| &= |x_n y_n - xy + 0 + 0 + 0| \\
&= |x_n y_n - xy + (xy - yx) + (x_n y - y x_n) + (y_n x - x y_n)| \\
&= |(x_n y_n - x y_n - x_n y + xy) + (y x_n - yx) + (x y_n - xy)| \\
&= |(x_n - x)(y_n - y) + y(x_n - x) + x(y_n - y)| \\
&\leq |(x_n - x)(y_n - y)| + |y(x_n - x) + x(y_n - y)| \\
&\leq |(x_n - x)(y_n - y)| \\
&< \sqrt{\varepsilon}\sqrt{\varepsilon} \\
&= \varepsilon,
\end{aligned}$$
for all $n \geq N$. This gives $x_n y_n \to xy$.[35]

4. There exists an $N_1, N_2 \in \mathbb{N}$ such that
$$n \geq N_1 \implies |x_n - x| < \frac{|x|}{2},$$
$$n \geq N_2 \implies |x - x_n| < \frac{x^2}{2}\varepsilon.$$
Note that we can manipulate one of this inequalities using the triangle inequality:
$$\begin{aligned}
|x_n - x| &< \frac{|x|}{2} \\
|x| - |x_n| &< \frac{|x|}{2} \\
||x| - |x_n|| &< \frac{|x|}{2} \\
-\frac{|x|}{2} < |x| - |x_n| &< \frac{|x|}{2} \\
\frac{|x|}{2} < |x_n| &< \frac{3|x|}{2} \\
\frac{1}{|x_n|} &< \frac{2}{|x|}
\end{aligned}$$
If we let $N = \max\{N_1, N_2\}$, then
$$\begin{aligned}
\left|\frac{1}{x_n} - \frac{1}{x}\right| &= \left|\frac{1}{x}\frac{1}{x_n}(x - x_n)\right| \\
&= \frac{1}{|x|}\frac{1}{|x_n|}|x - x_n| \\
&< \frac{1}{|x|}\frac{1}{|x|}\frac{|x|^2\varepsilon}{2} \\
&= \varepsilon,
\end{aligned}$$
for all $n \geq N$. This gives $1/x_n \to 1/x$.

[35] A very common "trick" used in almost all nontrivial proofs involving ε is to add and subtract the same term within an absolute value, in effect adding zero. You can then rearrange the terms and use the triangle inequality to "separate" terms in the absolute value. I call is a "trick", because when I first saw someone do it I thought "well how the hell was I ever supposed to know to do that without ever seeing it done!"

□

Remark 3.6. Part 1 and 2 of Theorem 3.2 allow us immediately to perform subtraction using limits. Parts 3 and 4 allow us to do the same with division.

Remark 3.7 (Linearity of Limits)**.** The first two parts of Theorem 3.2 allow us to conclude that limits are linear. This shouldn't be news, as it comes up in any standard calculus course, but it has powerful implication. We will use limiting processes to define differentiation and integration, so the linearity of limits will give rise to the linearity of these two operations as well.

Proposition 3.4. Suppose $\{\mathbf{x}_n\}$ is a sequence in $\mathbb{R}^k$ where $\mathbf{x}_n = (x_{1,n}, \ldots, x_{k,n})$. The sequence $\{\mathbf{x}_n\}$ converges to $\mathbf{x} = (x_1, \ldots, x_n)$ *if and only if* $x_{j,n} \to x_j$ for $j = 1, \ldots, k$.

Proof.

($\Longrightarrow$) Suppose $\mathbf{x}_n \to \mathbf{x}$. For all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$|\mathbf{x}_n - \mathbf{x}| < \varepsilon$$

for all $n \geq N$. This gives that $x_{j,n} \to x_j$, as

$$\begin{aligned} |x_{j,n} - x_j| &< \left[(x_{1,n} - x_1)^2 + \ldots + (x_{k,n} - x_k)^2\right]^{1/2} \\ &= |\mathbf{x}_n - \mathbf{x}| \\ &< \varepsilon \end{aligned}$$

for $j = 1, \ldots, k$.

($\Longleftarrow$) Suppose $x_{j,n} \to x_j$ for $j = 1, \ldots, k$. For all $\varepsilon > 0$, there is some $N \in \mathbb{N}$ such that

$$|x_{j,n} - x_j| < \frac{\varepsilon}{\sqrt{k}}$$

for all $n \geq N$. For all $n \geq N$ we have

$$\begin{aligned} |\mathbf{x}_n - \mathbf{x}| &= \left[(x_{1,n} - x_1)^2 + \ldots + (x_{k,n} - x_k)^2\right]^{1/2} \\ &< \left[(\varepsilon/\sqrt{k})^2 + \ldots + (\varepsilon/\sqrt{k})^2\right]^{1/2} \\ &= \left(\frac{k\varepsilon^2}{k}\right)^{1/2} \\ &= \varepsilon. \end{aligned}$$

Therefore $\mathbf{x}_n \to \mathbf{x}$.

□

Proposition 3.4 establishes that a sequence of vectors converge if and only if each component converges. This is not terribly surprising.

3.3 Subsequences

Given some sequence $\{x_n\}$ in X, it's possible to restrict our attention to only certain terms of $\{x_n\}$. For example, if we have $x_n = (-1)^n$ in $\mathbb{R}$, we could look at every other term and in effect have a sequence of all 1's or all -1's. Clearly, both these sequences that are "contained" in $x_n = (-1)^n$ would converge. This case is particularly interesting because the sequence $x_n = (-1)^n$, despite us being able to "throw away" certain terms and have a convergent sequence as a result. As we'll see, this is not some fluke occurrence, and it's related to the fact that this sequence lives in a compact space! This will be the first of many nice/cool results that will follow from compactness.

Definition 3.4. Given a sequence $\{x_n\}$ in X, consider a sequence $\{n_k\}$ such that $n_1 < n_2 < \cdots$. The sequence $\{x_{n_k}\}$ is called a *subsequence* of $\{x_n\}$. If $\{x_{n_k}\}$ converges, its limit is called a *subsequential limit* of $\{x_n\}$.

Remark 3.8. If $\{x_n\}$ corresponds to the function $f : \mathbb{N} \to X$, then we can think of a subsequence $\{x_{n_k}\}$ corresponding to some function $g : \mathbb{N} \to f(\mathbb{N})$, where $f(\mathbb{N}) \subseteq X$. In this sense, a subsequence of $\{x_n\}$ is a sequence in the range/image of $\{x_n\}$.

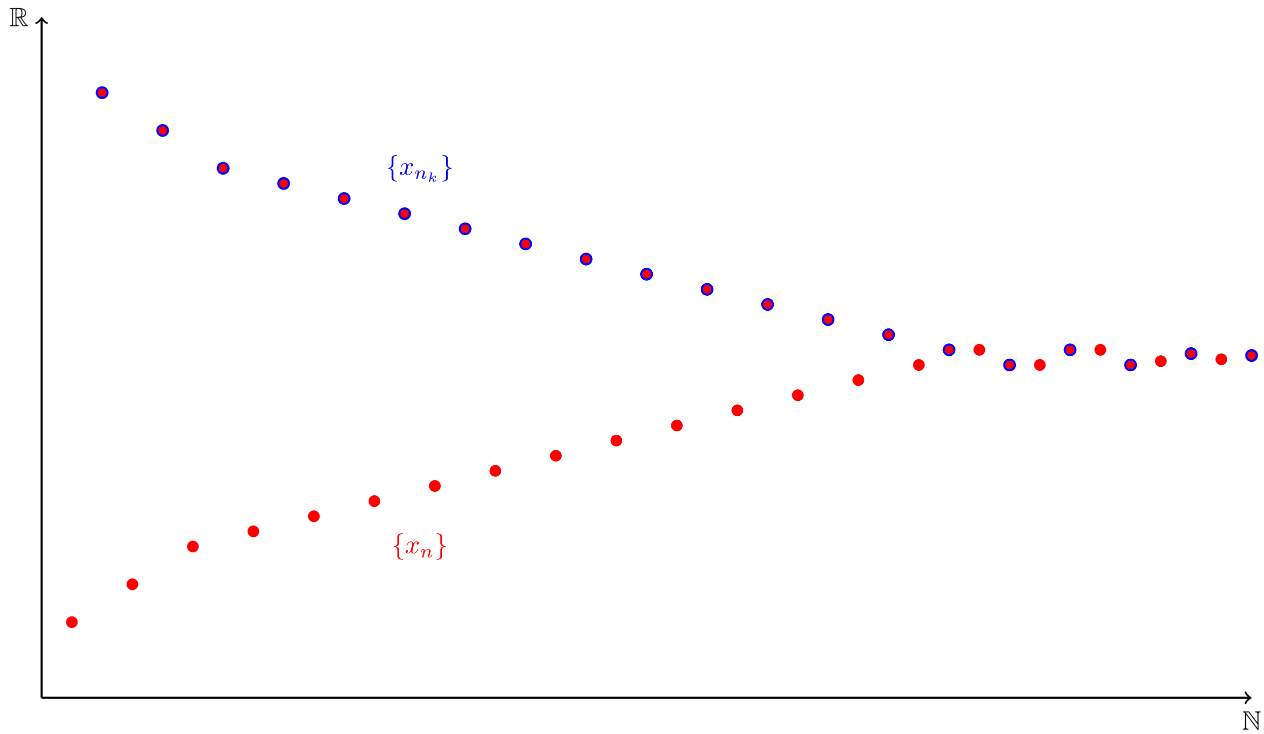

Figure 31: A sequence $\{x_n\}$ in $\mathbb{R}$, with a subsequence $\{x_{n_k}\}$.

A logical question to ask is how do we relate the limit of some converget sequence with the its subsequential limits? Is there at least one subsequential limit of a sequence that is the same as the limit of a sequence? Is the even stronger result that every subsequential limit of a sequence the same as the limit of the sequence possible?

Theorem 3.3. A sequence $\{x_n\}$ converges to x *if and only if* every subsequence of $\{x_n\}$ converges to x.

Proof.

($\Longrightarrow$) Suppose $x_n \to x$. For all ε, there exists an N such that $|x_n - x| < \varepsilon$ for all $n \geq N$. If x_{n_k} is some arbitrary subsequence of $\{x_n\}$, then $|x_{n_k} - x| < \varepsilon$ for all $n_k \geq N$. Therefore $x_{n_k} \to x$.

($\Longleftarrow$) Suppose every subsequence of $\{x_n\}$ converges to x. This means that the trivial subsequence of $\{x_{n_k}\} = \{x_n\}$ converges.[36]

$\square$

Corollary 3.2. A sequence x_n in a metric space X diverges *if and only if* it has a divergent subsequence, *or* more than one subsequential limit.

[36] This trivial subsequence comes from taking $\{n_k\} = \mathbb{N}$.

Theorem 3.3 and it's corollary relate the limit of a convergent sequence to its subsequential limits, and says they are the all the same. The proof of this is very brief, but the result is sweeping. *Every* single possible subsequence, a collection which is uncountably infinite, has the same limit.

Example 3.17. Let $\{x_n\} = \{1, 2, 3, 1, 2, 3, 1, 2, 3, \ldots\}$ be a bounded sequence in $\mathbb{R}$. This sequence has subsequences that converge to 1, 2, and 3, so $\{x_n\}$ diverges. Alternatively, we could say that $\{x_n\}$ diverges because it has a divergent subsequence in $\{x_{n_k}\} = \{1, 2, 1, 2, 1, 2, \ldots\}$.

A more interesting case to consider than that of a convergent sequence, is a divergent sequence. How do subsequences of divergent sequences behave. Do divergent sequences has subsequential limits? This question is answered with one of the most important theorems in analysis.[37]

Theorem 3.4. If $\{x_n\}$ is a sequence in a compact metric space X, then some subsequence of $\{x_n\}$ converges in X.

Proof. Let E be the range of $\{x_n\}$. If E is finite,[38] then we can construct a constant subsequence $\{x_{n_k}\}$ such that

$$x_{n_1} = x_{n_2} = \cdots = x,$$

where $x \in E$. This constant sequence converges.

If E is infinite, then E has some limit point $x \in X$ by the Bolzano-Weierstrass Property (Proposition 2.6). By Theorem 3.1, there exists a sequence $\{y_n\}$ in E such that $y_n \to x$. But E is just the range of $\{x_n\}$, so any sequence $\{y_n\}$ in E can be expressed as a subsequence $\{x_{n_k}\}$ (see Remark 3.8). Therefore x is a subsequential limit of x_n. $\square$

Corollary 3.3 (Bolzano-Weierstrass Theorem)**.** Every bounded sequence in $\mathbb{R}^k$ has a convergent subsequence.

Proof. A bounded subset of $\mathbb{R}^k$ is compact, so any such bounded sequence converges. $\square$

Remark 3.9 (Equivalence Forms of Compactness in Metric Space)**.** Recall Proposition 2.6 and Remark 2.10. These pertained to the fact that in a compact metric space, any subset has a limit point in the compact set. It was then discussed that there are more general settings in point-set topology where this may not be true for compact sets, and for that reason the property given by Proposition 2.6 is often called limit point compactness. A similar situation holds for Theorem 3.4. If every convergent sequence in some space has at least one subsequential limit, the space is sometimes called *sequentially compact*. In metric spaces, sequential compactness is equivalent to compactness, so knowing the difference isn't important.[39] Many times, the definition of sequential compactness is given as the definition of compactness, as it is easier to digest than the actual definition. In sum: a ***metric space*** is compact *if and only if* it is sequentially compact *if and only if* it is limit point compact. All of this will be made formal in Section 15

Example 3.18. The sequence $x_n = 1/n$ in $(0, 1]$ diverges (see Example 3.8). It also has no convergent subsequence, so by Theorem 3.4, $(0, 1]$ is not compact. We already knew this though, as $(0, 1]$ is not closed.

Example 3.19. The sequence $x_n = (-1)^n$ in $\mathbb{R}$ is bounded (see Example 3.14). Therefore it has at least one subsequential limit. Those limits are 1 and -1.

Remark 3.10. Much like the proof of Theorem 3.1, we do not find an explicit formula for a convergent subsequence when proving The Bolzano-Weierstrass Theorem. As of now, we don't have a blueprint for finding some convergent subsequence for a sequence in $\mathbb{R}^k$, we only know it is out there. Later on, we'll introduce the concept of lim sup and lim inf, and show a specific type of subsequence that will *always* converge.

[37] How can you tell if a result is important? A good rule of thumb is that something is important if it has some specific name.

[38] Remember that a sequence is infinite even if it ranges is finite. If it's range is finite, that just means it takes on at least one of these values an infinite number of times.

[39] This is why no introductory analysis course distinguishes them.

3.4 Cauchy Sequences

Up until now, showing a sequence converges has been purely a theoretical exercise. Proving x_n converges to x using the definition of convergence requires we know x, but how would we know the limit of a sequence if we didn't know it converged?[40]

Example 3.20. Let $x_n = (1 + 1/n)^n$ be a sequence in $\mathbb{R}$. In Example 3.7, I claimed this converged to the constant e, but this is not clear at all from looking at the sequence.

In this subsection we'll develop a way of verifying of a sequence converges without knowing the limit. It's hard to emphasize just how useful this is in actual applications of analysis, as it's rarely clear what some relevant sequence would converge to. The method we develop will not work in every metric space, but will work in the most common spaces. We start by introducing a second type of convergence.

Definition 3.5. A sequence $\{x_n\}$ in a metric space X is a *Cauchy sequence (in X)* if for all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ if $n \geq N$ and $m \geq N$.

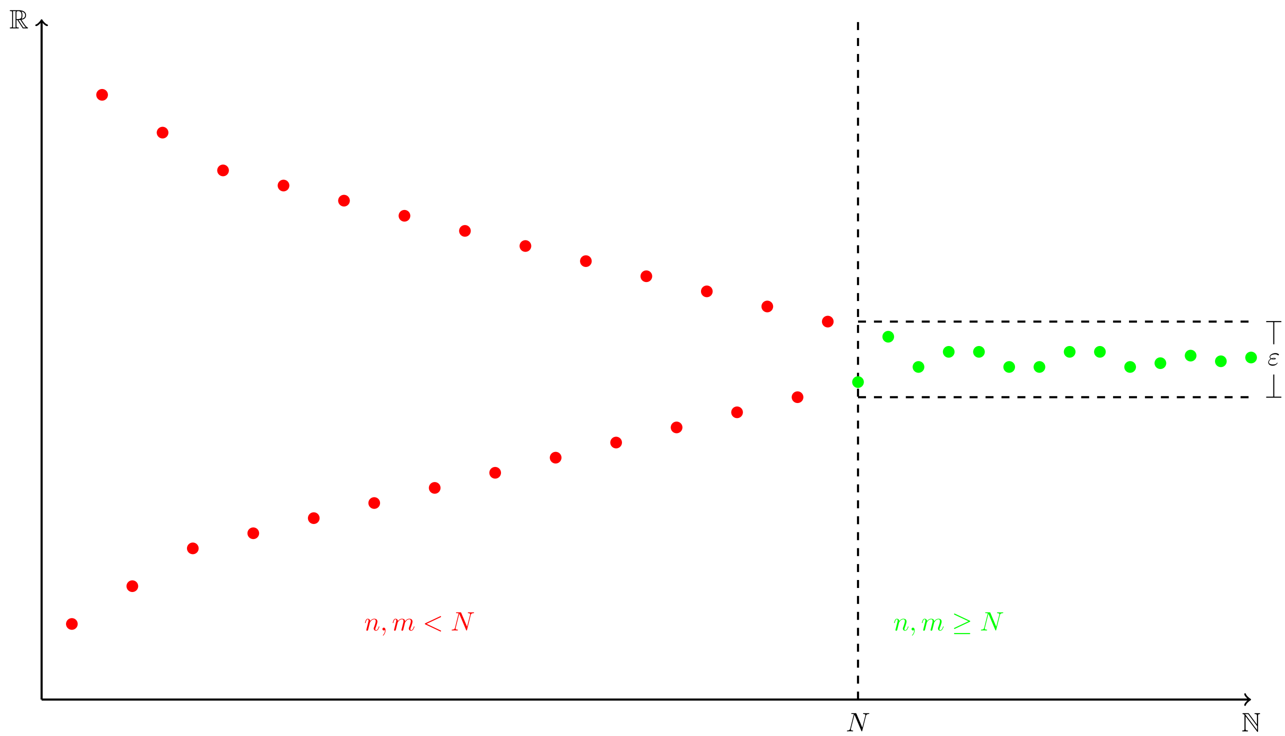

Figure 32: A Cauchy sequence in $\mathbb{R}$.

The terms in a Cauchy sequence get arbitrarily closer to each other over time. This is opposed to a convergent sequence, where the terms get arbitrarily close to some limit. This difference is the reason we never refer to a Cauchy sequence having a limit.[41]

Example 3.21. Let $x_n = 1/2^n$ be a sequence in $\mathbb{R}$. We can show that this sequence is a Cauchy sequence. If we let $N = -\ln(\varepsilon/4)/\ln(2)$, then we have

$$|x_n - x_m| = \left|\frac{1}{2^n} - \frac{1}{2^m}\right| \leq \frac{1}{2^N} + \frac{1}{2^N} = \frac{1}{2^{-\ln(\varepsilon/4)/\ln(2)}} + \frac{1}{2^{-\ln(\varepsilon/4)/\ln(2)}} = 2 \cdot 2^{\log_2(\varepsilon/4)} = \frac{\varepsilon}{2} < \varepsilon$$

[40] This is similar to something you may have seen in an intro stats course. You can use a z-test for the population mean, but that requires you know the population standard deviation. In what world would you know the population standard deviation but not the population mean?!

[41] Unless of course it also happens to be convergent. No matter how small we let ε be, we can always find an $N \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N$.

for all $n, m \geq N$. Therefore $x_n = 1/2^n$ is a Cauchy sequence.

While a Cauchy sequence does not have a limit, we can formulate a convergent sequence using a Cauchy sequence. If the terms in a Cauchy sequence are getting arbitrarily close, then one may suspect that if we look at the "tail" of such a sequence, the maximum distance between them is shrinking. This idea gives rise to the next definition and theorem, which will prove an alternate definition for a Cauchy sequence that uses the definition of a convergent sequence.

Definition 3.6. Let E be a nonempty subset of a metric space X. The *diameter* of E is the supremum of the set of $d(x, y)$ for all $x, y \in E$.

$$\operatorname{diam} E = \sup\{d(x, y) \mid x, y \in E\}$$

Remark 3.11. Whenever you see a supremum, it's worth asking yourself "does this necessarily exist? If so, how do I know this?" In this case, the answer is yes. In Definition 3.6, the set E is a subset of $[0, \infty]$, because $d(x, y) \in [0, \infty]$ for all $d(x, y)$ by the definition of d. This in turn gives $E \subseteq \mathbb{R}$, and any subset of $\mathbb{R}$ has a supremum.

Example 3.22. Let $E = [a, b] \subseteq \mathbb{R}$. We have $\operatorname{diam} E = b - a$. If we instead have $E = (b - a)$, then we still have $\operatorname{diam} E = b - a$, even though this diameter is never realized, because a supremum of a set needn't be in the set.

Lemma 3.1. If x_n is a sequence in X, and $E_n = \{x_N, x_{N+1}, x_{N+2}, \ldots\}$, then $\{x_n\}$ is a cauchy sequence *if and only if* $\lim_{N\to\infty} \operatorname{diam} E_n = 0$.

Proof.

($\Longrightarrow$) Suppose $\{x_n\}$ is a Cauchy sequence. For all $\varepsilon > 0$, there exists some $N \in \mathbb{N}$ such that

$$d(x_n, x_m) < \frac{\varepsilon}{2}$$

for all $n, m \geq N$. We use the supremum to define $\operatorname{diam} E_n$, so we know there exists $x', y' \in E_n$ such that

$$0 \leq \operatorname{diam} E_N = \sup_{x,y\in E_N} d(x, y) < d(x', y') + \frac{\varepsilon}{2}.$$

By the definition of E_n, x' and y' take the form x_n and p_m for some $n, m \geq N' \geq N$. Therefore

$$\operatorname{diam} E_n < d(x_n, x_m) + \frac{\varepsilon}{2} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This means that $\lim_{N\to\infty} \operatorname{diam} E_n = 0$.

($\Longleftarrow$) Suppose $\lim_{N\to\infty} \operatorname{diam} E_n = 0$. For all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$d(\operatorname{diam} E_n, 0) = \operatorname{diam} E_n < \varepsilon$$

for all $n \geq N$. If we have $n, m \geq N$, then $x_n, x_m \in E_N$. Therefore

$$d(x_n, x_m) \leq \sup_{x,y\in E_n} d(x, y) = \operatorname{diam} E_N < \varepsilon$$

for all $n, m \geq N$. This gives that $\{x_n\}$ is a Cauchy Sequence

$\square$

Example 3.23. Let $x_n = 1/n$ in $\mathbb{R}$. For $E_n = \{1/N, 1/(N + 1), 1/(N + 2), \ldots\}$, we have

$$\operatorname{diam} E_n = \sup\{d(x, y) \mid x, y \in E_n\} = \sup\{d(1/n, y) \mid y \in E_n\} = 1/N.$$

This means $\lim_{N\to\infty} \operatorname{diam} E_n = 0$, so $\{x_n\}$ is a Cauchy sequence by Theorem 3.5.

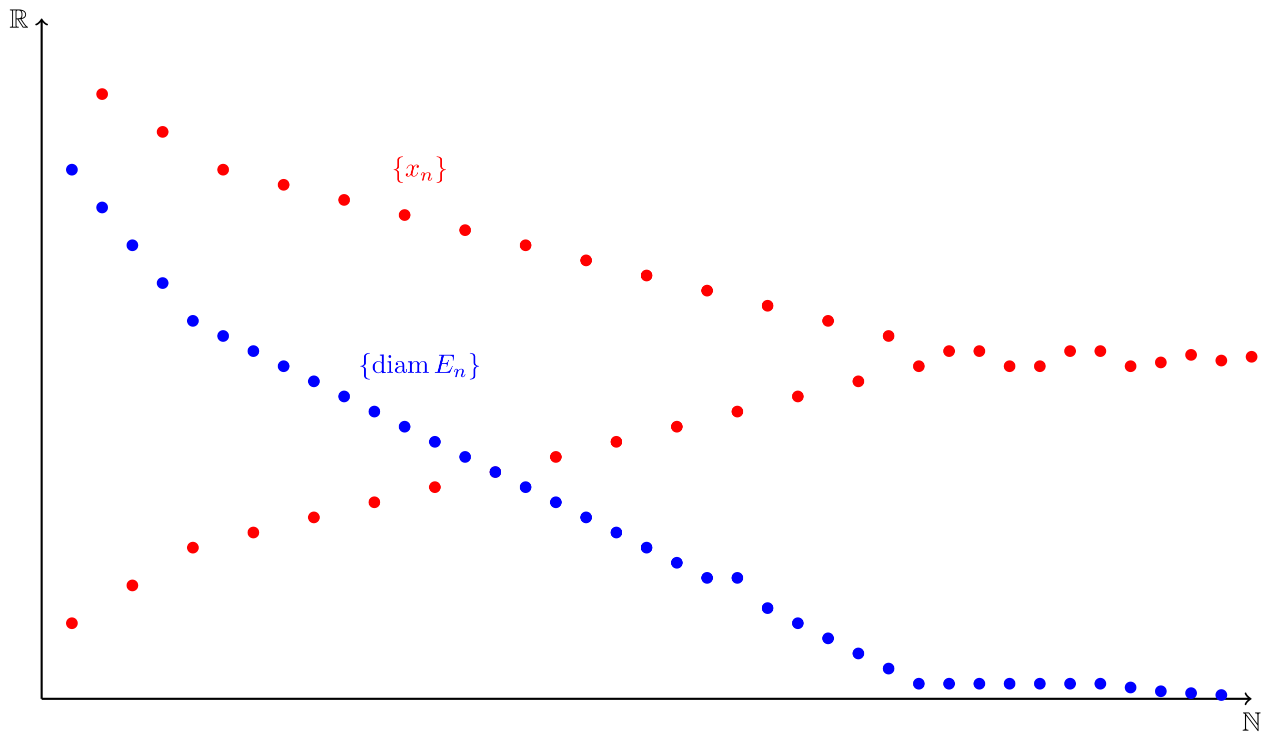

Figure 33: A Cauchy sequence in $\{x_n\}$ in $\mathbb{R}$, and the sequence $\{\operatorname{diam} E_n\}$ as defined in Theorem 3.5

Considering the diameter of a Cauchy sequence may seem redundant and overly complicated, but it allows us to prove many results relating to Cauchy sequences. Their introduction is more a means to and end than a definition that is important in its own right. That being said, we do need to introduce a couple of properties of the diameter in order to get mileage out of the concept when proving results.

Lemma 3.2 (Properties of Diameter)**.**

1. If $\bar{E}$ is the closure of a set E in a metric space X, then $\operatorname{diam} E = \operatorname{diam} \bar{E}$.
2. If K_n is a sequence of compact sets in X such that $K_{n+1} \subseteq K_n$ for all $n \in \mathbb{N}$ and if
$$\lim_{N\to\infty} \operatorname{diam} K_n = 0,$$
then $\cap_{n=1}^{\infty} K_n$ consists of exactly one point.

Proof.

1. We have that $E \subseteq \bar{E}$, so $\operatorname{diam} E \leq \operatorname{diam} \bar{E}$.[42] Now we will show that $\operatorname{diam} E \geq \operatorname{diam} \bar{E}$, which gives the equality. Fix a specific $\varepsilon > 0$, and let $x, y \in \bar{E}$. The point x and y are limit points of E, so there exists $x', y' \in E$ such that $d(x, x') < \varepsilon$ and $d(y, y') < \varepsilon$. The triangle inequality gives
$$d(x, y) \leq d(x, x') + d(x', y') + d(y', y) < 2\varepsilon + d(x', y') \leq 2\varepsilon + \operatorname{diam} E.$$
Our selection of $x, y \in \bar{E}$ and ε were arbitrary, so $\operatorname{diam} \bar{E} \leq \operatorname{diam} E$.
2. By Lemma 2.3, $\cap_{i=1}^{\infty} K_n \neq \emptyset$. For the sake of contradiction, suppose $\cap_{i=1}^{\infty} K_n \neq$ contains more than one point. If this is the case, then $\operatorname{diam} \cap_{i=1}^{\infty} K_n \neq> 0$. For all $n \in \mathbb{N}$, $\cap_{i=1}^{\infty} K_n \neq\subseteq K_n$, so $\operatorname{diam} K_n \geq \operatorname{diam} \cap_{i=1}^{\infty} K_n \neq$. This contradicts the assumption that $\lim_{N\to\infty} \operatorname{diam} K_n \to 0$.

[42] As much as I hate to borrow condescending adjectives from Rudin (1976), this is "clear".

□

With the introduction of Cauchy sequences comes the million dollar question that underlies this whole subsection. Do Cauchy sequences converge? The answer is in fact, no. Most of the time, Cauchy sequences do converge, in general this is not true. What is true however, is that all convergent sequences are Cauchy sequences. The task of finding a method to prove convergence without knowing a limit becomes a matter of finding a general class of functions for which all Cauchy sequences converge. Once we know this, we can simply verify such a sequence is Cauchy and know it converges.

Example 3.24. Recall the sequence $x_n = 1/n$ in the space $(0,1]$. This sequence does not converge because $0 \notin (0,1]$. Nevertheless, it is a Cauchy sequence. Cauchy sequences do not have an explicit limit, so excluding 0 from the space $\{x_n\}$ is in does not affect whether or not $\{x_n\}$ is Cauchy.

Example 3.24 shows that a Cauchy sequence need not be convergent, but we can show that any convergent sequence is a Cauchy sequence.

Theorem 3.5. Let $\{x_n\}$ be a sequence in a metric space X. If x_n converges to some limit $x \in X$, then $\{x_n\}$ is a Cauchy sequence.

Proof. For all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon/2$ for all $n \geq N$. The triangle inequality gives

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for all $m, n \geq N$. Thus $\{x_n\}$ is Cauchy. □

Example 3.25. Example 3.4, 3.5, 3.6, 3.7, and 3.9 also double as examples of Cauchy sequences by Theorem 3.6.

We now turn to one of the most important results involving sequences. If we restrict our attention to compact spaces, then all Cauchy sequences converge. This allows us to prove a sequence in a compact space converges without having any knowledge of its limit!

Theorem 3.6. If X is a compact metric space and if $\{x_n\}$ is a Cauchy sequence in X, then $\{x_n\}$ converges to some point of X.

Proof. Let $\{x_n\}$ be a Cauchy sequence in a compact space X. For $n\mathbb{N}$, define $E_n = \{x_N, x_{N+1}, x_{N+2}, \ldots\}$. By Lemma 3.1 and Lemma 3.2,

$$\lim_{N\to\infty} \operatorname{diam} E = \lim_{N\to\infty} \operatorname{diam} \bar{E} = 0.$$

The set $\bar{E}_n$ is a closed subset (Lemma 3.2) of a compact space, so it is compact (Proposition 2.4). We also have $E_{n+1} \subseteq E_n$ for all $n \in \mathbb{N}$ so $\bar{E}_{n+1} \subseteq \bar{E}_n$.

If $\bar{E}_{n+1} \subseteq \bar{E}_n$ for all $n \in \mathbb{N}$, where $\bar{E}_n$ is compact and $\lim_{N\to\infty} \operatorname{diam} \bar{E} = 0$, there is a unique $x \in X$ such that $x \in \cap_{n=1}^{\infty}$ (Lemma 3.2).

Let $\varepsilon > 0$. By $\lim_{N\to\infty} \operatorname{diam} \bar{E} = 0$, There exists some $N' \in \mathbb{N}$ such that $\operatorname{diam} \bar{E}_N < \varepsilon$ for all $N \geq N'$. Our unique point x is in $\bar{E}_n$, so $d(x,y) \leq \varepsilon$ for all $y \in \bar{E}$, and hence for every $y \in E_n$. This is equivalent to saying $x_n \in N_\varepsilon(x)$ for all $n \geq N_0$. This gives $x_n \to x$. □

Theorem 3.7 (Cauchy Criterion)**.** In $\mathbb{R}^k$, every Cauchy sequence converges.

To show this result, we will show that a Cauchy sequence in $\mathbb{R}^k$ is bounded.

Proof. FINISH □

The Cauchy Criterion is what we have been building to. To prove a sequence converges in $\mathbb{R}^k$, we just need to show it is a Cauchy sequence. This is often much easier and not as time consuming. It's important to remember that Theorem 3.6 is a more general result. It's so useful, we even have a name for metric spaces where Cauchy sequences always converge.

Definition 3.7. A metric space is *complete* is every Cauchy sequence converges.

Example 3.26. The real line $\mathbb{R}$ is complete by the Cauchy Criterion. This example also shows that the converse of Theorem 3.6 does not hold. The real line is not compact, but is nevertheless complete.

Example 3.27. The rationals $\mathbb{Q}$ are not complete. For example, the the sequence $x_n = (1 + 1/n)^n$ does not converge in $\mathbb{Q}$, but it is Cauchy.

Example 3.28. The set of all real valued continuous functions defined on the domain $[a, b] \subseteq \mathbb{R}$, denoted $\mathscr{C}([a, b])$, is complete. Right now, this example seems very abstract, as we haven't even discussed sequences of functions. We will return to this later on.

Remark 3.12. Early on, we referred to $\mathbb{R}$ as "complete" in the sense that it had no gaps. While this is related to Definition 3.7, they're not the same. Formally defining the first and distinguishing it from Definition 3.7 would be a long detour. If it is ever unclear which one is being referred to, then I'll try to specify.

3.5 Monotonic Sequences

While the Cauchy Criterion simplifies showing convergence in $\mathbb{R}^n$, is it possible to arrive at even stronger results by restricting our attention to a smaller class of real sequences? The answer is yes.

Many of the sequences in $\mathbb{R}$ we are first introduced to share a common element in the ordering of the elements. Many basic sequences either "grow" or "shrink" at each step. We will refer to this behavior as monotonicity.

Definition 3.8. A sequence $\{x_n\}$ in $\mathbb{R}$ is *monotonically increasing* if $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$.

Example 3.29. The sequence $x_n = n$ in $\mathbb{R}$ is not only monotonically increasing, but also divergent.

Example 3.30. The sequence $x_n = 1 - 1/n$ in $\mathbb{R}$ is monotonically increasing and converges to 1.

Definition 3.9. A sequence $\{x_n\}$ in $\mathbb{R}$ is *monotonically decreasing* if $x_n \geq x_{n+1}$ for all $n \in \mathbb{N}$.

Example 3.31. The sequence $x_n = 1/n$ in $\mathbb{R}$ is monotonically decreasing and converges to 0.

Example 3.32. The sequence $x_n = -n$ in $\mathbb{R}$ is monotonically decreasing and diverges.

If you take some time to think about the examples presented, it may become clear what the convergent sequences have in common. They are bounded! This is not coincidence. If a monotonic sequence is bounded, it "moves in one direction" towards its bound. It can neither "overcome" this bound, nor "change direction", so the only option is that it converges to it! As it turns out, the converse also happens to hold – Any convergent monotonic sequence is bounded.

Theorem 3.8 (Monotone Convergence Theorem)**.** Suppose $\{x_n\}$ in $\mathbb{R}$ is monotonic. Then $\{x_n\}$ converges *if and only if* it is bounded.

Proof.

($\Longrightarrow$) Suppose $\{x_n\}$ is monotonically increasing and bounded.[43] We have $x_n \leq x_{n+1}$, and $x_n \leq x$ for all $n \in \mathbb{N}$, where $x \in \mathbb{R}$ is the supremum of the range of $\{x_n\}$. For all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$x - \varepsilon < x_N \leq x,$$

otherwise $s - \varepsilon$ would be the supremum of the range of $\{x_n\}$. Since $\{x_n\}$ is monotonically increasing,

$$x - \varepsilon < x_n \leq x$$

for all $n \geq N$. But this gives that

$$|x_n - x| < \varepsilon$$

for all $n \geq N$. Therefore $x_n \to x$.

($\Longleftarrow$) Suppose $\{x_n\}$ is bounded and monotonic. All bounded sequences converge (Proposition 3.3).

$\square$

Not only have we shown our result, but in doing so, we showed that a bounded and monotonic sequence will converge to its supremum/infimum.

[43] If $\{x_n\}$ is monotonically decreasing, then the proof is analogous.

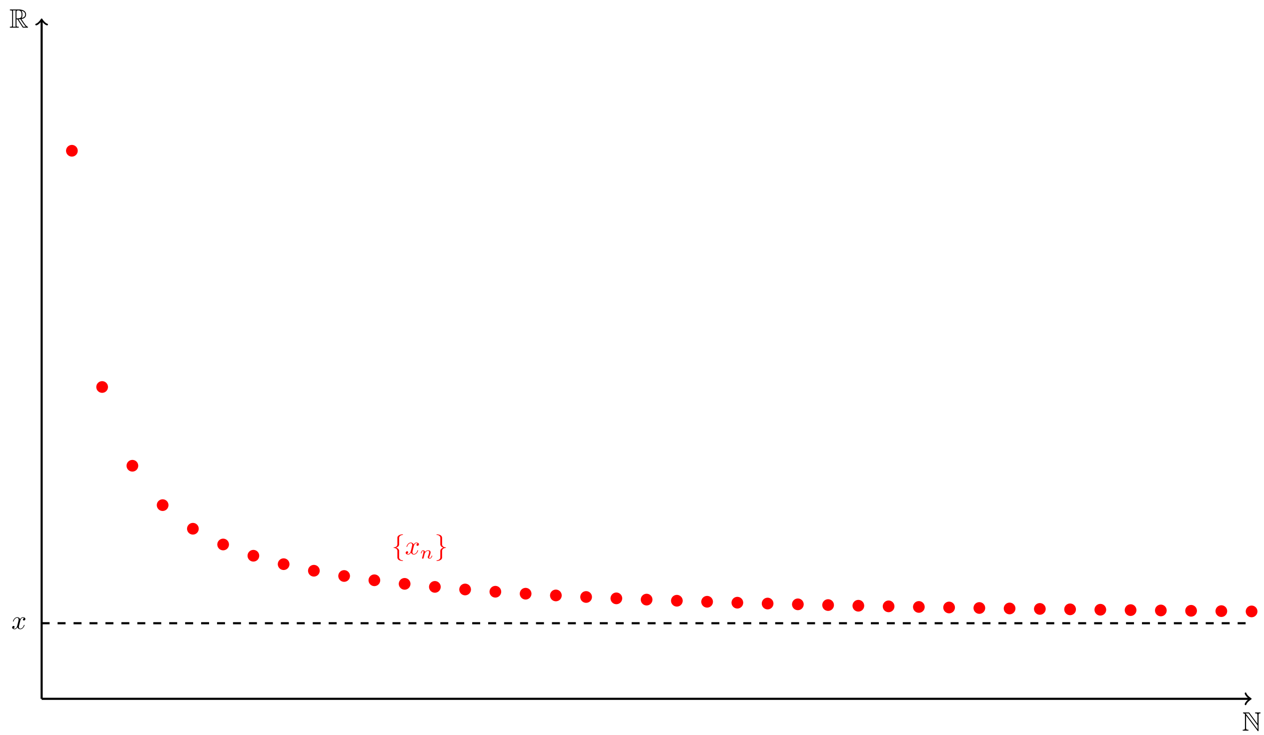

Figure 34: A monotonic and bounded sequence $\{x_n\}$ in $\mathbb{R}$. The sequence converges to the infimum of the sequences range, x.

Example 3.33. We can verify that $x_n = (1 + 1/n)^n$ converges by showing it is monotonically increasing and bounded. We can show it is bounded by using the inequality $\ln(1 + x) \leq x$.

$$\left(1 + \frac{1}{n}\right)^n = \exp\left(n \ln\left(1 + \frac{1}{n}\right)\right) \leq \exp\left(n \cdot \frac{1}{n}\right) = e.$$

To show the sequence is monotonic, we can use the inequality relating the arithmetic mean to the geometric mean (AM-GM inequality)[44] which gives,

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq (x_1 x_2 \ldots x_n)^{1/n}.$$

If we let $x_1 = 1$, $x_2 = x_3 = \cdots = x_{n+1} = 1 + 1/n$, then

$$\begin{aligned}
(x_1 x_2 \ldots x_{n+1})^{\frac{1}{n+1}} &\leq \frac{1}{n+1}(x_1 + x_2 + \cdots + x_{n+1}) \\
\left[1\left(1 + \frac{1}{n}\right)\cdots\left(1 + \frac{1}{n}\right)\right]^{\frac{1}{n+1}} &\leq \frac{1}{n+1}\left[1 + \left(1 + \frac{1}{n}\right) + \ldots + \left(1 + \frac{1}{n}\right)\right] \\
\left(1 + \frac{1}{n}\right)^{\frac{n}{n+1}} &\leq \frac{1 + n\left(1 + \frac{1}{n}\right)}{n+1} \\
&= 1 + \frac{1}{n+1} \\
\left(1 + \frac{1}{n}\right)^n &\leq \left(1 + \frac{1}{n+1}\right)^{n+1}.
\end{aligned}$$

Hence $x_n \leq x_{n+1}$. We've only shown that $\{x_n\}$ converges. In order to show that $x_n \to e$, we would need to show that e is a least upper bound.

[44]This is a very useful inequality to know when working with proves involving boundedness and monotonicity.

Remark 3.13 (Inequalities)**.** Proofs involving inequalities can be an acquired taste. They often rely on really clever algebraic manipulations and using a set of common inequalities such as the AM-GM inequality. For this reason, I think they're a complete pain in the ass. Instead of relying on mathematical intuition and insight, you just need to make some guess about how to manipulate an equation.

Remark 3.14. It's worth recapping what we've done with sequences up until this point, and the motivation behind our results. One of the main goals when working with sequences is proving/disproving convergence. Unfortunately, showing a sequence converges only using the definition of convergence can be hard, as it requires us to know the limit of a function. If we know our sequence is in a specific type of metric space, or has a certain property, then you can use Theorem 3.6, Theorem 3.7, and/or Theorem 3.8 to verify convergence!

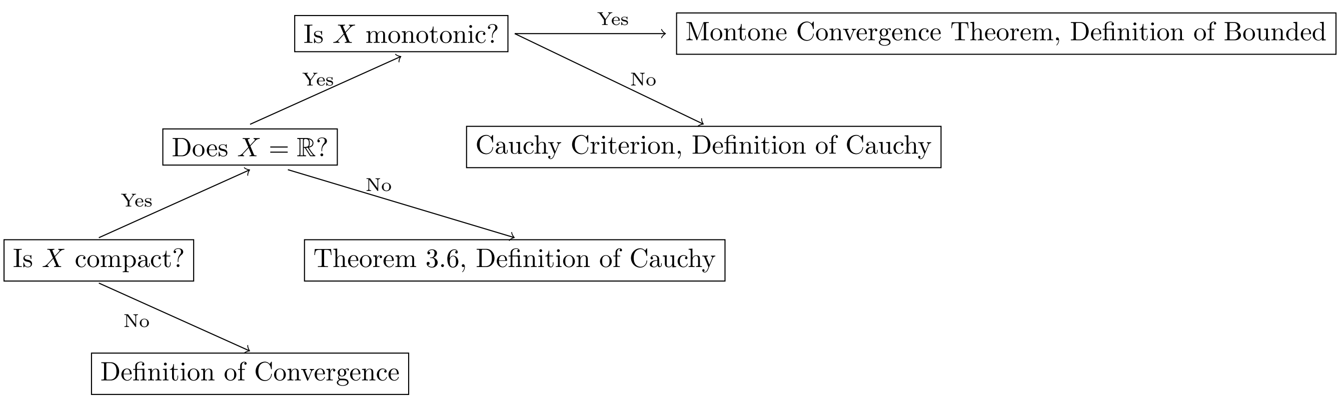

Figure 35: If $\{x_n\}$ is a convergent sequence in X, how do we prove that $\{x_n\}$ converges?

3.6 Sequences and Infinity

We have yet to discuss what happens when divergent sequences "go off" to infinity.

Definition 3.10. Let $\{x_n\}$ be a sequence in $\mathbb{R}$. If for all $M \in \mathbb{R}$ there exists an $N \in \mathbb{N}$ such that $x_n \geq M$, we write

$$\lim_{n\to\infty} x_n = \infty.$$

Similarly, if for every $M \in \mathbb{R}$ there exists an $N \in \mathbb{N}$ such that $n \geq N$ implies $x_n \leq M$, we write

$$\lim_{n\to\infty} x_n = -\infty.$$

These definition make a slight abuse of notation, as they refer to the limit of a sequence despite the sequence in question diverging.

Example 3.34. Let $x_n = n$ be a sequence in $\mathbb{R}$. For all $M \in \mathbb{R}$, let $N = M + 1$. We have

$$s_n \geq s_N = N = M + 1 > M$$

for all $n \geq N$. Therefore $x_n \to \infty$.

Remark 3.15 (Extended Real Numbers)**.** If we were instead in the extended real numbers (Definition 1.19) $\overline{\mathbb{R}}$, then the type of sequences given in Definition 3.10 would converge, because $\{-\infty, \infty\} \subset \overline{\mathbb{R}}$. In fact, because every set in $\overline{\mathbb{R}}$ is bounded, many of the results pertaining to sequences would be more general in $\overline{\mathbb{R}}$. Any proposition or theorem that requires a sequence to be bounded in $\mathbb{R}$, would not make this requirement in $\overline{\mathbb{R}}$, as it would be implicitly met.

3.7 lim sup and lim inf

It can be insightful to study how the bounds of sequences change as n grows. For sequences in $\mathbb{R}$, doing this illuminates interesting relationships between converge, bounds, limits of bounds, and subsequential limits.

A sequence is bounded if its range is a bounded set. If this sequence is in $\mathbb{R}$, then we know something very powerful about the range of the sequence – its supremum and infimum exist in $\mathbb{R}$ (Theorem 1.2). The same can be said for the supremum and infimum of the set of all elements in the range of x_n that have yet to be realized, $\{x_m \mid m \geq n\}$. The set $\{x_m \mid m \geq n\}$ is bounded above or below by x_n, so the supremum or the infimum (or both) of $\{x_m \mid m \geq n\}$ exist. We will define the supremum and infimum of a set using this set of values $\{x_m \mid m \geq n\}$.

Definition 3.11. Let x_n be a sequence in X. The *infimum of a sequence* is

$$\inf x_n = \inf_{m \geq n} x_m = \inf\{x_m \mid m \geq n\}.$$

Definition 3.12. Let x_n be a sequence in X. The *supremum of a sequence* is

$$\sup x_n = \sup_{m \geq n} x_m = \sup\{x_m \mid m \geq n\}.$$

The supremum and infimum of a sequence form their own sequences as the set $\{x_m \mid m \geq n\}$ depends on the value of n. Figure 36 illustrate the supremum and infimum of a sequence in $\mathbb{R}$.

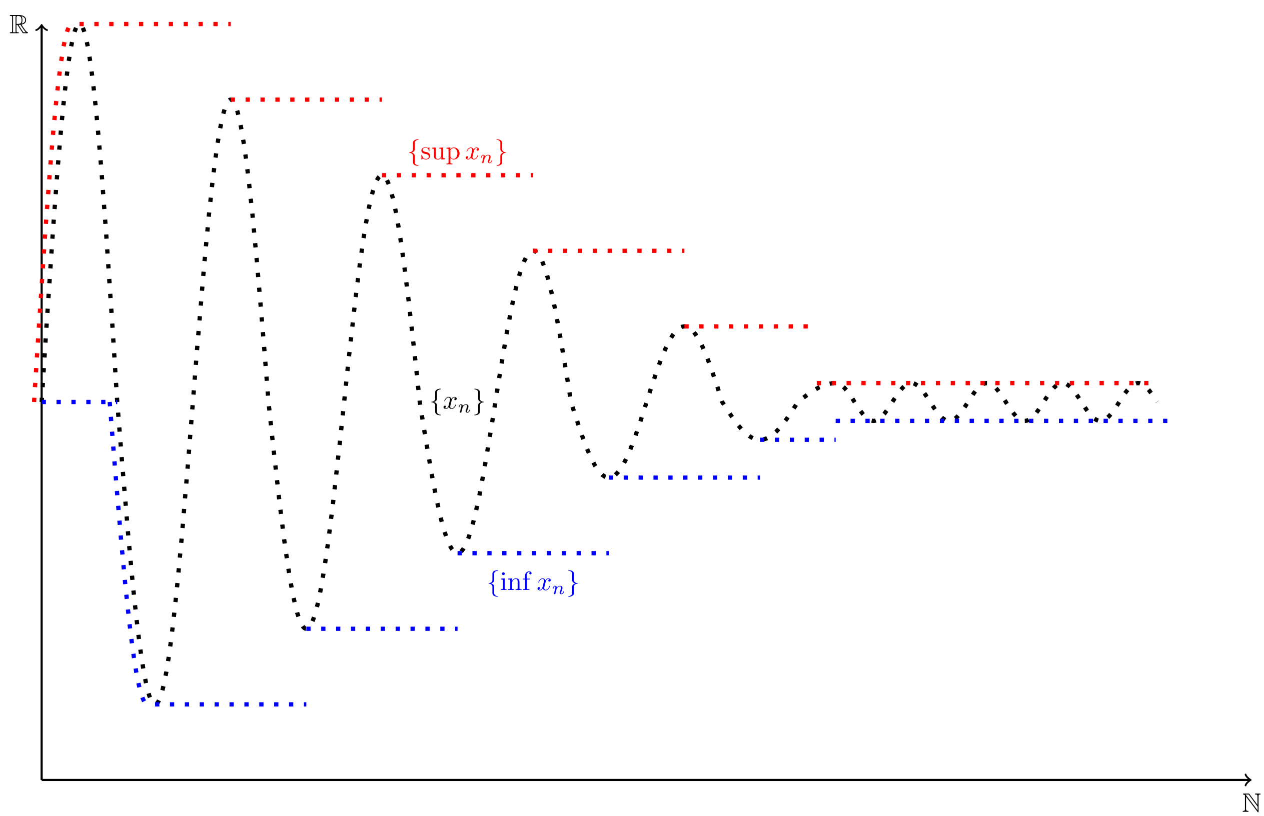

Figure 36: A sequence $\{x_n\}$ in $\mathbb{R}$, with the sequence $\{\sup x_n\}$ and $\{\inf x_n\}$ shown in red and blue, respectively. For any $n \in \mathbb{N}$, we take the supremum and infimum of all the points of $\{x_m \mid m \geq n\}$. Graphically this means we take the supremum and infimum of all the points to the right of a certain value of $\mathbb{N}$.

Example 3.35. Let $x_n = 1/n$ be a sequence in $\mathbb{R}$. We have $\{x_m \mid m \geq n\} = \{1/m \mid n \geq m\}$. This gives $\sup x_n = 1/n$ and $\inf x_n = 0$, each being its own sequence in $\mathbb{R}$.

Example 3.36. If x_n be a monotonically increasing sequence, then $\inf x_n = x_n$ as $x_n \le x_m$ for all $m \ge n$. Similarly, if x_n is monotonically decreasing, then $\sup x_n = x_n$.

Example 3.37. If $x_n = \sin x$, then $\inf x_n = -1$ and $\sup x_n = 1$ for all $n \in \mathbb{N}$.

Now we will define the limits of $\sup x_n$ and $\inf x_n$.

Definition 3.13. Let $\{x_n\}$ be a sequence in X. The *limit inferior* of $\{x_n\}$ is the limit of $\{\inf x_n\}$, and is written as $\liminf\limits_{n\to\infty} x_n$ or $\varliminf\limits_{n\to\infty} x_n$.

Definition 3.14. Let $\{x_n\}$ be a sequence in X. The *limit superior* of $\{x_n\}$ is the limit of $\{\sup x_n\}$, and is written as $\limsup\limits_{n\to\infty} x_n$ or $\varlimsup\limits_{n\to\infty} x_n$.

Example 3.38. The sequence $x_n = 1/n$ has a limit inferior of 0, and a limit superior of 0.

Example 3.39. The sequence $x_n = \sin x$ has a limit inferior of -1, and a limit superior of 1.

You may have noticed two things: the sequences $\{\sup x_n\}$ and $\{\inf x_n\}$ seem to always converge, and sometimes the limit inferior and limit superior are equal. The observations are consequences of our next results.

Proposition 3.5. If $\{x_n\}$ is a bounded sequence in $\mathbb{R}$, then $\liminf\limits_{n\to\infty} x_n$ and $\limsup\limits_{n\to\infty} x_n$ exist. That is, the sequences $\{\inf x_n\}$ and $\{\sup x_n\}$ converge.

Proof. We will prove the result for the limit superior.[45] Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}$. By the least-upper-bound property, $\sup x_n$ exists for all $n \in \mathbb{N}$, as the range of x_n is bounded. The sequence $\sup x_n$ is bounded because x_n is bounded. By the Monotone Convergence Theorem, is suffices to show that $\sup x_n$ is monotonically decreasing. We have $\{x_n \mid m \ge n+1\} \subset \{x_n \mid m \ge n\}$, so any upper bound of $\{x_n \mid m \ge n\}$ is an upper bound of $\{x_n \mid m \ge n+1\}$. This includes the least-upper-bound, which in this case is $\sup x_n$. By the definition of the least-upper-bound, $\sup x_{n+1}$ is less than all other upper-bounds of $\{x_n \mid m \ge n+1\}$, therefore $\sup x_n \ge \sup x_{n+1}$. This gives that $\{\sup x_n\}$ is monotonically decreasing. ☐

Remark 3.16 (Monotonicity and Limits of Bounds)**.** The proof of the last result took advantage of the fact that $\sup x_n$ was monotonically decreasing and bounded, allowing the use of the Monotonic Convergence Theorem (Theorem 3.8). When we proved this result, we showed that a monotonically decreasing sequence will converge to its infimum. This means that $\{\sup x_n\}$ converges to its infimum. For this reason the limit superior is sometimes written as $\inf \sup x_n$. Similarly, the limit inferior can be written as $\sup \inf x_n$.

Proposition 3.6. Let $\{x_n\}$ be a bound subsequence in $\mathbb{R}$. Then

1. There exist subsequences which converges to $\limsup x_n$ and $\liminf x_n$.
2. For all $c \in \mathbb{R}$, if there exists a subsequence which converges to c, then $\liminf x_n \le c \le \limsup x_n$.

If we let E be the set of subsequential limits corresponding to a bounded $\{x_n\}$ in $\mathbb{R}$, then Proposition 3.6 concludes that $\{\liminf x_n, \limsup x_n\} \subset E$, $\liminf x_n = \inf E$, and $\limsup x_n = \sup E$.

Example 3.40. Let $x_n = 1/n$ be a sequence in $\mathbb{R}$. It is convergent and bounded. We know from Example 3.37 that $\liminf x_n = \limsup x_n = 0$. By Proposition 3.6, 0 is the only subsequential limit of $\{x_n\}$. We already knew this though, because that is precisely what Theorem 3.3 asserts.

Example 3.41. For $x_n = \sin x$ in $\mathbb{R}$, the set of subsequential limits is a subset of $[-1, 1]$.

Proposition 3.7. Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}$. The sequence converges *if and only if*

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n.$$

Proof.

[45] The proof for the limit inferior is very similar.

($\Longrightarrow$) Suppose $\{x_n\}$ converges to $x \in \mathbb{R}$. By Theorem 3.3, every subsequence of $\{x_n\}$ converges to x. By Proposition 3.6, $\liminf x_n$ and $\limsup x_n$ are subsequential limits of $\{x_n\}$. This means they both are x, as x is the only subsequential limit of $\{x_n\}$. Therefore $\liminf x_n = \limsup x_n$.

($\Longleftarrow$) Assume that $\limsup x_n = \liminf x_n = x$. For all $\varepsilon > 0$,

$$x - \varepsilon < \liminf x_n \le \limsup x_n < x + \varepsilon.$$

This allows us to choose an $N_1, N_2 \in \mathbb{N}$ such that

$$s - \varepsilon < x_n < s + \varepsilon$$

for all $n \ge \max\{N_1, N_2\}$. This gives that $|x_n - x| < \varepsilon$ for all such n, so $x_n \to x$.

□

3.8 Series

Now we consider series. As we'll see, everything we did with sequences caries over nicely to series. In fact, working with series will be even simpler.

A series is simply a special type of sequence which results from summing each element of a different sequence.

Definition 3.15. Given a sequence $\{x_n\}$ in a metric space X, we have a *series* in the form of

$$\sum_{n=1}^{\infty} x_n.$$

Example 3.42 (Geometric Series)**.** If we let $x_n = 1/2^n$, then

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots .$$

In order to define the convergence of a series, we will look at the behavior of the first n terms. That is, we are interested in

$$\begin{aligned} s_1 &= \sum_{k=1}^{1} x_k = x_1 \\ s_2 &= \sum_{k=1}^{2} x_k = x_1 + x_2 \\ s_3 &= \sum_{k=1}^{3} x_k = x_1 + x_2 + x_3 \\ s_4 &= \sum_{k=1}^{4} x_k = x_1 + x_2 + x_3 + x_4 \\ &\vdots \end{aligned}$$

If the sequence $\{s_n\}$ converges, then we will say the series $\sum_{n=1}^{\infty} x_n$ converges.

Definition 3.16. Let $\sum_{n=1}^{\infty} x_n$ be a series in X. Define the *partial sums* of the series to be the sequence

$$s_n = \sum_{k=1}^{n} x_k.$$

A series *converges* to $x \in X$ if the sequence $\{s_n\}$ converges to $x \in X$. In this case, we write

$$\sum_{n=1}^{\infty} x_n = x.$$

Remark 3.17 (Everything from Sequences Carries Over)**.** Every single result we established for sequences can be formulated in terms of series, because every single series can be expressed as a sequence of partial sums.

Showing a series converges using this definition is unnecessarily difficult. In $\mathbb{R}$ we can use a reformulation of the Cauchy Criterion

Theorem 3.9 (Cauchy Criterion for Series)**.** Let $\sum_{n=1}^{\infty} x_n$ be a sequence in $\mathbb{R}$. The series converges *if and only if* for all $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that

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

for all $m \geq n \geq N$.

Proof. This follows from the fact that

$$|s_n - s_m| = \left| \sum_{k=1}^{n} x_n - \sum_{k=1}^{m} x_m \right| = \left| \sum_{k=n}^{m} x_k \right|.$$

□

Corollary 3.4. If $\sum_{n=1}^{\infty} x_n$ converges, then $x_n \to 0$.

Proof. Let $m = n$ and use Theorem 3.9 □

Remark 3.18 (No Limit Needed!)**.** Theorem 3.9 is a direct consequence of the Cauchy Criterion, so we don't need to know the limit of a convergent sequence to prove it converges.

Example 3.43. The converse of Corollary 3.4 is not true! The harmonic series $\sum_{n=1}^{\infty} 1/n$ diverges despite the fact that $1/n \to 0$.

Example 3.44. The series $\sum_{n}^{\infty} 1/x!$ converges to e in $\mathbb{R}$. This sequence does not converge in $\mathbb{Q}$, as $e \notin \mathbb{Q}$.

3.9 Tests for Convergent Series

Whenever we work with a series $\sum_{n=1}^{\infty} x_n$, we want to find some way to relate a sequence involving x_n to the convergence of the series. There are three tests we can use to verify that a sequence in $\mathbb{R}$ converges by doing just this: the comparison test, the ratio test, and the root test. Another benefit of these tests, is none require us to know the limit of a series to prove it converges!

Theorem 3.10 (The Comparison Test)**.** Let $\sum_{n=1}^{\infty} x_n$ be a series in $\mathbb{R}$. If $|x_n| \leq y_n$ for $n \geq N_0$, where N_0 is some fixed integer, and if $\sum_{n=1}^{\infty} c_n$ converges, then $\sum_{n=1}^{\infty} x_n$ converges.

Proof. For all $\varepsilon > 0$, there exists an $N \geq N_0$ such that $m \geq n \geq N$ implies

$$\sum_{k=n}^{m} y_k \leq \varepsilon,$$

according to The Cauchy Criterion. Therefore we have

$$\left| \sum_{k=n}^{m} x_k \right| \leq \sum_{k=n}^{m} |x_k| \leq \sum_{k=n}^{m} y_k \leq \varepsilon,$$

so again by the Cauchy criterion, $\sum_{k=n}^{m} x_k$ converges. □

Example 3.45. The geometric series $\sum_{n=1}^{\infty} 1/3^n$ converges to $3/2$. We have

$$\frac{1}{3^n + n} < \frac{1}{3^n}$$

for all $n \in \mathbb{N}$, so the series $\sum_{n=1}^{\infty} 1/(3^n + n)$ converges by the Comparison Test.

Example 3.46. The harmonic series $\sum_{n=1}^{\infty} 1/n$ diverges. We have

$$\frac{n}{n^2 - \cos^2 n} > \frac{n^2}{n} = \frac{1}{n}.$$

Therefore by the converse of Theorem 3.10, the series $\sum_{n=1}^{\infty} \frac{n}{n^2 - \cos^2 n}$ diverges.

Theorem 3.11 (The Ratio Test)**.** If $\sum_{n=1}^{\infty} x_n$ is a series in $\mathbb{R}$, then

1. the series converges if $\lim_{n\to\infty} |x_{n+1}/x_n| < 1$;
2. the series diverse if $|x_{n+1}/x_n| \geq 1$ for all $n \geq N_0$ for a fixed $N_0 \in \mathbb{N}$.

Proof. Suppose $\lim |x_{n+1}/x_n| = \beta < 1$. For ε This means we can find an integer $N \in \mathbb{N}$ such that $|x_{n+1}/x_n| < \beta$ for all $n \geq N$. This can be rewritten as $|x_{n+1}| < \beta |x_n|$. If we do this for $n = N, N+1, N+2, \ldots, N+k$, then

$$\begin{aligned}
&|x_{N+1}| < \beta |x_N|, \\
&|x_{N+2}| < \beta |x_{N+1}| < \beta^2 |x_N|, \\
&\qquad\vdots \\
&|x_{N+k}| < \beta^p |a_N|.
\end{aligned}$$

This gives

$$\sum_{k=N+1}^{\infty} |x_k| = \sum_{k=1}^{\infty} |x_{N+k}| < \sum_{k=1}^{\infty} \beta^k |x_N| = |x_N| \sum_{k=1}^{\infty} \beta^k = |x_N| \frac{\beta}{1-\beta}.$$

Therefore by the Comparison Test, $\sum_{n=1}^{\infty} x_k$ converges, because

$$\sum_{n=1}^{\infty} |x_n| \leq \sum_{k=N+1}^{\infty} |x_k|.$$

Now suppose that $\limsup |x_{n+1}/x_n| = \beta > 1$. This means that $|x_{n+1}| > |x_n|$ for a sufficiently large n, meaning $\lim\limits_{n\to\infty} x_n \neq 0$. By the converse of Corollary 3.4, the series diverges. □

Example 3.47. Does $\sum_{n=0}^{\infty} n!/5^n$ converge of diverge?

$$\lim_{n\to\infty} \left| \frac{x_{n+1}}{x_n} \right| = \lim_{n\to\infty} \left| \frac{(n+1)!}{5^{n+1}} \frac{5^n}{n!} \right| = \lim_{n\to\infty} \frac{(n+1)!}{5n!} = \lim_{n\to\infty} \frac{(n+1)n!}{5n!} = \lim_{n\to\infty} \frac{n+1}{5} = \infty$$

By the Ratio Test, the series diverges.

Example 3.48. If we have $\lim_{n\to\infty} |x_{n+1}/x_n| = 1$, then we aren't able to conclude anyhting about the convergence of a series. For example $\sum_{n=1}^{\infty} 1/n$ diverges, whereas $\sum_{n=1}^{\infty} 1/n^2$ converges. In both cases $\lim_{n\to\infty} |x_{n+1}/x_n| = 1$.

Theorem 3.12 (The Root Test)**.** If $\sum_{n=1}^{\infty} x_n$ is a series in $\mathbb{R}$, then

1. the series converges if $\lim_{n\to\infty} \sqrt[n]{|x_n|} < 1$;
2. the series diverges if $\lim_{n\to\infty} \sqrt[n]{|x_n|} > 1$;

3. we cannot determine anything if $\lim_{n\to\infty} \sqrt[n]{|x_n|} = 1$.

Proof. Let $\alpha = \lim_{n\to\infty} \sqrt[n]{|x_n|}$. If $\alpha < 1$, we can find a $\beta \in (\alpha, 1)$, and an $N \in \mathbb{N}$ such that $\sqrt[n]{|x_n|} < \beta$ for $n \geq N$. This inequality can be expressed as $|x_n| < \beta^n$. Because $\sum_{n=0}^{\infty} \beta^n$ converges, the Comparison Test gives that $\sum_{n=N}^{\infty} |x_n|$ converges. Since

$$\sum_{n=1}^{\infty} |x_n| = \sum_{n=N}^{N-1} |x_n| + \sum_{n=N}^{\infty} |x_n|,$$

the series $\sum_{n=1}^{\infty} |x_n|$ must converge, as the first sum is finite.

If $\alpha > 1$, then $|x_n| > 1^n = 1$, so $\lim\limits_{n\to\infty} x_n \neq 0$. By Corollary 3.4, the series diverges.

The series $\sum_{n=1}^{\infty} 1/n$ and $\sum_{n=1}^{\infty} 1/n^2$ both have $\alpha = 1$, despite the prior diverging and the latter converging. Therefore if $\alpha = 1$, we cannot make any statement about convergence. □

Example 3.49. Does the series $\sum_{n=0}^{\infty} \left(\frac{5n-3n^3}{7n^3+3}\right)^n$ converge or diverge?

$$\lim_{n\to\infty} \sqrt[n]{|x_n|} = \lim_{n\to\infty} \left|\left(\frac{5n-3n^2}{7n^3+2}\right)^n\right|^{1/n} = \lim_{n\to\infty} \left|\frac{5n-3n^2}{7n^3+2}\right| = \frac{3}{7} < 1$$

By the Root Test, the series converges.

Remark 3.19 (Slightly More General Tests)**.** We can make the Ratio Test and Root Test a bit more powerful by using lim sup instead of just the limit. The proofs would be modified slightly. This will rarely make a difference.

Remark 3.20 (Ratio vs. Root)**.** The Ratio Test is much easier to use than the Root Test. Calculating ratios tends to be way easier than the nth root of a value. That being said, the Root Test is more powerful in the sense that the Ratio Test will always agree with the Root Test. Furthermore, there are cases when the Ratio Test is inconclusive, but the Root Test is not. A good example of this is found in Example 3.35 in Rudin (1976).

3.10 Exercises

rudin, 3.4 part b
constant sequences in $\mathbb{Z}$
subsequential limits closed
series of nonnegative terms converges iff partial sums bounded

4 Continuity

We spent a fair amount of time studying what it means for a sequence or series to get "arbitrarily close" to some point. We now want to do something similar with more general functions, and rigorously define continuity. Continuity makes nearly all of analysis possible, and without it we would be hopeless. Knowing the properties related to continuity will prove useful when proving many results, as many results will pertain exclusively to continuous functions.

4.1 Limits of Functions

We begin by defining what a limit of a function is.

Definition 4.1. Let X and Y be metric spaces, and $E \subset X$. Suppose there is some function $f : E \to Y$, and a limit point p of E and some $L \in Y$ satisfying: for all $\varepsilon > 0$ there exists a corresponding $\delta > 0$ such that $d_Y(f(x), L) < \varepsilon$ for all points $x \in E$ which satisfy $d_X(x, p) < \delta$. We say that L is *the limit of f as x approaches p*, and write

$$\lim_{x \to p} f(x) = L.$$

This definition seems much more complicated than that given for convergence, but the concepts conveyed are fairly similar. It may be helpful to return to the somewhat lame example of playing a game with your friend involving ε. Instead of a sequence you now play with some function $f : E \to Y$, and a fixed limit point $p \in E$.[46] Your friend gives you a value of ε, and he challanges you to find a δ such that $d_Y(f(x), L) < \varepsilon$ whenever $d_X(p, x) < \delta$. Note that we never require that $f(x) = L$ (see Figure 37). Our next several examples will make this more concrete.

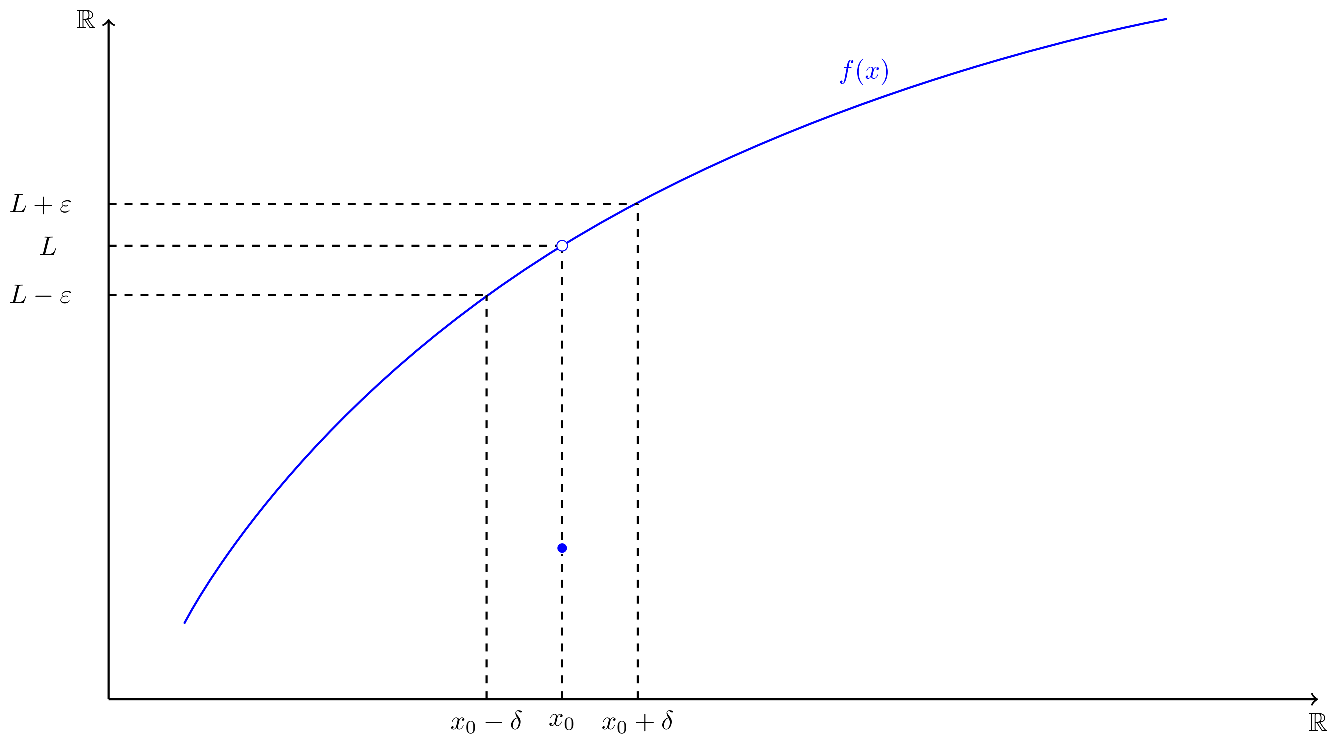

Figure 37: We have $\lim_{x \to x_0} f(x) = L$ for a real function. No matter how small ε gets, we can find a δ such that $|f(x) - L| < \varepsilon$ for all x satisfying $|x - x_0| < \delta$.

[46]It's important that p is a limit point. This means every neighborhood of p has infinitely many points, so we can always get arbitrarily close to it.

Example 4.1. Suppose we define $f : \mathbb{R} \to \mathbb{R}$ as some linear function $f(x) = mx + b$. For $L = mx_0 + b$ and $x_0 \in \mathbb{R}$, your friend gives you $\varepsilon = 0.1$. You need to find a δ such that

$$d_{\mathbb{R}}(f(x), L) = |mx + b - (mx_0 + b)| = |m(x - x_0)| = |m||x_0 - x| < 0.1.$$

Therfore you let $\delta = 0.1/|m|$. Instead of letting him keep naming ε, you tell him that you'll just set $\delta = \varepsilon/|m|$. This way whenever $d_{\mathbb{R}}(f(x), L) < \varepsilon$ we have

$$d_{\mathbb{R}}(f(x), L) = |mx + b - (mx_0 + b)| = |m(x - x_0)| = |m||x_0 - x| = |m| \cdot d_{\mathbb{R}}(x, x_0) < |m| \cdot \frac{\varepsilon}{|m|} = \varepsilon$$

whenever $d_{\mathbb{R}}(x, x_0) < \delta$. Therefore $\lim_{x\to x_0} f(x) = L$.

Example 4.2. Define $f : \mathbb{R} \to \mathbb{R}$ as

$$\begin{cases} 1 \text{ if } x > 0 \\ -1 \text{ if } x < 0 \end{cases}.$$

We can show that the limit at $x_0 = 0$ fails to exist by finding one such ε where Definition 4.1 does not holds. Let $\varepsilon = 1/2$, and suppose for contradiction $\lim_{x\to x_0} f(x) = L$. We have that $|f(x) - L| < \varepsilon = 1/2$ for all x such that $|x - 0| < \delta$. But we also have that

$$\begin{aligned} 2 &= |1 - (-1)| \\ &= |f(\delta/2) - f(-\delta/2)| \\ &= |f(\delta/2) - L + L - f(-\delta/2)| \\ &\leq |f(\delta/2) - L| + |L - f(-\delta/2)| \\ &\leq \frac{1}{2} + \frac{1}{2} \\ &= 1, \end{aligned}$$

which is a clear contradiction.

Example 4.3. Suppose we define $f : \mathbb{R} \to \mathbb{R}$ as $f(x) = \sin x$. We can verify that $\lim_{x\to 0} f(x) = 1$ despite the fact that $f(x)$ is undefined at 0. If we let $\delta = \sqrt{\varepsilon}$ and restrict our attention to the interval $(-\pi/2, \pi/2)$,[47] then for all $\varepsilon > 0$ we have

$$d_{\mathbb{R}}(f(x), L) = \left|\frac{\sin x}{x} - 1\right| < 1 - \cos x < 2\sin^2 \frac{x}{2} < \frac{x^2}{2} = \frac{|x - 0|^2}{2} = \frac{d_{\mathbb{R}}(x, 0)^2}{2} < \frac{(\sqrt{\varepsilon})^2}{2} < \varepsilon$$

whenever $d_{\mathbb{R}}(x, 0) < \delta$.

As you may suspect, their is a very strong link between the convergence of a sequence and the limit of a function. If we have $\lim_{x\to p} f(x) = L$, then p is a limit point of $E \subset X$ as defined in Definition 4.1. This limiting process lets x become arbitrarily close to p, much like a sequence would to its limit. As it turns out, there is a way to reformulate limits in terms of sequences if we use a sequence $\{p_n\}$ in E which converges to p.[48]

Theorem 4.1. Let X and Y be metric spaces, and $E \subset X$. Suppose there is some function $f : E \to Y$, and a limit point p of E. Then $\lim_{x\to p} f(x) = L$ *if and only if* $f(p_n) \to L$ (alternatively written as $\lim_{n\to\infty} f(p_n) = L$) for every non-constant sequence $\{p_n\}$ in E such that $p_n \to p$.[49]

Proof.

($\Longrightarrow$) Suppose $\lim_{x\to p} f(x) = L$. By Theorem 3.1, we can choose some arbitrary sequence $\{p_n\}$ in E such that $p_n \to p$. For all ε there exists a $\delta > 0$ such that $d_Y(f(x), L) < \varepsilon$ for all $x \in E$ satisfying $d_X(x, p) < \delta$. By the convergence of p_n, there exists an $N \in \mathbb{N}$ such that $d_X(p_n, p) < \delta$ for all $n > N$.[50] Therefore, for all $n > N$, we have $d_Y(f(p_n), L) < \varepsilon$. This gives

$$\lim_{n\to\infty} f(p_n) = L.$$

[47] This insures that $0 < \cos x < \sin x/x < 1$.

[48] Such a sequence is guaranteed to exist because p is a limit point of E (see Theorem 3.1).

[49] The use of the same limit notation here may obscure the fact that $\{f(p_n)\}$ is a sequence such that $f(p_n) \to L$.

[50] We take the δ from the definition of a limit of a function and we use it as the ε in the definition of a limit of a sequence.

($\Longleftarrow$) We will prove the contrapositive. Suppose that $\lim_{x\to p} f(x) \neq L$. There exists some $\varepsilon > 0$ such that for all $\delta > 0$, there exists some point $x \in E$ (which depends on δ) for which $d_Y(f(x), L) \geq \varepsilon$ but $d_X(x, p) < \delta$.[51] If we define $\delta_n = 1/n$ for all $n \in \mathbb{N}$, then there is sequence $\{p_n\}$ in E such that $d(p_n, p) < \delta_n = 1/n$. We have $p_n \to p$ despite the fact that $\lim_{n\to\infty} f(p_n) \neq L$.

□

Example 4.4. For the real function $f(x) = x$, it's clear that $\lim_{x\to 0} f(x) = 0$. By Theorem 4.1 we know that

$$\lim_{n\to\infty} f(1/n) = 0,$$

as $1/n \to 0$. Furthermore we know this is the case for any $\{x_n\}$ which converges to 0.

So what's the big deal with writing limits in terms of sequences? This just seems like another complex relationship to remember. Theorem 4.1 is important because it allows us to prove many results involving limits (and later continuity) using properties we already know from sequences. For example, when using Theorem 4.1 we know right away that limits of functions are unique because the limits of sequences are unique.

Corollary 4.1. If f has a limit at p, the limit is unique.

Proof. If $\lim_{x\to p} f(x) = L$, then by Theorem 4.1 the sequence $\{f(p_n)\}$ converges to L for all $\{p_n\}$ which converge to p. By Proposition 3.2, L is unique. □

The proof really amounts to nothing more than saying "combine Theorem 4.1 and Proposition 3.2". We can combine Theorem 4.1 with Theorem 3.2 to arrive at familiar properties of functions.

Theorem 4.2. Let X and Y be metric spaces, and $E \subset X$. Suppose there are some functions $f : E \to Y$ and $g : E \to Y$, and a limit point p of E. If $\lim_{x\to p} f(x) = A$ and $\lim_{x\to p} f(x) = B$, then

1. $\lim_{x\to p}(f + g)(x) = A + B$;
2. $\lim_{x\to p}(fg)(x) = AB$;
3. $\lim_{x\to p}(f/g)(x) = A/B$ if $B \neq 0$.

Proof. This follows from Theorem 4.1 and Theorem 3.2. □

4.2 Continuous Functions

The definition of a limit of a function as given in Definition 4.1 has two quirks. The first, which has been acknowledged twice, is that a function needn't actually take on the value of its limit at a point. We never require $f(p) = L$. Secondly, we do not require that $p \in E$, we instead require that p is a limit point of p. This means that not only can f be undefined at p, but we shouldn't even be suprised that it is undefined at p when $p \notin E$, as $f : E \to Y$. We want to rule both of these out. Essentially, we want to define "nice" functions at a point p to be those where $\lim_{x\to p} f(p) = L$. This is how we get our definition of continuity.

Definition 4.2. Let X and Y be metric spaces, $E \subset X$, $p \in E$, and $f : E \to Y$. Then f is *continuous at p* if for all $\varepsilon > 0$, there exists a corresponding $\delta > 0$ such that

$$d_Y(f(x), f(p)) < \varepsilon$$

for all points $x \in E$ which satisfy $d_X(x, p) < \delta$. If f is continuous at every point of E, then f is continuous (on E).

This definition is illustrated in Figure 38.

Remark 4.1. Go back and review Remark 2.4. This issue came up with sequences. We saw sequences that converged in one metric space, but not another. Similarly, a function can be continuous in one metric space but not another. It also can be continuous on one subset of a metric space, but not another. If a subset $E \subset X$ is never specified, then we assume that $E = X$. This is why we usually consider all of $\mathbb{R}$ when determining if a real function is continuous.

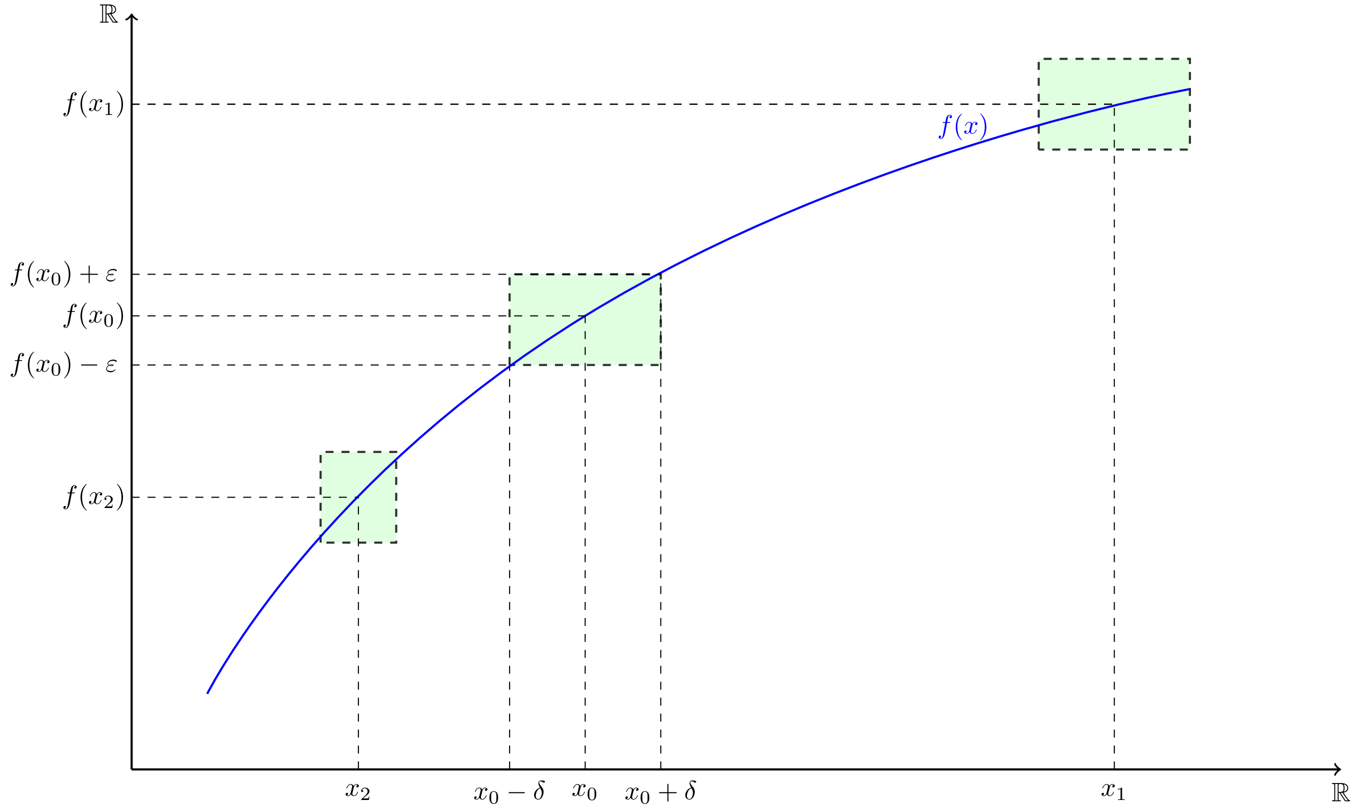

Figure 38: The function f is continuous at x_0, x_1, and x_2. Given some $\varepsilon > 0$, there exist a delta such that $|f(x) - f(y)| < \varepsilon$ for all $|x - y| < \delta$. Each green box has height of 2ε, and the width for the 2δ at the corresponding point. Note that the value of δ is the same for x_0 and x_1, but different for x_2. For a fixed ε, the delta for one point may differ depending on the value of x. If some value of δ is too large, then the box would be too wide, and the function $f(x)$ would "escape" from the top and/or bottom of the box. This graphical feature would imply a violation of continuity, so we need to shrink δ until the function "escapes" from the sides (which is what we did for x_2).

Example 4.5. Suppose we define $f : \mathbb{R} \to \mathbb{R}$ as some linear function $f(x) = mx + b$. Let $p \in \mathbb{R}$. We can mimic our work in Example 4.1 to show this function is continuous on all of $\mathbb{R}$. For $p \in \mathbb{R}$ satisfying $|x - p| < \varepsilon/|m|$, we have

$$d_{\mathbb{R}}(f(x), f(p) = |mx + b - (my + b)| = |m(x - p)| = |m||x - p| = |m| \cdot d_{\mathbb{R}}(x, p) < |m| \cdot \frac{\varepsilon}{|m|} = \varepsilon.$$

Therefore f is continuous at p, for all $p \in \mathbb{R}$

Example 4.6. Let $f : \mathbb{R} \to \mathbb{R}$ be define as

$$\begin{cases} 1 \text{ if } x > 0 \\ -1 \text{ if } x \leq 0 \end{cases}.$$

This function is continuous on the set $\mathbb{R}\backslash\{0\}$. It fails to be continuous on all of $\mathbb{R}$, as there is a "jump" from -1 to 1 at $x = 0$.

Example 4.7. The function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^2$ is continuous on $\mathbb{R}$. Let $p \in \mathbb{R}$. For all $\varepsilon > 0$, define $\delta = \min\{1, \varepsilon/(2|p| + 1)\}$. We have

$$|f(x) - f(p)| = |x^2 - p^2| = |x - p| \cdot |x + p| < \delta(2|p| + 1) = \varepsilon,$$

[51] This is just the negation of Definition 4.1.

for $x \in \mathbb{R}$ which satisfy $|x - p| < \delta$. Therefore f is continuous at p, for all $p \in \mathbb{R}$.

Example 4.8 (Thomae's function)**.** Define the real function $f : [0, 1] \to [0, 1]$ as

$$f(x) = \begin{cases} \frac{1}{q} \text{ if } x = \frac{p}{q} \text{ for } p, q \in \mathbb{Z} \\ 0 \text{ if } x \neq \mathbb{Q} \end{cases} \qquad ,$$

where we assume $p/q \in \mathbb{Q}$ is in simplest terms. The function f is continuous for the irrational numbers in $[0, 1]$ but discontinues for the rational numbers in $[0, 1]$.

Suppose $r \in [0, 1]$ is irrational, giving $f(r) = 0$. Fix $\varepsilon > 0$. By the Archimedean Property, there is some $m \in \mathbb{N}$ such that $1/m\varepsilon$, and some $k_m \in$ such that

$$r \in \left(\frac{k_m}{m}, \frac{k+1}{m} \right).$$

Define the following values:

$$d_m = \min \left\{ \left| r - \frac{k}{m} \right|, \left| r - \frac{k+1}{m} \right| \right\},$$
$$\delta = \min\{d_1, \ldots, d_m\}.$$

If $x \in [0, 1]$ is a rational number with $|x - r| < \delta$, then $x = p/q$ and $q > m$, so

$$|f(x) - f(r)| = \left| \frac{1}{q} - 0 \right| = \frac{1}{q} < \frac{1}{m} < \varepsilon,$$

whenever $|x - r| < \delta$. If $x \in [0, 1]$ is irrational, then

$$|f(x) - f(r)| = |0 - 0| = 0 < \varepsilon,$$

regardless of δ. Thus, for any $x \in [0, 1]$ satisfying $|x - r| < \delta$, $|f(x) - f(r)| < \varepsilon$. The function is continuous at any irrational r.

Now suppose $r = p/q \in [0, 1]$ is rational. We have $f(r) = 1/q$. Let $\varepsilon = 1/2q$ and define $r_k = r + \frac{1}{k\sqrt{2}}$. The number r_k is irrational, so $f(r_k) = 0$. Can we find a δ such that

$$|f(x) - f(r)| = \left| f(x) - \frac{1}{q} \right| < \varepsilon = \frac{1}{2q}?$$

No we cannot. We can always find and x_k such that

$$|f(x_k) - f(r)| = \frac{1}{q} > \frac{1}{2q}.$$

Remark 4.2 (Pathological Examples)**.** Thomae's function is constructed for the express purpose of being an unintuitive example. Such examples are known as pathological in math, and are particularly common in analysis. While such examples would never arise in the real world, they often give insight into just what is possible. We'll see several more: a function that is continuous everywhere but differentiable nowhere, a set that we can not possible know the size of, a set that is uncountably infinite but has no size, etc.

Remark 4.3 (Can δ depend on p?)**.** Yes! In Example 4.5, our selection of δ was independent of x, and works for every point in $\mathbb{R}$. This was not the case for Example 4.7, but that is fine. Our selection of δ will change for each value of $p \in \mathbb{R}$. All that matters is *for each* $\varepsilon > 0$, we can find a correspond δ. We'll come back to this idea shortly.

Remark 4.4 (It Only Takes One ε, Discontinuity)**.** Showing a function is *not* continuous only requires we find a single ε such that $|f(x) - f(y)| \geq \varepsilon$ for all $|x - y| < \delta$ for $y \in E$. This follows from the negation of Definition 4.1. If we let $x \to 0$, then we end up with $\delta = 0$, a clear contradiction.

Our next makes explicit the fact that if f is continuous at a point, then the limit of the function exists at that point.

Theorem 4.3. Let X and Y be metric spaces, $E \subset X$, $p \in E$ be a limit point of E, and $f : E \to Y$. Then f is continuous *if and only if* $\lim_{x \to p} f(x) = f(p)$.

Proof. Simply compare the definition of a limit (Definition 4.1), with the definition of continuity (Definition 4.2). □

Corollary 4.2 (Continuous Functions Preserve Limits)**.** Let X and Y be metric spaces, and $E \subset X$. Suppose there is some function $f : E \to Y$, and a $p \in E$ which is a limit point of E. Then f is continuous at p *if and only if* $f(p_n) \to f(p)$ for every non-constant sequence $\{p_n\}$ in E such that $p_n \to p$. That is

$$\lim_{n\to\infty} f(p_n) = f\left(\lim_{n\to\infty} p_n\right) = f(p).$$

Proof. This follows from Theorem 4.1 and Theorem 4.3. □

Corollary 4.3 (Properties of Continuous Functions)**.** Let X and Y be metric spaces, and $E \subset X$. Suppose there are some functions $f : E \to Y$ and $g : E \to Y$, and a point $p \in E$. If f and g are continuous at p, then

1. $f + g$ is continuous at p;
2. fg is continuous at p;
3. f/g is continuous at p if $p \neq 0$.

Proof. This follows from Theorem 4.2 and Theorem 4.3. □

Remark 4.5 (Moving Around Limits)**.** Corollary 4.2 is the first time we've encountered a type of question analysis will provide answers for again and again – when can I move a limit into a function/operation? The majority of Section 7 will be dedicated to these type of questions. Corollary 4.2 tells us that a function is continuous if and only if the limit of a function is the function of the limits.

Example 4.9. Define $f : \mathbb{R} \to \mathbb{R}$ as

$$\begin{cases} x \text{ if } x \in \mathbb{Q} \\ 0 \text{ if } x \notin \mathbb{Q} \end{cases}.$$

This function is continuous at 0, because for all real $x_n \to 0$, $f(x_n) \to f(0) = 0$. This happens to be the only point in $\mathbb{R}$ at which f is continuous. Let $c \in \mathbb{Q}\backslash\{0\}$. There is a sequence of irrational numbers $\{x_n\}$ such that $x_n \to c$. Despite this, $f(x_n) = 0 \to 0 \neq c$, so f is not continuous at c. Similarly, if $c \in \mathbb{R}\backslash\mathbb{Q}$, then there exists a sequence of rationals x_n such that $x_n \to c$. Despite this, $f(x_n) = x_n \to c \neq f(c) = 0$, so f is not continuous at c.

While Corollary 4.3 proves valuable when showing a function is continuous, when combined with our next result, we can show that almost any function we can write down is continuous.

Theorem 4.4 (Composition Preserves Continuity)**.** Let X, Y, and Z be metric spaces, and $E \subset X$. If $f : E \to Y$ and $g : f(E) \to Y$ are continuous at $p \in E$ and $f(p) \in f(E)$ respectively, then the function $h : g(f(E)) \to Z$ defined as

$$h(x) = g(f(x)) = (g \circ f)(x)$$

is continuous at p.

Proof. Fix $\varepsilon > 0$. Since g is continuous at $f(p)$, there exists a δ' such that

$$d_Z(g(y), g(f(p))) < \varepsilon$$

if $d_Y(y, f(p)) < \delta'$ for $y \in f(E)$. Since f is continuous at p, there exists $\delta > 0$ such that

$$d_Y(f(x), f(p)) < \delta'$$

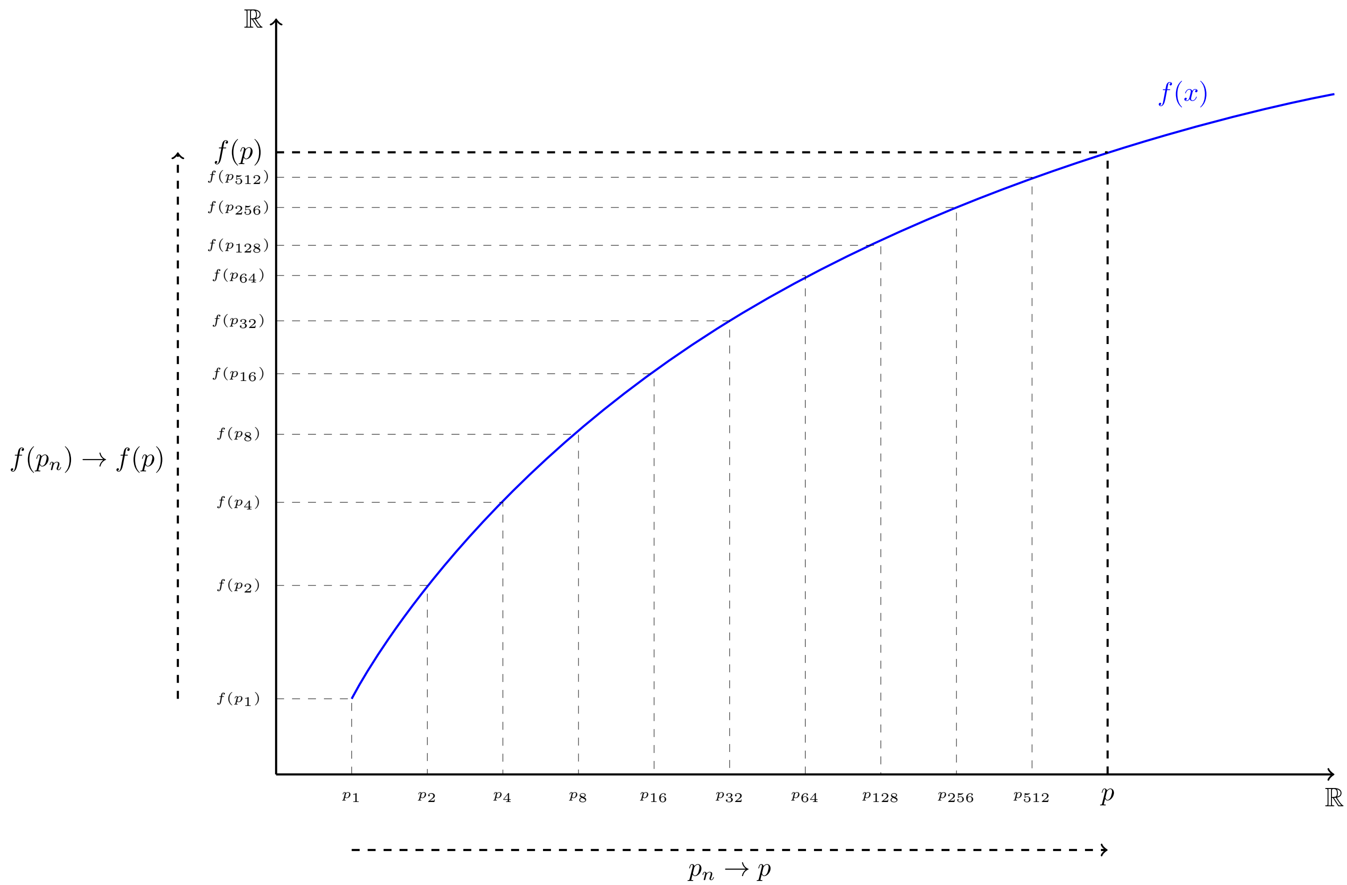

Figure 39: For $p_n \to p$, we have $f(p_n) \to f(p)$, therefore f is continuous at p.

if $d_X(x,p) < \delta$ and $x \in E$.[52] These inequalities give

$$d_Z(g(f(x)), g(f(p))) = d_Z(h(x), h(p)) < \varepsilon$$

if $d_X(x,p) < \delta$ and $x \in E$. Therefore $h = g \circ f$ is continuous at p. $\square$

If $h = g \circ f$ is continuous, then any composition of h with another continuous function will also be continuous. If we keep composing the result with continuous functions, we will always have a continuous function. For this reason, nearly every function encountered in actual applications will be continuous.

Example 4.10. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = \left(\frac{e^{\sin x}}{\cos x}\right)^4.$$

If we let $g(x) = \cos x$, $h(x) = \sin x$, $j(x) = e^x$, $k(x) = x^4$ (all continuous functions from $\mathbb{R}$ to $\mathbb{R}$), then

$$f(x) = k \circ ((j \circ h)/g).$$

By Theorem 4.4 and Corollary 4.3, $f(x)$ is continuous.

If a continuous function is a "nice" mapping from some metric space $E \subset X$ to another metric space Y, it's worth asking what happens to sets when they mapped using f? If E is open, will the image $f(E)$ be

[52] We took the δ' from the definition of g's continuity, and let $\varepsilon = \delta'$ for the definition of f's continuity. This means that any $f(x)$ which satisfy $d_Y(f(x), f(p)) < \delta'$ will satisfy $d_Y(y, f(p)) < \delta'$.

open? Does this hold for closed sets? What about our favorite sets, compact sets? We will now begin to put continuity in conversation with Section 2, and the point-set topology of metric spaces. We'll end this subsection by introducing a *very important* theorem related to the topology of a metric space. We will then introduce a special type of continuity, before seeing how continuity interacts with compactness.

Theorem 4.5. Let X and Y be metric spaces. The function $f : X \to Y$ is continuous on X *if and only if* $f^{-1}(V)$ is open in X for every open set V in Y.

Proof.

($\Longrightarrow$) Suppose f is continuous on X and V is an open set in Y. We need to show that $f^{-1}(V)$ is open by showing every point of the set is an interior point. Let $p \in f^{-1}(V) \subset X$, and $f(p) \in V$. The set V is open, so there exists an $r = \varepsilon > 0$ such that $N_\varepsilon(f(p)) \subset V$. Alternatively we may write, $y \in V$ for all $d_Y(f(p), y) < \varepsilon$. The function f is continuous at p because $p \in X$, so there exists a $\delta > 0$ such that $d_Y(f(x), f(p)) < \varepsilon$ is $d_X(x, p) < \delta$. In terms of sets, this means that $N_\delta(p) \subset f^{-1}(V)$, so $p \in f^{-1}(V)$ is an interior point. Therefore $f^{-1}(V)$ is open.

($\Longleftarrow$) Suppose $f^{-1}(V)$ is open in X for all open sets V in Y. For $p \subset X$ and $\varepsilon > 0$, let $V = N_\varepsilon(f(p))$. Alternatively, V is the set of all y such that $d_Y(y, f(p)) < \varepsilon$. The set V is a neighborhood, so it is open, hence $f^{-1}(V)$ is open. If $f^{-1}(V)$ is open, then there exists a δ such that $N_\delta(p) \subset f^{-1}(V)$. In other words, as soon as $d_X(p, x) < \delta$, $x \in f^{-1}(V)$. Because $x \in f^{-1}(V)$, $f(x) \in f(V)$, so we have $d_Y(f(x), f(p)) < \varepsilon$ as soon as $d_X(p, x) < \delta$. This means f is continuous on X. =

□

Corollary 4.4. Let X and Y be metric spaces. The function $f : X \to Y$ is continuous on X *if and only if* $f^{-1}(V)$ is closed in X for every closed set V in Y.

Proof. The complement of an open set is closed, and $f^{-1}(V^c) = (f^{-1}(V))^c$. □

Theorem 4.5 is illustrated in Figure 40 for a real function.

Remark 4.6 (Topological Definition of Continuity)**.** We've already referenced several times the fact that we are restricting our attention to metric spaces, but that point-set topology can be studied in a more general setting. As it turns out, when working in a topological space that is not a metric space, we define a continuous function as one with the property established in Theorem 4.5. If you were to ever take a proper topology course, continuity is never discussed in terms of $\varepsilon - \delta$, but is instead related to open sets.

Example 4.11. Theorem 4.5 only says that the preimage of an open set will be open when a function is continuous. It never says that the image of and open set is open. If we define $f : \mathbb{R} \to \mathbb{R}$ as $f(x)^2$, and let $E = (-1, 1)$ be an open set, then

$$f(E) = [0, 1).$$

Therefore $f(E)$ is not open, despite f being continuous on $\mathbb{R}$.

Example 4.12. Theorem 4.5 greatly simplifies the proof of Theorem 4.4. For an open $E \subset Z$, $g^{-1}(E) \subset Y$ is open by the continuity of g. The set $f^{-1}(g^{-1}(E)) \subset X$ is open because f is continuous. Therefore, for $h = g(f(x))$, $h^{-1}(E) \subset X$ is open, making h continuous.

4.3 Uniform Continuity

Before exploring compactness and continuity, we need to address an earlier remark, namely Remark 4.3. We noticed that the value of δ which corresponds to ε sometimes depends on where we are in the domain of a function. While does not violate continuity, it is a bit odd. If f is continuous on a set E, then we can always find a corresponding δ for each ε, but that δ may not work for all values of E. What we are really doing is find a corresponding δ for each ε *and* for each $p \in E$. However, Example 4.5 showed us that sometimes our corresponding δ works *for all* values on the domain. This is a special case that merits its own definition.

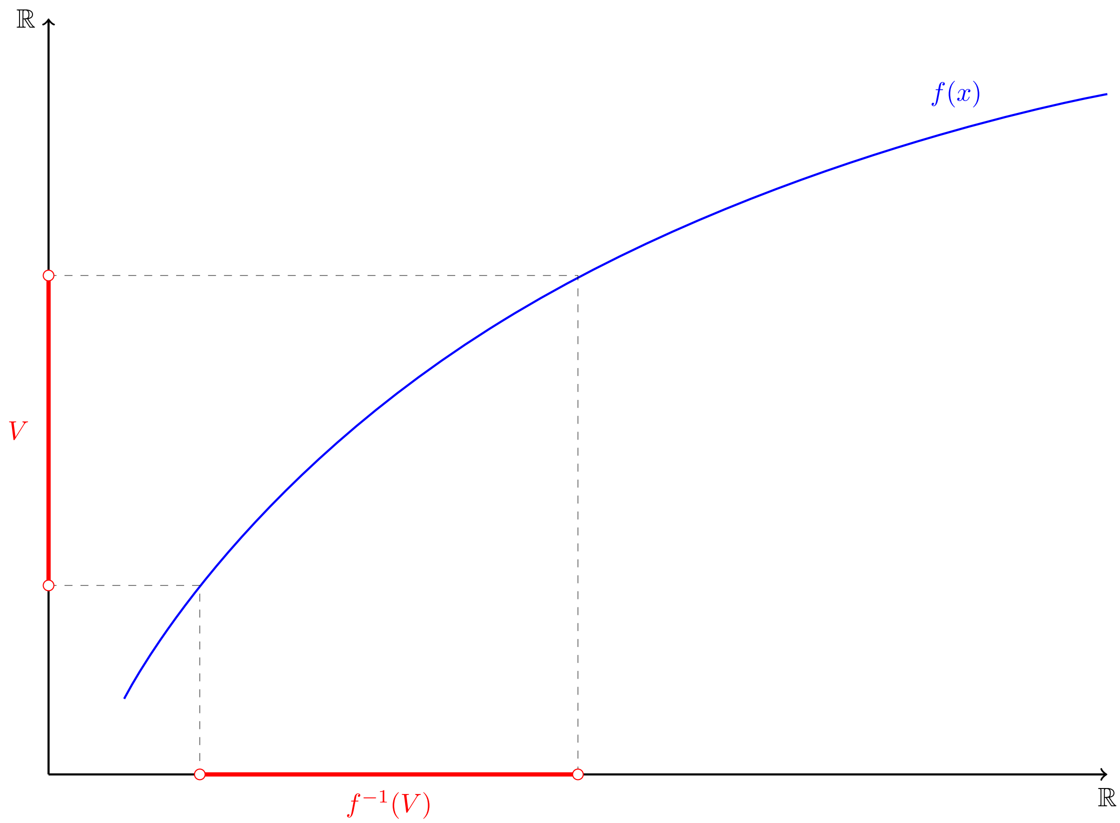

Figure 40: A continuous function $f : \mathbb{R} \to \mathbb{R}$, and the open interval $V \subset \mathbb{R}$. Because f is continuous, $f^{-1}(V)$ is open in $\mathbb{R}$.

Definition 4.3. Let X and Y be metric spaces. The function $f : X \to Y$ is *uniformly continuous (on X)* if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $d_X(x, p) < \delta$ implies $d_Y(f(x), f(p)) < \varepsilon$ *for all* $x, p \in X$.

There are two major differences between uniform continuity and continuity. We already discussed the first − we find a single value δ which works for all elements of f's domain. This difference arises from the fact that in Definition 4.2, we only looked at $p \in E$ which satisfy $d_X(x, p) < \delta$. We don't do this in definition 4.3, as we look at *all* $p \in E$. The second is that uniform continuity pertains to sets, not points. While a function can be continuous at a single point, saying f is uniformly continuous at a point it vacuously true and meaningless, as the concept deals with multiple points in a domain. It also should be clear that any function which is uniformly continuous is continuous.

It cannot be emphasized enough how much stronger uniform continuity is than continuity. If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then for every $\varepsilon > 0$ (of which there are an uncountably infinite amount), we may need to find a corresponding δ for each point in $\mathbb{R}$! It is feasible that we have an uncountably infinite amount of δ for each ε. If instead f is uniformly continuous, for each fixed $\varepsilon > 0$ (of which there are still an uncountably infinite amount), then we know there is one "silver bullet" δ that works for all of $\mathbb{R}$. This difference gives rise to an interesting geometric interpretation presented in Figure 41 and Figure 42.

Example 4.13. Let $f : [0, \infty) \to \mathbb{R}$ be defined as $f(x) = \sqrt{x}$. The function is uniformly continuous on $[0, \infty)$. If we let $\delta = \varepsilon^2$, then for all $|x - p| < \delta$, we have

$$|f(x) - f(p)|^2 = |\sqrt{x} - \sqrt{p}|^2 \leq |\sqrt{x} - \sqrt{p}| \cdot |\sqrt{x} + \sqrt{p}| = |x - p| < \varepsilon^2.$$

Taking the square root of this inequality gives $|f(x) - f(p)| < \varepsilon$ for all $|x - p| < \delta$. Our choice of δ does not depend on $p \in \mathbb{R}$, so f is uniformly continuous.

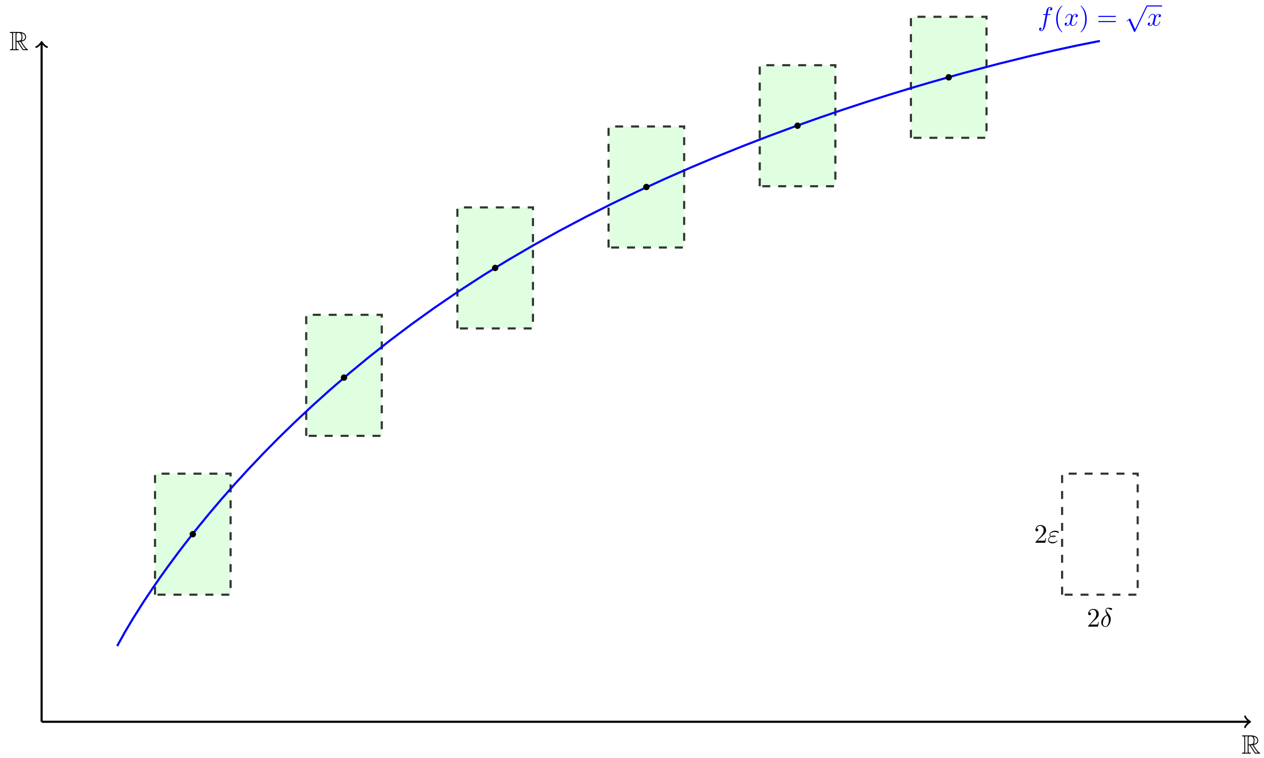

Figure 41: Let $f : (0, 1] \to \mathbb{R}$ be defined as $f(x) = \sqrt{x}$. The function is uniformly continuous. For each fixed ε, we can center a box of height ε at a point $(x, f(x))$. We can find some width of the box δ such that the function will never "escape" from the top or bottom of the box, no matter the point $(x, f(x))$.

Example 4.14. The function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = 1/x$ is not uniformly continuous on $(0, 1]$. Suppose for contradiction that f is uniformly continuous on $(0, 1)$. Let $\varepsilon = 1/n$. There exists a δ such that

$$|f(x) - f(p)| = \left|\frac{1}{x} - \frac{1}{p}\right| = \left|\frac{x-p}{xp}\right| < \frac{\delta}{xp} < \varepsilon = \frac{1}{n}$$

for all $x, p \in (0, 1)$ satisfying $|x - p| < \delta$. Therefore we need $\delta < |xp|/n$ for all $x, y \in (0, 1)$ satisfying $|x - p| < \delta$. If we let $y = x + \delta/2$, then $\delta < |xp|/n$ becomes

$$\delta < \frac{2x^2}{2n - x}$$

for all $x \in (0, 1)$.

4.4 Continuity and Compactness

If continuous functions are "nice", and compact sets are "nice", then it should not come as a surprise that continuous functions behave very well with compact sets. We will first define what it means for a function to be bounded, and then we will jump into a series of results about compactness.

Definition 4.4. Let $E \subset \mathbb{R}$. The real function $f : E \to \mathbb{R}$ is *bounded* if there exists an $M \in f(x)$ such that $|f(x)| \leq M$ for all $x \in E$.

Figure 43 shows a bounded real function.

Example 4.15. The function $\sin(x)$ is bounded on all of $\mathbb{R}$, as $|\sin x| \leq 1$ for all $x \in \mathbb{R}$.

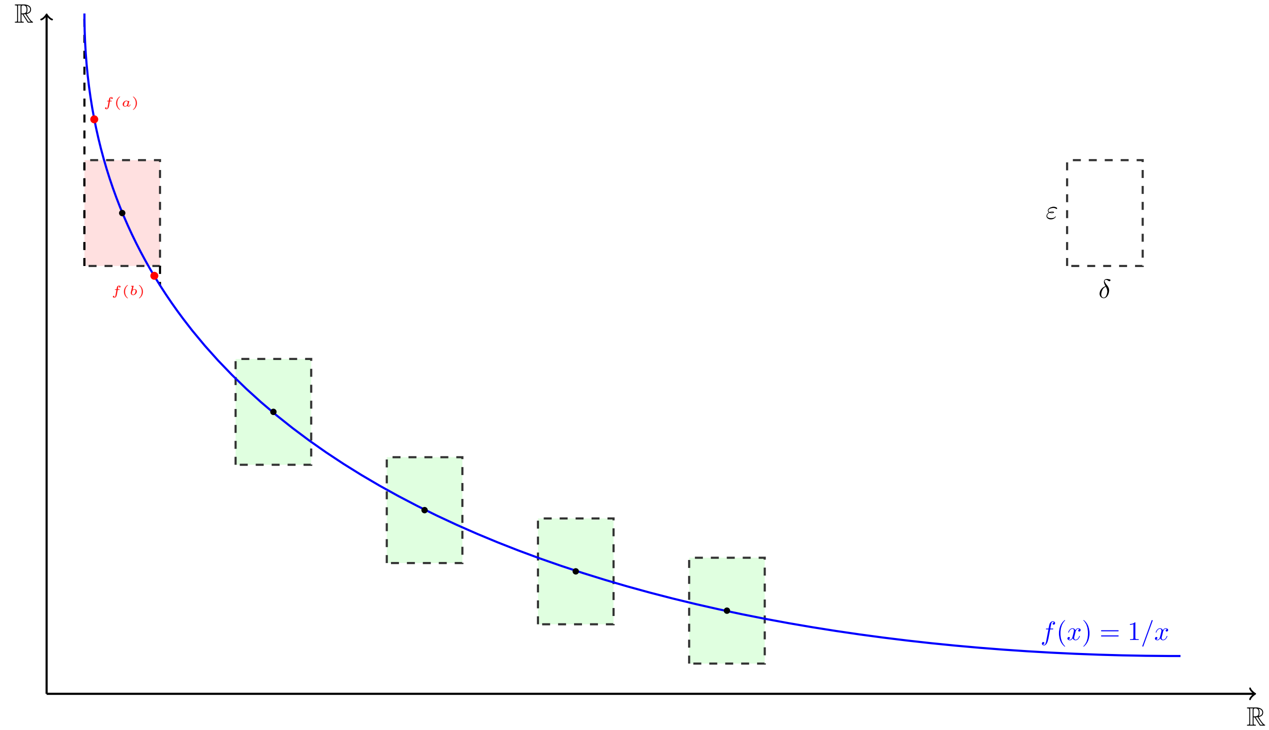

Figure 42: The function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = 1/x$ is not uniformly continuous. For our set value of ε and δ the function "escapes" from the top and bottom of the red rectangle. This is because there exists $a, b \in \mathbb{R}$ such that $|f(a) - f(b)| \geq \varepsilon$ despite the fact that $|a - b| < \delta$. We could make the corresponding δ smaller, but the function eventually "escape" as we move to the left.

Example 4.16. The function $f(x) = x^2$ on $\mathbb{R}$ is continuous but not bounded.

Theorem 4.6 (Continuity Preserves Compactness)**.** Let $f : X \to Y$ be a continuous function where X and Y are metric spaces. If X is compact, then $f(X)$ is compact.

Proof. We will show that an arbitrary open cover $\{V_\alpha\}$ of $f(X)$ has a finite subcover. Since f is continuous, $f^{-1}(V_\alpha)$ is open for all α (Theorem 4.5). The set X is compact, so there exists finitely many indices $\alpha_1, \ldots, \alpha_n$ such that

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

But we have $f(f^{-1})(E) \subset E$ for any $E \subset Y$, so we can take the image of the finite subcover of X and conclude

$$f(X) \subset V_{\alpha_1} \cup \ldots \cup V_{\alpha_n}.$$

Therefore $f(X)$ is compact. □

This result is similar to Theorem 4.5, but they "go in different directions". The preimage of an open set is open for a continuous f. Now we are saying that the image of a compact set is compact for a continuous f. It's tempting to say this holds for open sets, but it is not true. In general openness is not preserved by continuous functions (Example 4.11).

Corollary 4.5. Let $E \subset \mathbb{R}$. If the real function $f : E \to \mathbb{R}$ is continuous and E is closed and bounded, then $f(E)$ is closed and bounded.

Example 4.17. Let $f : (0, \infty) \to \mathbb{R}$ be the continuous function $f(x) = 1/x$. The set $(1, \infty)$ is neither closed nor bounded, so $f^{-1}((1, \infty)) = (0, 1)$ is neither closed nor bounded.

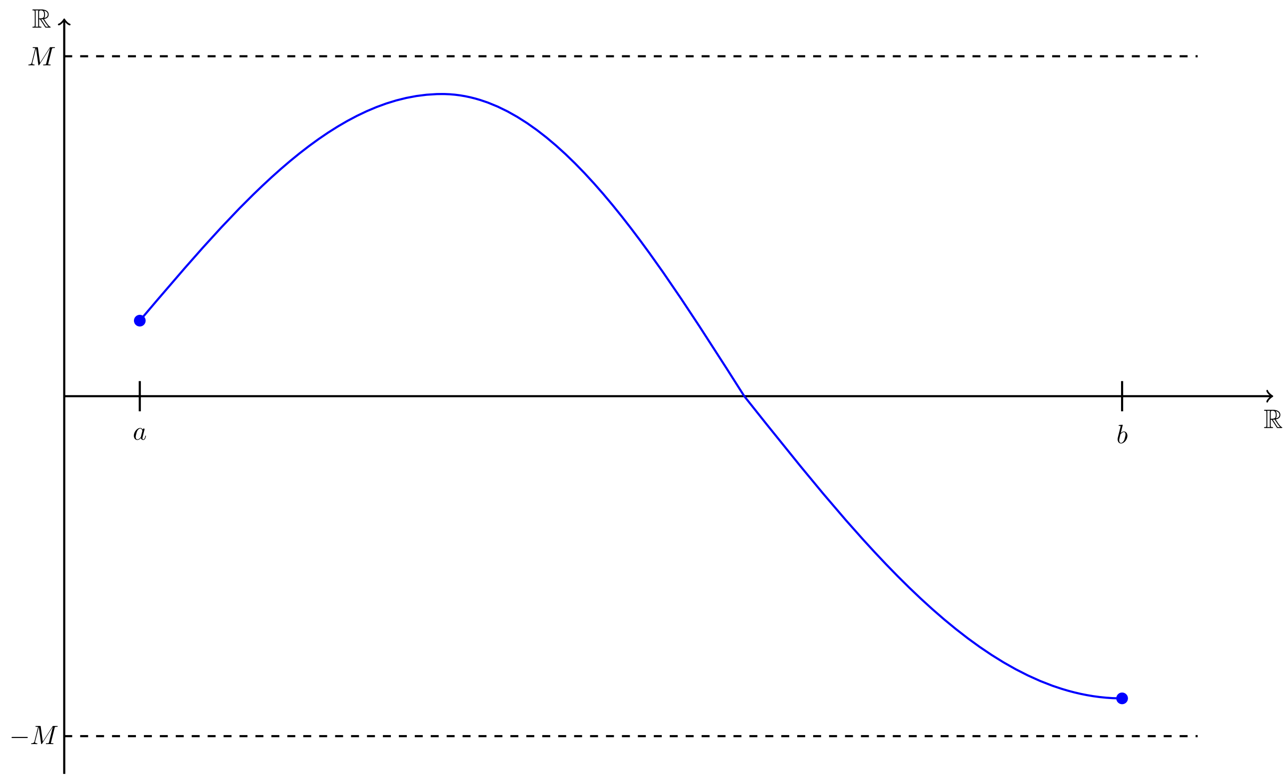

Figure 43: A bounded function $f : [a, b] \to \mathbb{R}$

Our next theorem proves useful in many applications, as it allows us to determine when a function has a maximum and minimum on an intervals.

Theorem 4.7 (Extreme Value Theorem)**.** Suppose f is a continuous real function on a compact metric space X, and

$$M = \sup_{x \in X} f(x),$$
$$m = \inf_{x \in X} f(x).$$

Then there exists points $x, y \in X$ such that $f(x) = M$ and $f(y) = m$.

Proof. Continuous functions preserve compactness (Theorem 4.6), so $f(X) \subset \mathbb{R}$ is closed and bounded. A closed and bounded set contain their infimum and supremum, so $f(X)$ contains M and m. □

Remark 4.7 (sup or max?)**.** It may not be clear when we use maximum and when we use supremum. A maximum of some set or function is always attained. This is not always the case for a supremum. If we know that $\sup E \in E$, then we are free to write $\max E = \sup E$, but this won't always hold. For instance, $(0, 1)$ has a supremum of 1, but no well defined maximum. If you're ever unsure if the supremum is attained in the set or by the function, use sup. A maximum is always a supremum, so it technically is not incorrect. The same holds for inf and min.

Theorem 4.7 tells us that if a function is defined on a compact space, then it *must* achieve a maximum and minimum on that domain. Just knowing that such points exist is a great deal of information.

Example 4.18. The function $\sin(x)$ does not achieve a maximum or minimum on the interval $(0, \pi/2)$, despite it being bounded. We can always find some larger value of $f(x)$ as we get arbitrarily close to $\pi/2$, or some smaller value of $f(x)$ as we get arbitraily close to 0. This is the exact type of behavior we ruled out when defining compactness! By defining $\sin(x)$ on $[0, \pi/2]$, we now achieve a maximum and minimum on the interval.

Example 4.19. The function $f(x) = x^2$ achieves a minimum on $\mathbb{R}$, but it fails to achieve both a maximum and a minimum on all of $\mathbb{R}$. This is because $\mathbb{R}$ is neither closed nor bounded, rendering it not compact.

Recall the fact that $f(x) = 1/x$ is not uniformly continuous on $(0,1)$. In Figure 42 we argued that no matter the value of δ given for a fixed ε, we could move the rectangle of height ε and length δ to the left until the function "escaped" from the top and bottom. Us being able to keep moving the rectangle to the left is a result of $(0,1)$ not being closed. We can always get a little bit closer to 0 without leaving the domain. As we do this, the function will take on larger values indefinitely because $f((0,1)) = (1,\infty)$ is unbounded. These two observations seem to hint at the fact that if eliminate these behaviors, a function will always be uniformly continuous. As it turns out, compactness does just this, and the proof of this is one of the more elegant proofs in analysis.[53]

Theorem 4.8. Let f be a continuous function which maps a compact metric space X to a metric space Y. Then f is uniformly continuous on X.

Proof. Fix $\varepsilon > 0$. The function f is continuous on X, so for each point $p \in X$ there is a $\delta(p)$ such that

$$d_Y(f(p), f(x)) < \varepsilon/2$$

whenever $x \in X$ and $d_X(p,x) < \frac{1}{2}\delta(p)$.[54] Let

$$J(p) = \left\{ x \in X \mid d_X(x,p) < \frac{1}{2}\delta(p) \right\}.$$

We have $p \in J(p)$ for all $p \in X$, so

$$X \subset \bigcup_{p\in X} J(p),$$

where $J(p)$ is open for all $p \in X$.[55] That is to say, $\{J(p)\}_{p\in X}$ is an open cover of X. The space X is compact, so there exists a finite set of points $p_1, \ldots, p_n$ such that

$$X \subset J(p_1) \cup \cdots \cup J(p_n).$$

If we let

$$\delta = \frac{1}{2}\min\{\delta(p_1), \ldots, \delta(p_n)\},$$

then $\delta > 0$.[56]

Let $x, p \in X$ such that $d_X(x,p) < \delta$. The point p is in X, so it must be "covered" by one of the open sets in the finite subcover $\{J(p_1), \ldots, J(p_n)\}$. That is, there is an $m \in \mathbb{N}$, where $1 \le m \le n$, such that $p \in J(p_m)$. Hence, $d_X(x,p_m) < \frac{1}{2}\delta(p_m)$. The triangle inequality gives

$$d_X(x,p_m) \le d_X(x,p) + d_X(p,p_m) < \delta + \frac{1}{2}\delta(p_m) \le \delta(p_m).$$

But then, we can use the triangle inequality and the continuity of f to conclude

$$d_Y(f(x), f(p)) \le d_Y(f(x), f(p_m)) + d_Y(f(p_m), f(p)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

so f is uniformly continuous. ∎

[53] Full disclosure: this is my favorite proof from basic real analysis.

[54] We are writing δ as a function of $p \in X$ to emphasize that all we know now if f is continuous.

[55] This follows from the fact that $J(p)$ is a neighborhood of p, and all neighborhoods are open sets. We could also write $J(p) = N_{1/2\delta(p)}(p)$.

[56] This finite set achieves is infimum so we use min. A minimum of a finite set of positive numbers in positive. This is not necessarily the case with an infimum of an infinite set of positive numbers even if that set is compact. For us to be able to conclude $\delta > 0$, it is crucial that the set $\{\delta(p_1), \ldots, \delta(p_n)\}$ is finite, that way we can take the minimum over a finite set corresponding to the finite subcover.

This proof may look technical, but the underlying idea is much simpler. Think about the function $f(x) = 1/x$ on $(0,1]$, and Figure 42. For a fixed ε, we can always move the rectangles to the left until the function "escapes" from the top or bottom. The closer we get to 0 on the $x-$axis, the smaller we need to make δ, and we can always move closer and closer to 0, so there isn't one δ that will work at every point. Suppose instead we looked at $f(x) = 1/x$ on the interval $[0.001, 1]$. We may be able to get closer and closer to 0.001, making δ smaller as we go to satisfy continuity, but because 0.001 is defined by f, we could just take the δ that works at 0.001 and use it for all of $[0.001, 1]$. If the δ needs to become small as we move our rectangle to the left, then the "leftmost" point must have the smallest δ, a choice of δ that will work for all of $[0.001, 1]$ and our fixed ε. The next example will walk through this process, and the proof of Theorem 4.8, with an actual function.

Example 4.20. Suppose $f : [-10, 10] \to \mathbb{R}$ is defined as $f(x) = x^2$. Example 4.7 showed that f is continuous. In this case we had $\delta = \min\{1, \varepsilon/(2|p| + 1)\}$. Fix $\varepsilon = 1$, so $\varepsilon/2 = 1/2$.[57] We can let $\delta(p) = 1/(4|p| + 1)$, and have

$$|f(x) - f(p)| < \varepsilon = 1/2$$

for all $p \in [-10, 10]$ satisfying $|x - p| < \frac{1}{2}\delta(p) < \delta(p)$. First, we should explicitly show that $\delta(p)$ will change with p. Suppose $p = 1$, giving $\delta(1) = 1/5$. For all $x \in [-10, 10]$ such that $|x - 1| < 1/5$, we do indeed have $|x^2 - 1| < 1/2$. Now let's try to use $\delta(1)$ for $p = 9$. The set of all $x \in [-10, 10]$ which satisfy $|x - 9| < 1/5$ is $(44/5, 46/5)$. For $91/10 \in (44/5, 46/5)$, we have

$$|f(91/10) - f(9)| = \left(\frac{91}{10}\right)^2 - 9^2 = 1.81 > \frac{1}{2} = \frac{\varepsilon}{2}.$$

This means that $\delta(1)$ won't work for $p = 9$! We now will mimic the proof of Theorem 4.8 and find a δ that will work for all $p \in [-10, 10]$.

Define

$$J(p) = \left\{ x \in [-10, 10] \;\middle|\; |x - p| < \frac{1}{2}\delta(p) = \frac{1}{8|p| + 2} \right\} = N_{1/2\delta(p)} = N_{1/(8|p|+2)}(p).$$

We have $p \in J(p)$ for all $p \in [-10, 10]$, so

$$[-10, 10] \subset \bigcup_{p \in [-10,10]} J(p).$$

For the finite set of points $P = \{-10, -9.9, \ldots, 9.9, 10\} = \{-10 + (0.1)n \mid n = 1, \ldots, 200\}$,[58] we have

$$[-10, 10] \subset \bigcup_{p \in P} J(p),$$

making $\{J(p)\}_{p \in P}$ a finite subscover of $[-10, 10]$. Define

$$\delta = \min_{p \in P}\{\delta(p)\} = \delta(10) = \delta(-10) = \frac{1}{81}.$$

We will show that this value of δ will work for a random point in $[-10, 10]$, say $p = 1.95$. We could verify this right away, but we'll instead follow the steps of the proof, even though they become painfully redundant when working with actual numbers. Now let $x \in [-10, 10]$ such that $|x - 1.95| < 1/81$.[59] We have $x \in [-10, 10]$, so it must be in one of the elements of the finite subcover $\{J(p)\}_{p \in P}$. We in fact have $1.95 \in J(1.9)$, hence

$$|1.95 - 1.9| < \frac{1}{2}\delta(1.95) = 0.0581,$$

[57] We use $\varepsilon = 1$ as the final value of ε used to show uniform continuity. The value $\varepsilon/2$ is used in the definition of continuity of x^2.

[58] Where the heck do I get this set? Well the minimum value of $1/(8|p| + 2)$ on the interval $[-10, 10]$ is about 0.012. If we round this down to 0.01, then all the $J(p)$ centered at points in $[-10, 10]$ that are distance 0.01 apart will be guaranteed to cover $[-10, 10]$, as we took the distance between points to be less than the smallest radius of $J(p)$.

[59] For the rest of this paragraph, whenever we refer to x, we mean these particular x satisfying $|x - 1.95| < 1/81$

so the continuity of f at $p = 1.95$ gives

$$|f(1.95) - f(1.9)| = 0.1925 < \frac{\varepsilon}{2} = \frac{1}{2}. \tag{5}$$

We also have

$$|x - 1.9| \leq |1.95 - x| + |1.95 - 1.9| < \delta + \frac{1}{2}\delta(1.95) = \frac{1}{81} + 0.0581 = 0.0704 < \delta(1.95) = .104,$$

so by the continuity of f at 1.9,

$$|f(x) - 1.9| < \frac{\varepsilon}{2} = \frac{1}{2}. \tag{6}$$

Combining Equation (5) and (6) gives

$$|f(x) - f(1.95)| \leq |f(x) - f(1.9)| + |f(1.9) - f(1.95)| < \frac{1}{2} + \frac{1}{2} = 1 = \varepsilon$$

for all $x \in [-10, 10]$ such that $|x - 1.95| < \delta = 1/81$. Our choice of δ does in fact work for $\varepsilon = 1$ and $p = 1.95$!

The choice of $p = 1.95$ does not matter. We could have picked *any* value in $[-10, 10]$, and $\delta = 1/81$ still would have worked for $\varepsilon = 1$. This should not come as a surprise, because we just picked *one of* the smallest value of $\delta(p)$ for $p \in [-10, 10]$. Why is it that we need compactness if we could have just let $\delta = \min_{p \in [-10,10]} \delta(p)$? This would work in this specific case because the range $\delta([-10, 10])$ is compact, due to $\delta(p)$ being continuous for $\varepsilon = 1$ (Theorem 4.6), and we the infimum of a compact set in $\mathbb{R}$ is an element of the set. In general this won't work, because we never explicitly said that $\delta(p)$ is a function, let alone a continuous function.

Example 4.21. *Any* real continuous function $f : [a, b] \to \mathbb{R}$ is uniformly continuous on $[a, b]$, because $[a, b] \subset \mathbb{R}$ is compact.

Example 4.22. A compact domain is a sufficient condition for uniform continuity, but it is not a neccesary condition. Any linear real function $y = mx + b$ is uniformly continuous on all of $\mathbb{R}$, and $\mathbb{R}$ is not compact.

4.5 Intermediate Value Theorem

We now will turn out attention to one of the major theorems presented in a calculus course. The Intermediate Value Theorem is perhaps the quintessential result of continuity. In calculus, you may have been taught that a continuous function is any function you could draw without picking up your pencil. If this is the case, and the range of your function starts at $f(a)$ and ends at $f(b)$, then your function will of course take on every value between $f(a)$ and $f(b)$.

Theorem 4.9. Let f be a real function continuous on the interval $[a, b]$. If $f(a) < f(b)$ and if c is a number satisfying $f(a) < c < f(b)$, then there is a $x \in (a, b)$ such that $f(x) = c$.

Proof. Suppose $f : [a, b] \to \mathbb{R}$ is continuous, and $f(a) < f(b)$. Let $c \in \mathbb{R}$ such that $f(a) < c < f(b)$. Define the set

$$S = \{x \in [a, b] \mid f(x) \leq c\}.$$

The set $S \subset [a, b]$ is nonempty[60] and bounded above by b, so $s = \sup S$ exists by the least-upper-bound property. We claim that $f(s) = c$. We will show that $f(s) \not< c$ and $f(s) \not> c$.

Suppose $f(s) > c$, and let $\varepsilon = f(s) - c > 0$. By the continuity of f, there exists a $\delta > 0$ such that

$$|f(x) - f(s)| < f(s) - c = \varepsilon$$

for all $x \in [a, b]$ which satisfy $|x - s| < \delta$. For all such x we can conclude $f(x) > c$,[61] so $x \notin S$. But if this holds for any $x \in [a, b]$ such that $|x - s| < \delta$, then $s - \delta$ is an upper bound of S. This contradicts $s = \sup S$.

[60] $a \in S$

[61] $|f(x) - f(s)| < f(s) - c$ implies $f(x) - f(s) < f(s) - c$ or $f(s) - f(x) < f(s) - c$. Either way, $f(x) > c$.

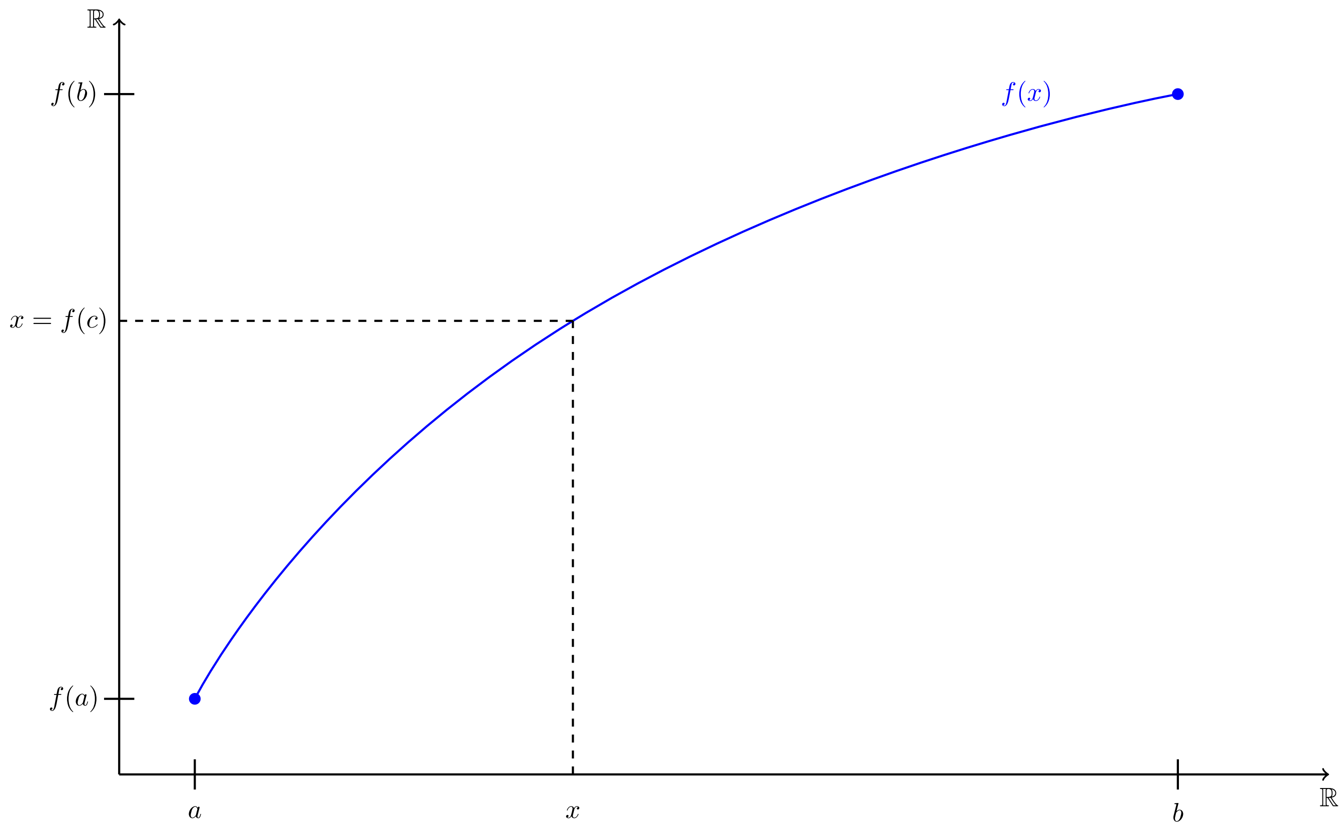

Figure 44: We have a real function $f : [a, b] \to \mathbb{R}$. The Intermediate Value Theorem says for all $f(a) < c < f(b)$, there is a $x \in (a, b)$ such that $f(x) = c$.

Suppose $f(s) > c$, and let $\varepsilon = c - f(s) > 0$. By the continuity of f, there exists a $\delta > 0$ such that

$$|f(x) - f(s)| < c - f(s) = \varepsilon$$

for all $x \in [a, b]$ which satisfy $|x - s| < \delta$. For all such x we can conclude $f(x) < y$, so $x \in S$. This implies that $s + \delta/2 \in S$ is an upper bound, which contradicts $s = \sup S$ being an upper bound. $\square$

The Intermediate Value Theorem may seem like a parlor trick, but it is useful in many proofs. There are many proofs that rely on us being able to find a certain value in an interval. If we have information about some function, and know the Intermediate Value Theorem holds, then we may be able to use if to find the value we're interested in. Many proofs are about "converting" information. We may have information about f, but need information about points in the domain of f. In a sense, Theorem 4.9 allows us to convert information about f *into* the information about the domain.

Remark 4.8 (Bounded Intervals, Figures)**.** Figure 44 is one of the first times where it's been important that a function is shown on a bounded interval $[a, b]$. For all future figures, if a function is on some bounded interval, or we are restricting our attention to a bounded interval, I will try to denote that with solid dots at the "beginning" and "end" of the function, or mark a and b on the $x-$axis. We saw this with Figure 43 as well. If no such markings are present, then I'm trying to convey that f is defined on all of $\mathbb{R}$.

Example 4.23 (Continuous Functions on $\mathbb{Q}$)**.** Let $f : [a, b] \to \mathbb{R}$ be continuous such that $f([a, b]) \subset \mathbb{Q}$.[62] The function f must be constant, that is for all $x \in [a, b]$, $f(x) = c$ for some $c \in \mathbb{R}$. Without loss of generality, assume $f(a) < f(b)$. By the Intermediate Value Theorem, the function takes on every value in the interval

[62] The function is still real, as $\mathbb{Q} \subset \mathbb{R}$.

$(f(a), f(b))$. This interval contains irrational numbers, as the irrational numbers are dense in $\mathbb{R}$. This is a contradiction, so f is constant.

Example 4.24 (Roots of a Polynomial)**.** Suppose

$$p(x) = a_0 + a_1 x + \ldots + a_k x^k$$

is an odd polynomial (k is an odd number) with real coefficients $a_i \in \mathbb{R}$. Taking the limit of $p(x)$ as x goes to infinity and negative infinity gives

$$\lim_{x \to -\infty} f(x) = -\infty,$$
$$\lim_{x \to \infty} f(x) = \infty.$$

This is the first time we've seen limits that go off to infinity, but if we combine what we know about sequences that diverge to infinity and Theorem 4.1, we can conclude that the limit does not exist, despite us using notation that would indicate it does. These attempts to take limits also lets us know that at some point in $\mathbb{R}$, $p(x)$ switches signs. Therefore by the Intermediate Value Theorem, there exists at least one root $x_0 \in \mathbb{R}$ such that $p(x_0) = 0$.[63]

Example 4.25. The converse of the intermediate value theorem is not true. That is, just because for any two point $x_1 < x_2$ and a number c in between $f(x_1)$ and $f(x_2)$ we are able to find a point $x \in (x_1, x_2)$ such that $f(x) = c$, that *does not mean* f is continuous. Define $f : \mathbb{R} \to \mathbb{R}$ to be

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 \text{ if } x = 0 \end{cases}.$$

This function satisfies the aforementioned property, but is not continuous on $\mathbb{R}$, as there is a discontinuity at 0.

4.6 Discontinuities

We've used the word discontinuous several times, but now we'll take the time to define it, and classify two types of discontinuities. A function that is discontinuous at a point is simply not continuous at the point. The formal definition we will present negates the statement given in Definition 4.2

Definition 4.5. Let X and Y be metric spaces, $E \subset X$, $p \in E$, and $f : E \to Y$. Then f is *discontinuous at p* if for all $\delta > 0$, there exists a single $\varepsilon > 0$ such that

$$d_Y(f(x), f(p)) \geq \varepsilon$$

for a single $x \in E$ which satisfies $d_X(x, p) < \delta$. If f is discontinuous at least one point in E, then f is *discontinuous (on E)*.

We can classify three different types of discontinuities if we introduce the notion of a right-hand limit and a left-hand limit for a real function.

Definition 4.6. Let f be a real function defined on (a, b), and consider any point p such that $a \leq p < b$. If for all sequence $\{p_n\}$ in (p, b) such that $p_n \to p$, we have $f(p_n) \to L$, then we write $f(p+) = L$, or

$$\lim_{x \to p^+} f(x) = p.$$

We say that L *is the right-hand limit* of $f(x)$ as $x \to p$.

[63] We do not know for sure how many roots of $p(x)$ are in $\mathbb{R}$, but we know it's at least one. If we wanted to look for the other roots, we would find exactly k of them in $\mathbb{C}$! This follows from the Fundamental Theorem of Algebra which says that any complex polynomial of degree k has k roots in $\mathbb{C}$. If these sorts of facts interest you, see Dummit and Foote (2004).

Definition 4.7. Let f be a real function defined on (a, b), and consider any point p such that $a < p \leq b$. If for all sequence $\{p_n\}$ in (a, p) such that $p_n \to p$, we have $f(p_n) \to L$, then we write $f(p-) = L$, or

$$\lim_{x \to p^-} f(x) = p.$$

We say that L *is the left-hand limit* of $f(x)$ as $x \to p$.

An alternate definition of these limits would use ε and δ. For example, $f(p-) = L$ if for all $\varepsilon > 0$, we have $|f(p) - L| < \varepsilon$ for all $x \in (a, b)$ satisfying $p - \delta < x < p$. The only difference between this and the definition of a limit is we restrict our attention to the x that are within a distance of δ to the left of p. The definition for the right hand limit would be the same, except we look at the interval $p < x < x + \delta$It can be proven that if $f(x+) = f(x-)$, then $\lim_{x \to p} f(x) = p$. The first type of discontinuity is the type we saw in Figure 37. This type it special, as it does not prohibit the existence of a limit at a point.

Definition 4.8. Let f be a real function defined on (a, b). If f is discontinuous at a point x, and if $f(x+) = f(x-)$, then we say f has a *removable discontinuity* at x.

With removable discontinuities, it's important to remember that f still needs to be defined at the point of discontinuity. It doesn't mean anything to say that f has a discontinuity at a point where it is not defined. Our second type of discontinuity is the type shown in Figure 45.

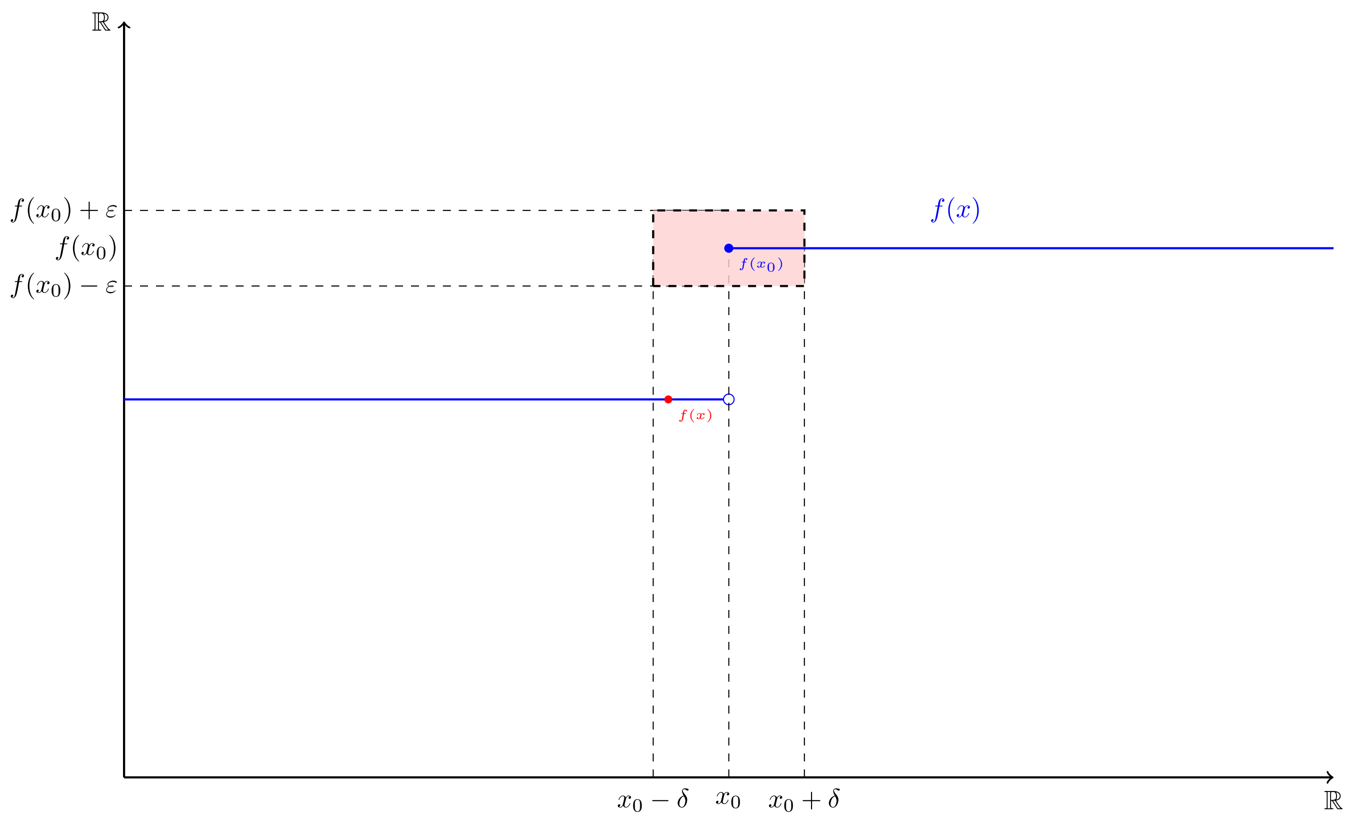

Figure 45: The real function $f : \mathbb{R} \to \mathbb{R}$ shown here is discontinuous at x_0. No matter small we take δ to be, we can find at least one $\varepsilon > 0$ and at least one x satisfying $|x - x_0| < \delta$ such that $|f(x) - f(x_0)| \geq \varepsilon$. This is also an example of a jump discontinuity.

Definition 4.9. Let f be a real function defined on (a, b). If f is discontinuous at a point x, and if $f(x+) \neq f(x-)$, then we say f has a *jump discontinuity* at x.

Lastly we treat the case shown in Figure 46.

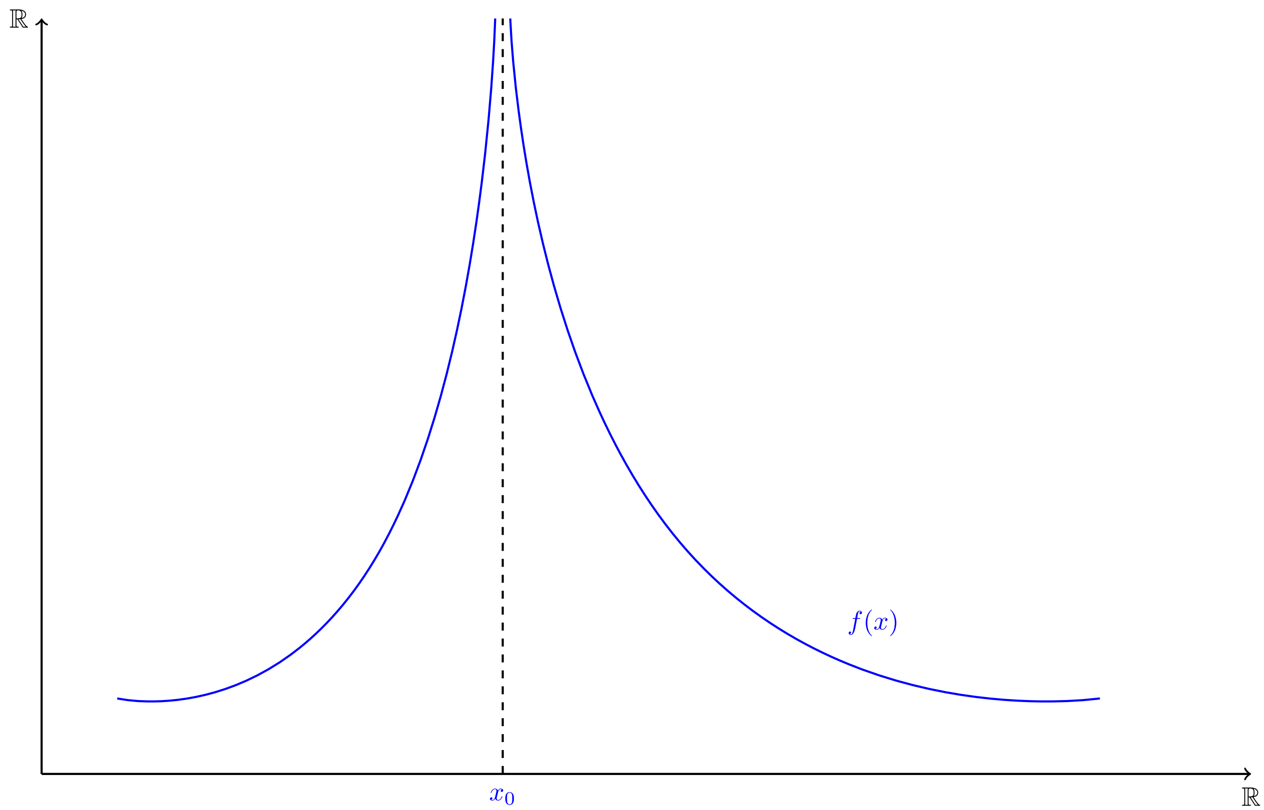

Figure 46: This real function has an essential discontinuity at the point x_0. In this particular instance, neither $f(x_0-)$, nor $f(x_0+)$ exist.

Definition 4.10. Let f be a real function defined on (a, b). If f is discontinuous at a point x, and if either $f(x-)$, or $f(x+)$ (or both) do not exist, then we say f has an *essential discontinuity* at x.

Example 4.26 (Dirichlet Function, Nowhere Continuous)**.** Define $f : \mathbb{R} \to \mathbb{R}$ as

$$f(x) = \begin{cases} 1 \text{ if } x \in \mathbb{Q} \\ 0 \text{ if } x \notin \mathbb{Q} \end{cases} .$$

This function takes the value 1 at every rational, and 0 at every irrational. This function happens to be discontinuous on all of $\mathbb{R}$! Let ε be any number in$(0, 1]$ such that $1/2$. If $x \in \mathbb{Q}$, for any value of $\delta > 0$, we can always find some irrational $y \notin \mathbb{Q}$ such that $|x - y| < \delta$, as the irrational numbers are dense in the rationals. This means we would have

$$|f(x) - f(y)| = |0 - 1| = 1 > \frac{1}{2} = \varepsilon,$$

for all δ! A similar argument holds if $x \notin \mathbb{Q}$, as the rational numbers are dense in the reals. Another way to see this is by letting x_n be a sequence of rationals which converge to a real number x.[64] We have $f(x_n) = 1$ for all x_n, so

$$f\left(\lim_{n\to\infty} x_n\right) = f(x) = 0 \neq 1 = \lim_{n\to\infty} 1 = \lim_{n\to\infty} f(x_n),$$

so f is not continuous at x by Corollary 4.3. We could show a similar result by using the density of the irrationals in the rationals to construct a sequence of irrationals which converge to an arbitrary rational. We also have that each discontinuity is an essential discontinuity.

[64] We know such a sequence exists by Corollary 3.1.

4.7 Monotonicity

Finally, we'll discuss a special group of functions that are either always weakly increasing or weakly decreasing. These functions will play a very important role in the study of integration, and will have an interesting property related to differentiation.

Definition 4.11. Let f be a real function defined on (a, b). The function f is *monotonically increasing* on (a, b) if $a < x < y < b$ implies $f(x) \leq f(y)$.

Definition 4.12. Let f be a real function defined on (a, b). The function f is *monotonically decreasing* on (a, b) if $a < x < y < b$ implies $f(x) \geq f(y)$.

Example 4.27. The following real functions are monotonically increasing on the entirety of their domains: $f(x) = e^x$, $f(x) = \ln x$, $f(x) = x$, $f(x) = \sqrt{x}$. The following functions are monotonically decreasing on the entirety of their domains: $f(x) = -x$, $f(x) = 1/x$, $f(x) = \arccos(x)$.[65]

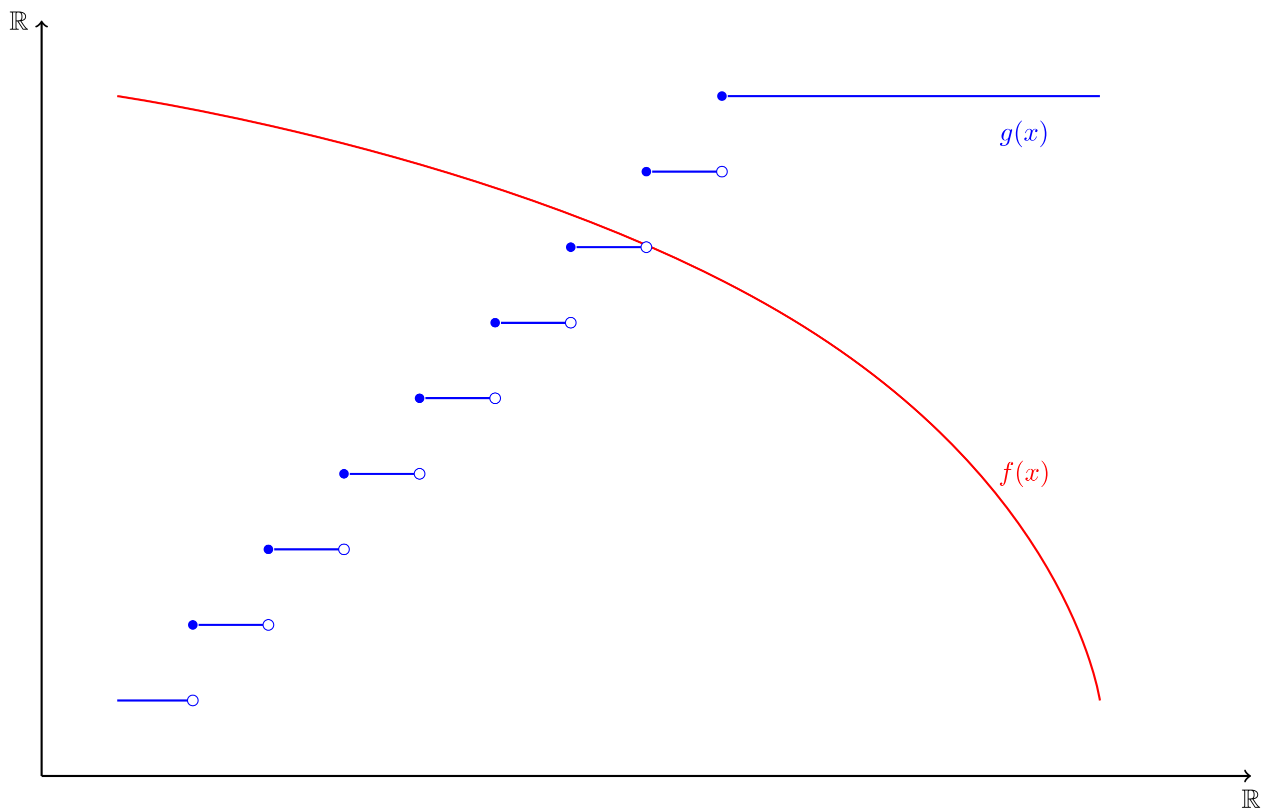

Figure 47: We have two real functions, $f : \mathbb{R} \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$. The function f is monotonically decreasing, whereas the function $f(x)$ is monotonically increasing. Note that g is not continuous on $\mathbb{R}$.

If a function is either monotonically increasing or monotonically decreasing, we may just refer to it as *monotonic*. Neither Definition 4.11 nor Definition 4.12 make any mention of continuity. As Figure 47 shows, monotonicity does not imply continuity. In fact, neither of the presented definitions seem to have anything whatsoever to do with continuity. Monotonicity alone is not a strong enough condition for us to make meaningful statements about continuity, *but* it does provide us with information about discontinuities. If a function is monotonic, we can deduce three facts about any discontinuities it may have.

[65] You should avoid writing $\cos^{-1}(x)$, as it is not clear if that either means $\arcsin(x)$ or $1/\cos(x)$.

Proposition 4.1. Let f be a monotonically increasing real function on (a, b). The limits $f(x+)$ and $f(x-)$ exist at every point of (a, b). Furthermore,

$$\sup_{a<t<x} f(t) = f(x-) \leq f(x) \leq f(x+) = \inf_{x<t<b} f(t).$$

We also will have $f(x+) \leq f(y-)$ for all $a < x < y < b$.

An analogous result holds for monotonically decreasing functions, and the proof is similar.

Proof. The range $f((a, x)) \subset \mathbb{R}$ has an upper bound in $f(x)$. By the completeness of $\mathbb{R}$, there exists some $A = \sup f((a, x))$. By the definition of the supremum, $A \leq f(x)$. We will show that $A = f(x-)$.

Fix $\varepsilon > 0$. By the definition of A as a least-upper-bound, there exists some $\delta > 0$ such that $a < x - \delta < x$ and

$$A - \varepsilon < f(x - \delta) \leq A.$$[66]

The function f is monotonic, so for all t satisfying $x - \delta < t < x$, we have

$$f(x - \delta) \leq f(t) \leq A.$$

These two inequalities imply that for all $t \in (a, b)$ satisfying $x - \delta < t < x$, we have

$$|f(t) - A| < \varepsilon,$$

so

$$A = \sup f((a, b)) = \sup_{a<t<x} f(t) = f(x-)$$

as desired. To show that $\sup_{x<t<b} f(t) = f(x+)$, we let $B = \inf f((x, b))$, and repeat this process. This gives the inequality presented in the proposition.

If we now let $a < x < y < b$, then we can apply the presented inequality to the point $x \in (a, y)$ and write

$$f(x+) = \inf_{x<t<y} f(t).$$

Applying it to $y \in (x, b)$ gives

$$f(y-) = \sup_{x<t<y} f(t).$$

This infimum and supremum are taken over the same set, so the infimum is less that the supremum, which gives the second desired inequality,

$$f(x+) = \inf_{x<t<y} f(t) \leq \sup_{x<t<y} f(t) = f(y-).$$

□

The first inequality of Proposition 4.1 is shown in Figure 48. Perhaps more useful than Proposition 4.1, is one of its immediate consequences. The sup and inf in question will always exist, implying the existence of both $f(x-)$ and $f(x+)$. In light of definition 4.10, we arrive at a corollary.

Corollary 4.6. A monotonic function in $\mathbb{R}$ has no essential discontinuous.

This result makes sense if we think about Figure 46. For a function to have an essential discontinuity like this, we need it to increase to infinity from the left or right. This means the function must decrease once we move past the discontinuity. This would contradict monotonicity because the function would increase for certain values of its domain, and then decrease for others.

Perhaps the most interesting result pertaining to discontinuities of monotonic functions, is that they will always form a countable set. We saw functions with an uncountable number of discontinuities (Example 4.26), but we'll see that monotonicity does not allow for this. This fact will become especially useful when we develop the theory of Riemann integration.

[66] We found a value $x - \delta$ such that $f(x - \delta)$ is in $f((a, x))$, but just barely below the set's supremum or equal to it. It's so close to that supremum, that $f(x - \delta) \in (A - \varepsilon, A]$ for our arbitrarily small ε.

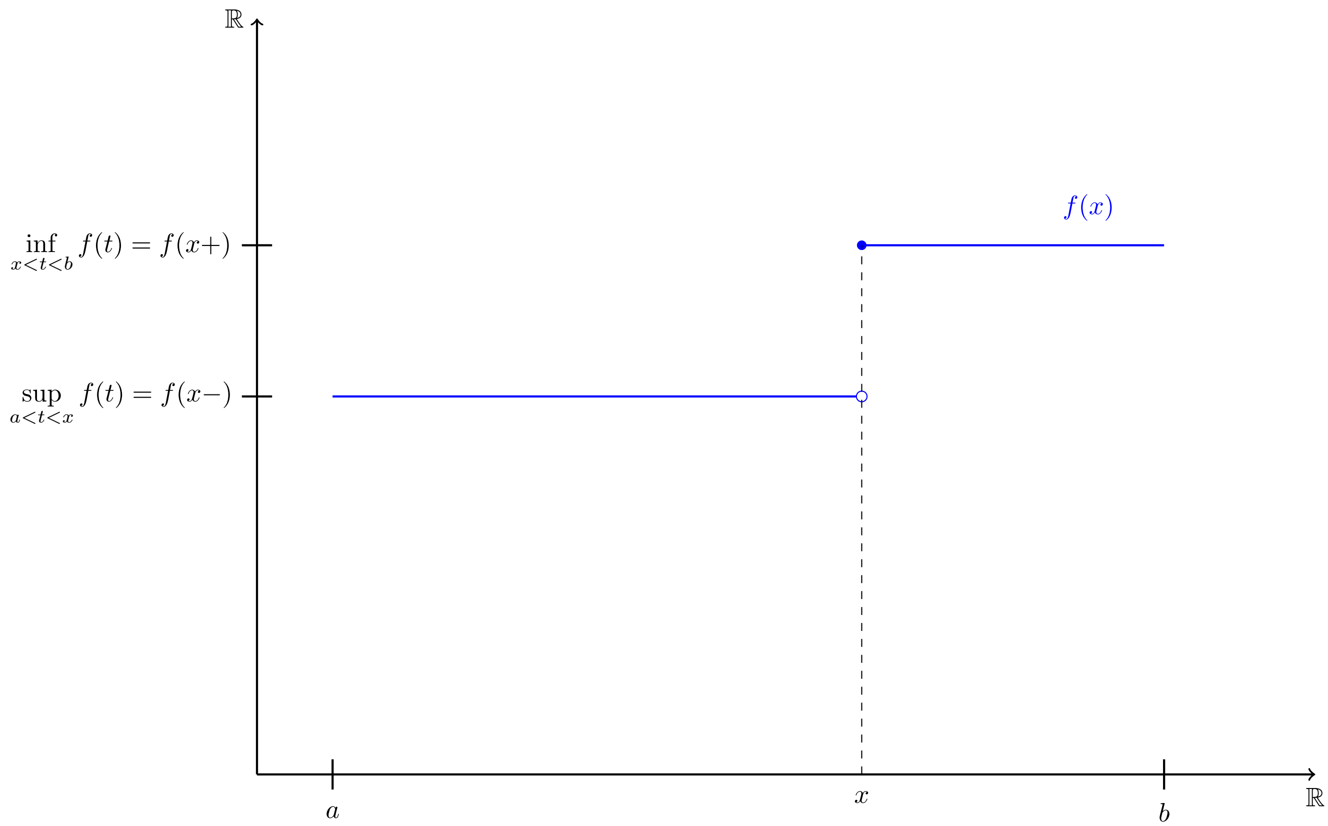

Figure 48: The first inequality of Proposition 4.1 illustrated.

Proposition 4.2. Let f be a monotonic function on (a, b).The set of points (a, b) at which f is discontinuous is at most countable.

Proof. We will prove this result for a monotonically increasing function. Let E be the set of points at which f is discontinuous. To show E is countable, it suffices to show that we can write a bijection from E to some countable set.[67]

The rationals are dense in $\mathbb{R}$, so we can always find a rational in number in between $f(x-)$ and $f(x+)$. For all $x \in r(x)$ we can associate a rational number $r(x)$ such that

$$f(x-) < r(x) < f(x+).$$

Since $x_1 < x_2$ implies $f(x_1+) \leq f(x_2-)$ by Proposition 4.1, $r(x_1) \neq r(x_2)$ if $x_1 = x_2$. We therefore have a bijection from E to a subset of $\mathbb{Q}$ which is countable.[68] □

4.8 Exercises

min and max continuous

examples of open close compact bounded and complete not being preserved in wrong direction

proof of 4.8 with real functions, examples where inf won't work for a general comapct X.

Uniform cont and cauchy

left and right limits equal

[67] This works because sets having equal cardinality is transitive. See Proposition 1.5.

[68] We used the density of $\mathbb{Q}$ in $\mathbb{R}$ to make the mapping surjective, and then used Proposition 4.1 to insure injectivity.

5 Differentiation

We now will treat the first of two major topics in calculus – differentiation. For this whole section, we will be dealing with real functions of a single variable. Most of the results should be very familiar, and as such the number of examples will be limited. For the most part, differentiation as you saw it in calculus is rigorously defined. The only things that should be new are the proofs of results, and the emphasis put on certain topics. When people first take calculus, most of it is just learning how to take derivatives. That won't be so important here. The emphasis will instead be placed on the theorems, and how to use them to prove certain results.

5.1 The Definition of a Derivative

We begin by defining the rate of change of a function between two points.

Definition 5.1. Let f be a real function defined on $X \subset \mathbb{R}$. For any $x \in X$ we define the *difference quotient* $\phi(t)$ as

$$\phi(t) = \frac{f(t) - f(x)}{t - x}$$

where $t \in X$ and $t \neq 0$.

For a fixed value x, the difference quotient captures the rate of change of a function between points x and t. If X is an open interval, then φ will not be defined at the endpoints of X, as there are excluded from the set. We arrive at the definition of a derivative by letting t tend towards x, thereby let $|x - t|$ going to 0.

Definition 5.2. Let f be a real function defined on X. If $\lim_{t \to x} \phi(t)$ exists, then we write

$$f'(x) = \lim_{t \to x} \phi(t),$$

and call f' the *derivative of f*. The derivative f' is a function defined at every point where $\lim_{t \to x} \phi(t)$ exists. If f' exists at a point x we say f is *differentiable at x*. If f' exists at every point of $E \subset X$, we say f is *differentiable on E*.

We will sometimes refer to $f'(x)$ as the "instantaneous rate of change" of $f(x)$ at x.

Remark 5.1 (Is the Derivative a Paradox?)**.** As this video points out, calling the derivative the "instantaneous rate of change" can be interpreted as a paradox. The phrase "rate of change" implies that something actually changes over a time period. The derivative correspond to a single point. Does something change at one given point in time? The phrase "instantaneous rate of change" should be meaningless. In using it in reference to the derivative, we are giving it a definition that it would otherwise not have in the physical world.

Remark 5.2 (An Equivalent Definition)**.** The definition of $f'(x)$ as given in Definition 5.2, may not be the way you first saw the derivative defined in calculus. Instead, you may have seen

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.$$

This is *perfectly valid*, and the *exact same definition*. We arrive at this equation by letting $t - x = h$, and taking $h \to 0$. The variable h just represents the "distance" (or displacement) over which we are calculating the rate of change. The choice of presenting it as an alternative is based off the fact that Rudin (1976) and Tao (2016a) give the definition of $f'(x)$ in terms of the difference quotient. While all the results will hold regardless of what definition we use, sometimes one definition more be more suitable for a proof or example. In all these cases, we could simply take Definition 5.1, let $t - x = h$, and take $h \to 0$. This step won't be explicitly shown after this, so make sure this sits well. As we'll see later, using the definition where $h \to 0$ will help us build intuition when using the derivative to approximate functions. Much later on when we consider differentiation with multiple variables, we will also opt to use a definition with $h \to 0$.

Notation 5.1. There are many different ways to notate the derivative, the two most common being Lagrange's notation of f', and Leibniz's notation of $\frac{df}{dx}$. We will exclusively use Lagrange's notation for real functions of a single variable. This decision is motivated by the fact that f' does not make reference to the variable used to denote the input of f. We are free to write $f(x)$, $f(t)$, $f(\theta)$, etc. If we use $\frac{df}{dx}$, then we must always write $f(x)$. This is a matter of purely notational flexibility. Another reason to shy away from $\frac{df}{dx}$, is that it encourages one to interpret df and dx as infinitesimally small numbers. Historically, this is how Leibniz interpreted the derivative, but it was never formal in this context. We will make it formal when treating differential forms.

Example 5.1 (The Power Rule)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = x^n$. We will show that $f'(x) = nx^{n-1}$ for all $x \in \mathbb{R}$. First, we should point out that

$$(t-x)^n = (t-x)\left[x^{n-1} + tx^{n-2} + \cdots + t^{n-2}x + t^{n-1}\right] = (t-x)\sum_{k=0}^{n-1} t^k x^{(n-1)-t}.$$

We will need to use this to rewrite the denominator of the difference quotient.

$$\begin{aligned}
f'(x) &= \lim_{t\to x} \frac{f(t)-f(x)}{t-x} \\
&= \lim_{t\to x} \frac{t^n - x^n}{t-x} \\
&= \lim_{t\to x} \frac{(t-x)\left[x^{n-1} + tx^{n-2} + \cdots + t^{n-2}x + t^{n-1}\right]}{t-x} \\
&= \lim_{t\to x} \left[x^{n-1} + tx^{n-2} + \cdots + t^{n-2}x + t^{n-1}\right] \\
&= x^{n-1} + x\cdot x^{n-2} + \cdots + x^{n-2}\cdot x \\
&= nx^{n-1}
\end{aligned}$$

Example 5.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as e^x. What is $f'(x)$?

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{e^{x+h}-e^x}{h} = \lim_{h\to 0} \frac{e^x e^h - e^x}{h} = e^x \cdot \lim_{h\to 0} \frac{e^h - 1}{h}.$$

At this point we will let $u = e^h - 1$ and perform a substitution, noting that $h = \ln(u+1)$.[69] We also have $u \to 0$ as $h \to 0$.[70]

$$f'(x) = e^x\cdot\lim_{u\to 0} \frac{e^h-1}{h} = e^x\cdot\lim_{u\to 0} \frac{u}{\ln(u+1)} = e^x\cdot\lim_{u\to 0} \frac{1}{\frac{1}{u}\cdot\ln(u+1)} = e^x\cdot\lim_{u\to 0} \frac{1}{\left[\ln(u+1)^{\frac{1}{u}}\right]} = e^x\cdot\frac{1}{\ln\left[\lim\limits_{u\to 0}(u+1)^{\frac{1}{u}}\right]}$$

Why are we able to pass the limit into the natural log?[71] Finally note that

$$\lim_{u\to 0}(u+1)^{1/u} = \lim_{n\to\infty}(1+1/n)^{1/n} = e,$$

which we first took as a fact in Example 3.7.

$$f'(x) = e^x \cdot \frac{1}{\ln\left[\lim\limits_{u\to 0}(u+1)^{\frac{1}{u}}\right]} = e^x \cdot \frac{1}{\ln(e)} = e^x \cdot \frac{1}{1} = e^x$$

We have that the derivative of e^x is e^x.

[69] I find substitutions like this very unsettling. It feels like another "trick" used in proofs that you would only know if you've proved the result before. I also could be very alone in this regard, and just don't think "maybe I should try substitution" as much as I should.

[70] The limit of $u = e^h - 1$ as $h \to 0$ is $e^0 - 1 = 0$ because e^x is continuous. This gives $u \to 0$ as $h \to 0$.

[71] Because the natural log is continuous, and you can bring limits into continuous functions, as continuity preserves limits.

Theorem 5.1 (Differentiability implies Continuity). Let f be defined on X. If f is differentiable at $x \in X$, then f is continuous at x.

Proof. We can use Theorem 4.3, and show $\lim_{t\to x} f(t) = f(x)$ if f is differentiable at x. We have

$$\begin{aligned} f(t) - f(x) &= \frac{f(t)-f(x)}{t-x}(t-x) \\ \lim_{t\to x}[f(t)-f(x)] &= \lim_{t\to x} \phi(t)(t-x) \\ \lim_{t\to x} f(t) - f(x) &= f'(x)\cdot 0 \\ &= 0. \end{aligned}$$

If $\lim\limits_{t\to x} f(t) - f(x) = 0$, then $\lim\limits_{t\to x} f(t) = f(x)$. □

Where would this proof go wrong if we did not know f is differentiable at x?[72] It is quite easy to come up with examples where the converse of Theorem 5.1 fails.

Example 5.3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = |x|$. This function is continuous at 0 (and all of $\mathbb{R}$), but is not differentiable at 0. We have

$$\phi(t) = \frac{f(t)-f(0)}{t-0} = \frac{|t|}{t}.$$

If we take the left-hand limit and right-hand limit of the difference quotient, we will find they do not agree.

$$\begin{aligned} \lim_{t\to 0^+} \phi(t) &= \lim_{t\to 0^+} \frac{|t|}{t} = \lim_{t\to 0^+} \frac{t}{t} = \lim_{t\to 0^+} 1 = 1 \\ \lim_{t\to 0^-} \phi(t) &= \lim_{t\to 0^-} \frac{|t|}{t} = \lim_{t\to 0^-} \frac{-t}{t} = \lim_{t\to 0^-} 1 = -1 \end{aligned}$$

These two limits agreeing is both a necessary and sufficient condition for the limit existing. We therefore have that $\lim_{t\to 0} \phi(t)$ does not exist, so f' is undefined at 0. Note that f is still differentiable on $\mathbb{R}\backslash\{0\}$.

5.2 Familiar Properties of the Derivative

We will now prove the rules of differentiation that are presented in a calculus course. After proving the most basic properties, we will introduce the Chain Rule, and a basic version of the Inverse Function Theorem, two results that are not immediately obvious.

Theorem 5.2. Suppose f and g are defined on X, and are differentiable at $x \in X$.

1. If $f(x) = c$ for some constant $c \in \mathbb{R}$, then $f'(x) = 0$.
2. $(cf)'(x) = cf'(x)$.
3. (Sum Rule) $(f+g)'(x) = f'(x) + g'(x)$.
4. (Product Rule) $(fg)'(x) = f'(x)g(x) + g'(x)f(x)$.
5. (Quotient Rule) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x)-g'(x)f(x)}{g(x)^2}$.

Proof. The functions f and g are differentiable at x. By Theorem 5.1, these functions are continuous at x, so we will have $\lim_{t\to x} g(t) = g(x)$ and $\lim_{t\to x} f(t) = f(x)$ by Theorem 4.3. We will use these equalities in the proof of the Product Rule and the Quotient Rule.

1. $$\lim_{t\to x} \phi(t) = \lim_{t\to x} \frac{f(t)-f(x)}{t-x} = \lim_{t\to x} \frac{c-c}{t-x} = \lim_{t\to x} \frac{0}{t-x} = \lim_{t\to x} 0 = 0$$

[72] If $\lim\limits_{t\to x} \phi(t)$ did not exist, then there would have been no way to calculate $(\lim\limits_{t\to x} \phi(t))\cdot 0$.

2.

$$\lim_{t\to x}\phi(t)=\lim_{t\to x}\frac{cf(t)-cf(x)}{t-x}=\lim_{t\to x}c\cdot\frac{f(t)-f(x)}{t-x}=c\cdot\lim_{t\to x}\frac{f(t)-f(x)}{t-x}=cf'(x)$$

3.

$$\begin{aligned}\lim_{t\to x}\phi(t)&=\lim_{t\to x}\frac{(f+g)(t)-(f+g)(x)}{t-x}\\&=\lim_{t\to x}\frac{[f(t)+g(t)]-[f(x)+g(x)]}{t-x}\\&=\lim_{t\to x}\frac{f(t)-f(x)+g(t)-g(x)}{t-x}\\&=\lim_{t\to x}\frac{f(t)-f(x)}{t-x}+\lim_{t\to x}\frac{g(t)-g(x)}{t-x}\\&=f'(x)+g'(x)\end{aligned}$$

4.

$$\begin{aligned}\lim_{t\to x}\phi(t)&=\lim_{t\to x}\frac{(fg)(t)-(fg)(x)}{t-x}\\&=\lim_{t\to x}\frac{f(t)g(t)-f(x)g(x)+0}{t-x}\\&=\lim_{t\to x}\frac{f(t)g(t)-f(x)g(x)+[f(t)g(x)-f(t)g(x)]}{t-x}\\&=\lim_{t\to x}\frac{g(x)[f(t)-f(x)]+f(t)[g(t)-g(x)]}{t-x}\\&=g(x)\cdot\lim_{t\to x}\frac{f(t)-f(x)}{t-x}+\lim_{t\to x}f(t)\cdot\frac{g(t)-g(x)}{t-x}\\&=g(x)f'(x)+\lim_{t\to x}f(t)\cdot\lim_{t\to x}\frac{g(t)-g(x)}{t-x}\\&=f'(x)g(x)+g'(x)f(x)\end{aligned}$$

5.

$$\begin{aligned}\lim_{t\to x}\phi(t)&=\lim_{t\to x}\frac{(f/g)(t)-(f/g)(x)}{t-x}\\&=\lim_{t\to x}\frac{\frac{f(t)}{g(t)}-\frac{f(x)}{g(x)}+0}{t-x}\\&=\lim_{t\to x}\frac{\frac{f(t)}{g(t)}-\frac{f(x)}{g(x)}+\left[\frac{f(x)}{g(t)}-\frac{f(x)}{g(t)}\right]}{t-x}\\&=\lim_{t\to x}\frac{\frac{f(t)}{g(t)}-\frac{f(x)}{g(t)}+\frac{f(x)}{g(t)}-\frac{f(x)}{g(x)}}{t-x}\\&=\lim_{t\to x}\frac{\frac{1}{g(x)}\left[\frac{f(t)g(x)}{g(t)}-\frac{f(x)g(x)}{g(t)}\right]-\left[\frac{f(x)g(t)}{g(x)g(t)}-\frac{f(x)g(x)}{g(t)g(x)}\right]}{t-x}\\&=\lim_{t\to x}\frac{\frac{g(x)}{g(x)g(t)}\,[f(t)-f(x)]-\frac{f(x)}{g(x)g(t)}\,[g(t)-g(x)]}{t-x}\\&=\lim_{t\to x}\frac{1}{g(x)g(t)}\left[g(x)\cdot\lim_{t\to x}\frac{f(t)-f(x)}{t-x}-f(x)\cdot\lim_{t\to x}\frac{g(t)-g(x)}{t-x}\right]\\&=\frac{1}{g(x)g(x)}[g(x)f'(x)-f(x)g'(x)]\\&=\frac{f'(x)g(x)-g'(x)f(x)}{g(x)^2}\end{aligned}$$

$\square$

Example 5.4 (Linearity of Derivatives)**.** If we combine the Sum Rule and Part 2 of Theorem 5.2, we have that the operation of differentiation is linear. If f and g are defined on $[a,b]$ and are differentiable at $x \in [a,b]$, then for constants $c, d \in \mathbb{R}$ we have

$$(cf + dg)'(x) = cf'(x) + dg'(x).$$

Example 5.5 (Difference Rule)**.** Suppose f and g are defined on $[a,b]$, and are differentiable at $x \in [a,b]$. We can derive[73] the Difference Rule by using the linearity of differentiation.

$$(f - g)'(x) = (f + (-1)g)'(x) = f'(x) + (-1)g'(x) = f'(x) - g'(x)$$

Example 5.6 (Polynomials)**.** If $\mathscr{P}(\mathbb{R})$ is the set of all polynomials with real coefficients, then every element of $\mathscr{P}(\mathbb{R})$ is differentiable on all of $\mathbb{R}$. We can show this using the Power Rule, Sum Rule, and Part 2 of Theorem 5.2. We can write $p(x) \in \mathscr{P}(\mathbb{R})$ as

$$p(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n = \sum_{k=0}^{n} a_k x^k$$

for $a_k \in \mathbb{R}$ for all k. We have

$$p'(x) = a_1 + 2a_2 x + \cdots + (n-1)a_{n-1}x^{n-2} + na_n x^{n-1} = \sum_{k=1}^{n} k a_k x^{k-1}.$$

This is define for all $x \in \mathbb{R}$. The ease at which we can differentiate polynomials is one of the reasons we like working with them.

One of the benefits of composition preserving the continuity of two functions (Theorem 4.4) is that it means most function we work with are continuous. If you were tasked with writing down some random function, it could most likely would be the result of a composition of simpler functions. The ubiquity of functions composed via other functions makes it useful to know how to differentiate such functions. This is where the Chain Rule comes in. Before formally stating and proving the Chain Rule, it's worth building a ***non-technical***[74] intuition as to why it works.

Suppose you have a real function $g(y)$, and you want to find the rate of change with. We know that this rate is given as $g'(y)$. Now let there be a second function $f(x)$. You're now challenged with finding the rate of change of g with respect to x, after taking $y = f(x)$ to be an input of g. That is, we need to find $h'(x)$ for $h(x) = g(f(x))$. We have two pieces of information: $g'(y)$ (how g changes with respect to its input), and $f'(x)$ (how f changes with respect to its input). To measure the change of g at $f(x)$ with respect to x, we need to somehow convert a measure of change in $y = f(x)$ to a change in x. This can be achieved by using the change in y with respect to x as an exchange rate. If we think of this as a problem related to converting between two measures of change, then it seems the logical thing to do would be multiply g's change with respect to $y = f(x)$ by the conversion rate of $y = f(x)$'s change with respect to x.

$$g'(y) \cdot f'(x) = \frac{\text{change in } g}{\text{change with respect to } y = f(x)} \times \frac{\text{change in } y = f(x)}{\text{change with respect to } x} = \frac{\text{change in } g}{\text{change with respect to } x}$$

Writing everything in terms of x gives us $g'(f(x))f'(x)$.

Remark 5.3 (I Broke the Rules)**.** Okay so this explanation of the Chain Rule is kind of wrong. In Notation 5.1, I said that I would be avoiding the notation of $\frac{dy}{dx}$, as it encourages people to think of dy and dx as their own numbers which form a fraction. This is not what derivatives are, but I just treated them as fractions when explaining the Chain Rule. In fact, had I used Leibniz's notation, the Chain Rule becomes very aesthetically pleasing and easy to remember:

$$\frac{dg}{dx} = \frac{dg}{dy}\frac{dy}{dx}.$$

[73] Pun intended.

[74] I'll explain why my explanation is technically wrong afterwards.

Contrary to what it looks like, we aren't multiplying fractions when we write this. Treating dx and dy like their own numbers can lead to problems and at this point lacks the rigor that analysis aims to achieve. Remark 10.1.17 in Tao (2016a) makes similar points about this issue.

Theorem 5.3 (The Chain Rule)**.** Let X and Y be subsets of $\mathbb{R}$. If $f : X \to \mathbb{R}$ is continuous on X, $f'(x)$ exists at some point $x \in X$, and $g : Y \to \mathbb{R}$ is differentiable at $f(x)$. If we define $h(t) = g(f(t))$, then h is differentiable at x, and

$$h'(x) = g'(f(x))f'(x).$$

Proof. Let $y = f(x)$. The functions f and g are differentiable at x and y respectively.

$$\begin{aligned} f'(x) &= \lim_{t \to x} \frac{f(t) - f(x)}{t - x} \\ g'(y) &= \lim_{s \to y} \frac{g(s) - g(y)}{s - y} \end{aligned}$$

Define $u(t)$ and $v(s)$ as

$$u(t) = \begin{cases} \frac{f(t)-f(x)}{t-x} - f'(x) \text{ if } t \neq x \\ 0 \text{ if } t = x \end{cases},$$
$$v(s) = \begin{cases} \frac{g(s)-g(y)}{s-y} - g'(s) \text{ if } s \neq y \\ 0 \text{ if } s = y \end{cases},$$

for $t \in X$ and $s \in Y$. We have $\lim_{t \to x} u(t) = 0$ and $\lim_{s \to y} v(s) = 0$. These functions allow to rewrite the derivatives of g and f as

$$f(t) - f(x) = (t - x)[f'(x) + u(t)], \tag{7}$$
$$g(s) - g(y) = (s - y)[g'(y) + v(s)], \tag{8}$$

as we take the limits $t \to x$ and $s \to y$.[75] If we let $f(t) = s$ in (8), we can use (7) and (8) to write

$$\begin{aligned} h(t) - h(x) &= g(f(t)) - g(f(x)) \\ &= [f(t) - f(x)][g'(y) + v(s)] \\ &= (t - x)[f'(x) + u(t)][g'(y) + v(f(t))] \\ \frac{h(t) - h(x)}{(t - x)} &= [f'(x) + u(t)][g'(y) + v(f(t))] \end{aligned}$$

for $t \neq x$. If we let $t \to x$, then

$$\begin{aligned} h'(x) &= \lim_{t \to x} \frac{h(t) - h(x)}{(t - x)} \\ &= \lim_{t \to x} [f'(x) + u(t)][g'(y) + v(f(t))] \\ &= \left[f'(x) + \lim_{t \to x} u(t)\right] \left[g'(y) + \lim_{t \to x} v(f(t))\right] \\ &= [f'(x)]\,[g'(y) + v(y)] \\ &= f'(x)g'(y) \\ &= g'(f(x))f'(x). \end{aligned}$$

Note that to conclude $\lim_{t \to x} v(f(t)) = v(y)$, it must be the case that $\lim_{t \to x} f(t) = f(x) = y$. This follows by the assumed continuity of f on X. $\square$

[75] Equation (78) is simply $f(t) - f(x) = (t - x)\left[f'(x) + \frac{f(t)-f(x)}{t-x} - f'(x)\right]$ when $t \neq x$. When we let $t \to x$ we have $u(t) \to 0$, and we arive at $f(t) - f(x) = (t - x)f'(x)$ as $t \to x$. In other words, $f'(x) = \lim_{t \to x} \frac{f(t)-f(x)}{t-x}$, which is true. Equation (7) is therefore perfectly valid. The same argument works for Equation (7).

Our last rule concerning the calculation of derivatives pertains to the inverse of a function. If we let X and Y be subsets of $\mathbb{R}$, we may want to know the derivative of $f^{-1} : Y \to X$ at a point $y = f(x)$, but only know the derivative of $f : X \to Y$ at x. If we know that f^{-1} is differentiable, then this is a straightforward application of the chain rule and composition of f^{-1} and f.

$$\begin{aligned} (f^{-1} \circ f)'(x) &= (f^{-1})'(f(x))f'(x) \\ 1 &= (f^{-1})'(y)f'(x) \\ (f^{-1})'(y) &= \frac{1}{f'(x)} \end{aligned}$$

There is one shortcoming in this approach. In order to use the Chain Rule, we needed to assume that f^{-1} was differentiable. If we do not know beforehand that f^{-1} is invertible, then perhaps we cannot conclude $(f^{-1})'(y) = \frac{1}{f'(x)}$. As it turns out, we still can by The Inverse Function Theorem, which says that knowing f^{-1} is continuous suffices to conclude $(f^{-1})'(y) = \frac{1}{f'(x)}$. This result is rarely given this name in the context of a single variable, and in fact it may not show up in an analysis course to begin with. It's not very special for functions in $\mathbb{R}$. A far more general version of the Inverse Function Theorem will hold with functions in several variables. The introduction here is just to give some early exposure to the idea of the theorem (which I'm calling a proposition here as an allusion to it not being the full-fledged Inverse Function Theorem).

Proposition 5.1 (The Inverse Function Theorem)**.** Let X and Y be subsets of $\mathbb{R}$, and $f : X \to Y$ be an invertible function with inverse $f^{-1} : Y \to X$. Suppose $x \in X$ and $y \in Y$ such that $f(x) = y$. If f is differentiable at x_0, f^{-1} is continuous at y_0, and $f'(x_0) \neq 0$, then f^{-1} is differentiable and

$$(f^{-1})'(y) = \frac{1}{f'(x)}.$$

Proof. We need to show that

$$\lim_{s \to y} \frac{f^{-1}(s) - f^{-1}(y)}{s - y} = \frac{1}{f'(x)}.$$

By the assumed continuity of f^{-1}, we can show

$$\lim_{n \to \infty} \frac{f^{-1}(s_n) - f^{-1}(y)}{s_n - y} = \frac{1}{f'(x)},$$

for a sequence $\{s_n\}$ in Y which converges to s. Let $t_n = f^{-1}(s_n)$ be a sequence in X.By continuity, we have

$$\lim_{n \to \infty} t_n = \lim_{n \to \infty} f^{-1}(s_n) = f^{-1}\left(\lim_{n \to \infty} s_n\right) = f^{-1}(s) = t.$$

Because f is differentiable (and therefore continuous) at x, we have

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} = \lim_{n \to \infty} \frac{f(t_n) - f(x)}{t_n - x} = f'(x).$$

This difference quotient is not zero since $t_n \neq x$, and $f(t_n) \neq f(x)$ (f is a bijection after all). We can use the limit laws of Theorem 3.2 to invert the difference quotient and arrive at

$$\lim_{n \to \infty} \frac{t_n - x}{f(t_n) - f(x)} = \frac{1}{f'(x)}.$$

But since $t_n = f^{-1}(s_n)$ and $x = f^{-1}(y)$, we have $\lim\limits_{n \to \infty} \frac{f^{-1}(s_n) - f^{-1}(y)}{s_n - y} = \lim\limits_{s \to y} \frac{f^{-1}(s) - f^{-1}(y)}{s - y} = \frac{1}{f'(x)}$. $\square$

5.3 Local Extrema

The derivative gives us a useful tool for finding *local* minima and maxima of a function.

Definition 5.3. Let X be a subset of $\mathbb{R}$. The function $f : X \to Y$ has a *local maximum* at a point $p \in X$ if there exists $\delta > 0$ such that $f(x) \leq f(p)$ for all $x \in X$ with $d(p, x) < \delta$.

Definition 5.4. Let X be a subset of $\mathbb{R}$. The function $f : X \to Y$ has a *local minimum* at a point $p \in X$ if there exists $\delta > 0$ such that $f(x) \geq f(p)$ for all $x \in X$ with $d(p, x) < \delta$.

We use the word local to emphasize the fact that local extrema need not be *the* maximum or *the* minimum on X. We see this in Figure 49, f has a local maximum at p_0, despite not having a *global* maximum at p_0. All we care about is that there is some small neighborhood of X over f attains a maximum at p_0. This neighborhood is the set of all $x \in X$ such that $d(p_0, x) < \delta$ for $\delta > 0$, which can also be written as $N_\delta(p_0) \subset X$. We can relate the local extrema of a function to its derivative. Note that the proposition will

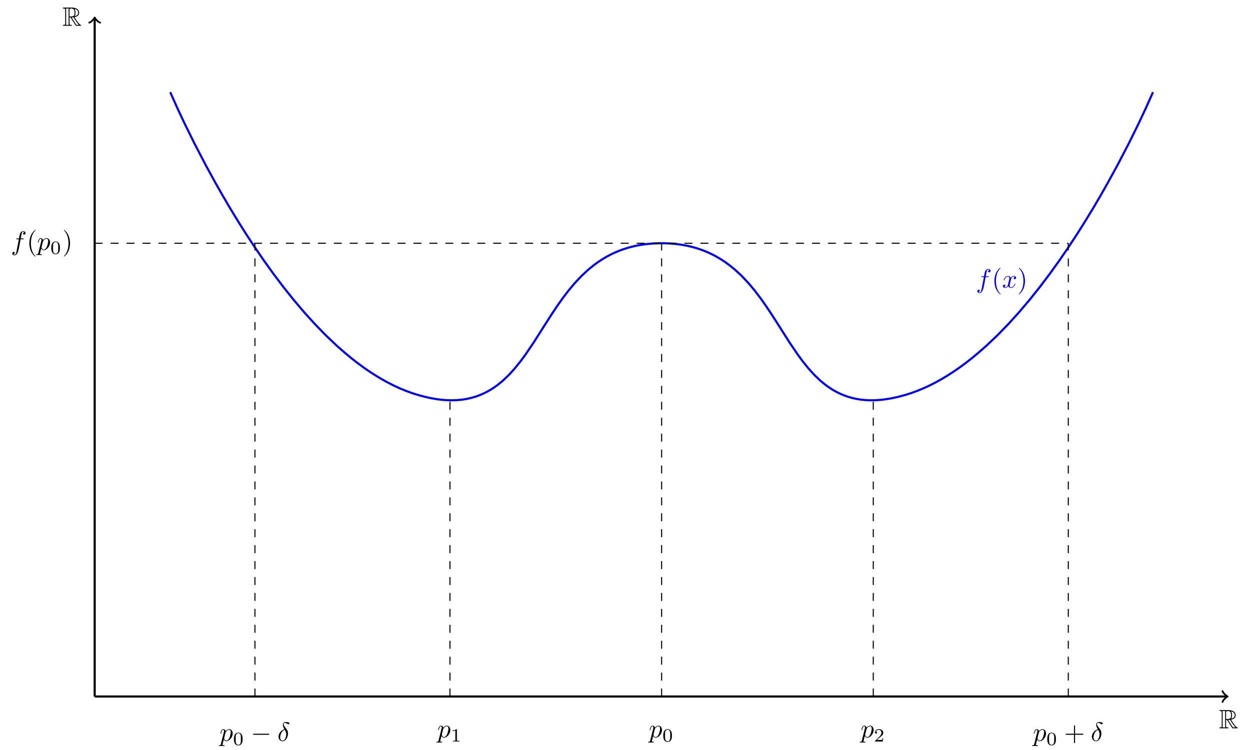

Figure 49: The function has a local maximum at p_0, and local minima at p_1 and p_2. The points $p_0 - \delta$ and $p_0 + \delta$ correspond to the endpoints of the interval $(p_0 - \delta, p_0 + \delta)$. We have $f(x) \leq f(p_0)$ for all $x \in (p_0 - \delta, p_0 + \delta)$, as prescribed in Definition 5.3.

present is a necessary condition for local extrema, but not a sufficient condition. Our main use for this will be proving other theorems. In many applied settings which rely on optimization, this result is assigned more significance.

Proposition 5.2 (Fermat's Theorem for Extrema)**.** Let $X = [a, b]$ be a subset set of $\mathbb{R}$. If $f : X \to \mathbb{R}$ has a local extrema at a point $p \in (a, b)$, and if $f'(p)$ exists, then $f'(p) = 0$.

Proof. We will prove the case for local maxima. Let $p \in (a, b)$ be a local maximum. There exists some δ such that $f(p) \geq f(x)$ for all x satisfying

$$x < p - \delta < x < p + \delta < b.$$

If $p - \delta < t < p$, then

$$\frac{f(t) - f(p)}{t - p} \geq 0$$

as $f(p) \geq f(t)$, and $p > t$. This means

$$\lim_{t \to p} \frac{f(t) - f(p)}{t - p} = f'(p) \geq 0.$$

If instead $p < t < p + \delta$, then

$$\frac{f(t) - f(p)}{t - p} \leq 0$$

as $f(p) \geq f(t)$, and $p < t$. This means

$$\lim_{t \to p} \frac{f(t) - f(p)}{t - p} = f'(p) \leq 0.$$

If $f'(p) \leq 0$ and $f'(p) \geq 0$, then $f'(p) = 0$. □

As the next three examples show, there are several drawbacks to Fermat's Theorem.

Example 5.7 (Saddle Point)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be the function $f(x) = x^3$. We have $f'(0) = 0$, but the point 0 is neither a local minimum nor a local maximum. This follows from the fact that Lemma 5.1 is a necessary condition, but not a sufficient condition.

Example 5.8 (Endpoints)**.** Let $f : [0, 1] \to \mathbb{R}$ be the function $f(x) = x$. This function has a local maximum at 1 and local minimum at 0. Lemma 5.1 does not hold in this case, because it assumes that $p \in (a, b)$.

Example 5.9 (Not Differentiable)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be the function $f(x) = |x|$. This function has a local minimum at 0, but $f'(0)$ is undefined. Lemma 5.1 does not hold in this case, as it assumes that $f'(p)$ exists.

5.4 Mean Value Theorems

As far as analysis is concerned, the Mean Value Theorem is the most important result involving differentiation. When discussing the Intermediate Value Theorem, it was noted that it gave us a way to "convert" information about f to information about the domain of f. A similar metaphor holds for the Mean Value Theorem, as it allows us to gather some information about f' is we know information about f (or vice-versa). If we know that f is differentiable on some interval $[a, b]$, then the Mean Value Theorem tells use the value of $f'(c)$ for $c \in [a, b]$, namely it is

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

We will present three versions of the Mean Value Theorem, each more general than the last.

Theorem 5.4 (Rolle's Theorem)**.** Let $a < b$, and let $f : [a, b] \to R$ be a continuous function which is differentiable on (a, b). Suppose also that $g(a) = g(b)$. Then there exists an $x \in (a, b)$ such that $g'(x) = 0$.

Proof. We have that f is continuous on $[a, b]$. By the Extreme Value Theorem (Theorem 4.7), f attains a maximum M at some point $x \in [a, b]$, and a minimum m at some point $y \in [a, b]$.

Case 1. If $x, y \in \{a, b\}$,[76] then f is constant on $[a, b]$ because $f(a) = f(b)$, so $f'(z) = 0$ for all $z \in (a, b)$.

Case 2. Suppose x is not an endpoint of $[a, b]$. We have a local maximum of f at x, so by Fermat's Theorem (Proposition 5.2), $f'(x) = 0$.

Case 3. Suppose y is not an endpoint of $[a, b]$. We have a local minimum of f at y, so by Fermat's Theorem (Proposition 5.2), $f'(y) = 0$.

□

[76] That is, they are the endpoints of the interval $[a, b]$.

Corollary 5.1 (Mean Value Theorem)**.** Let $a < b$, and let $f : [a, b] \to R$ be a continuous function which is differentiable on (a, b). Then there exists an $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof. Define a new function as

$$h(x) = f(x) - \frac{f(b) - f(a)}{b - a} \cdot x.$$

The function h is continuous on $[a, b]$, differentiable on (a, b),[77] and was defined so $h(a) = h(b)$. By Rolle's Theorem, there exists a $c \in (a, b)$ such that $h'(c) = 0$.

$$h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0 \implies f'(c) = \frac{f(b) - f(a)}{b - a}$$

□

A geometric interpretation of the Mean Value Theorem is found in Figure 50. The Mean Value Theorem

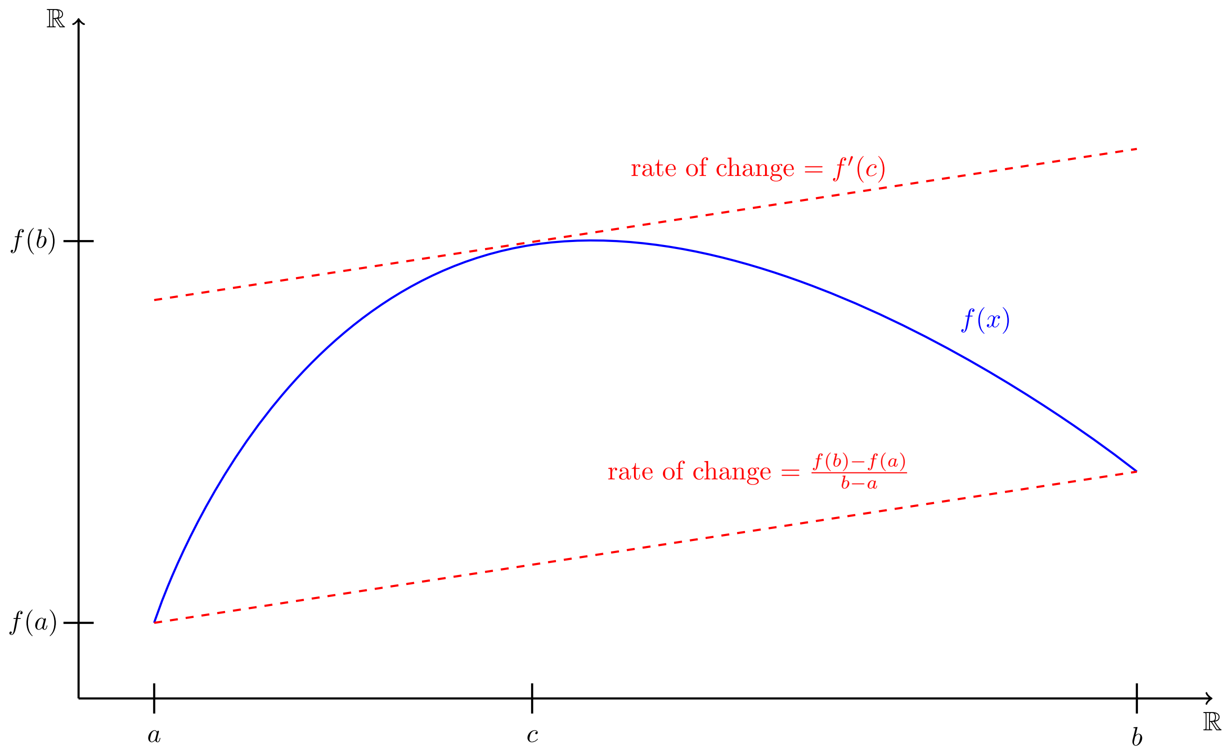

Figure 50: The Mean Value Theorem.

can be extended to the case of parametric curves.

Proposition 5.3 (Cauchy's Mean Value Theorem)**.** Let $a < b$, and let $f : [a, b] \to R$ and $g : [a, b] \to \mathbb{R}$ be continuous functions which are differentiable on (a, b). Then there exists an $c \in (a, b)$ such that

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

[77]The continuity and differentiability of h follow immediately from f.

Proof. Like we did while proving the Mean Value Theorem, define a function

$$h(x) = f(x) - \frac{f(b) - f(a)}{g(b) - g(a)} \cdot g(x).$$

If we apply Rolle's Theorem, the result follows. $\square$

Figure 51 illustrates Cauchy's Mean Value Theorem.

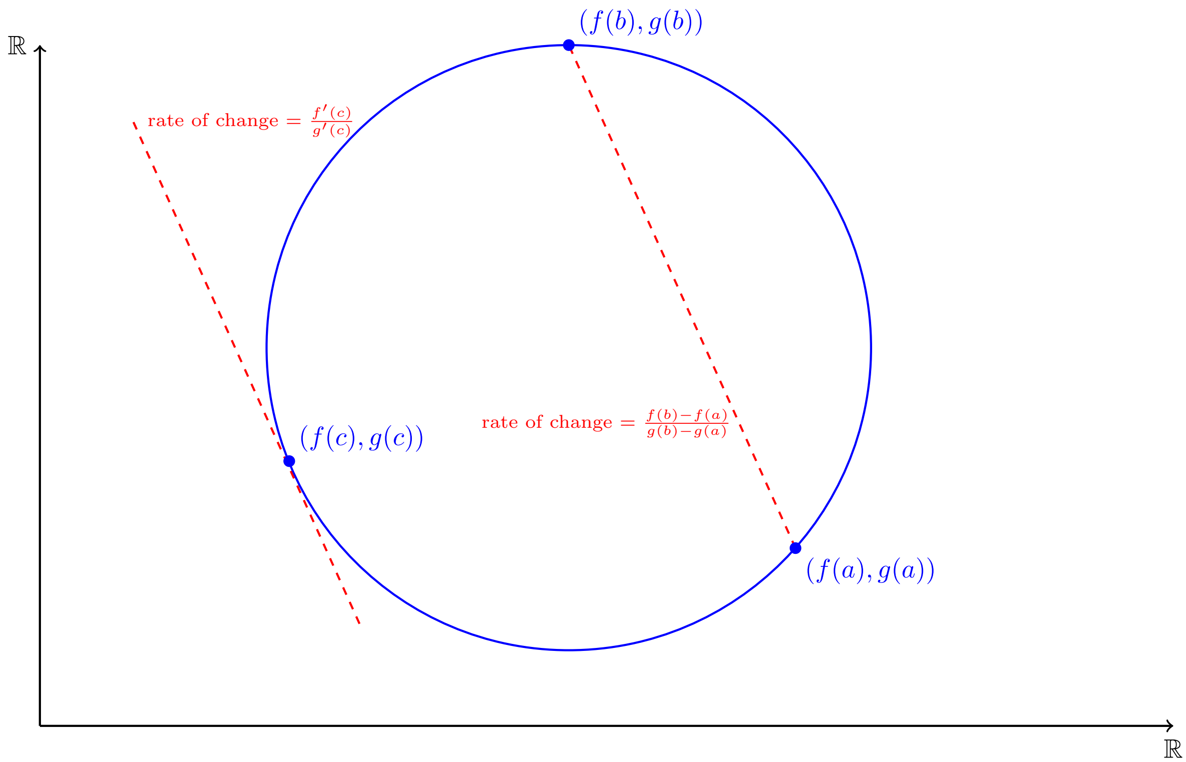

Figure 51: Cauchy's Mean Value Theorem. Note that the parameterization begins at $(f(a), g(a))$ and then is orientated clockwise.

Example 5.10. Let $f(t) = \cos(t)$ and $g(t) = \sin t$. If we let $a = 0$ and $b = \pi/2$. Then we can find a value of $c \in (0, \pi/2)$ such that

$$\frac{f'(c)}{g'(c)} = \frac{-\sin c}{\cos c} = \frac{0-1}{1-0} = -1.$$

You can verify that $c = \pi/4$ works.

5.5 L'Hôpital's Rule

Now we will prove a classic result from calculus that allows us to calculate a limit using derivatives.

Proposition 5.4 (L'Hôpital's Rule)**.** Suppose f and g are real and differentiable in (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$, where $-\infty \leq a < b \leq \infty$. Suppose we have

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = A.$$

If $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, or if $\lim_{x\to a} g(x) = \infty$, then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = A.$$

Proof. FINISH $\square$

5.6 Higher Order Derivatives

The derivative f' is a function in its own right. This means we could not only differentiate f' itself, but differentiate the resulting function. This idea gives rise to derivatives of higher order.

Definition 5.5. If f has a derivative f' on an interval, and if f' is itself differentiable, we write $(f')' = f''$, and call f'' the *second derivative of f*. Continuing in this manner gives

$$f, f', f'', f^{(3)}, \ldots, f^{(n)},$$

where $f^{(n)}$ is the *nth derivative of f*.

Remark 5.4 (Smoothness)**.** For $f^{(n)}(x)$ to exist at a point x, we need to be able to take the limit

$$\lim_{t \to x} \frac{f^{(n-1)}(t) - f^{(n-1)}(x)}{t - x}.$$

This means that not only is a requirement that $f^{(n-1)}(t)$ is differentiable at x, but also we need $f^{(n-1)}(t)$ to be defined in some neighborhood of x. What does it mean for $f^{(n-1)}(t)$ to exist in s neighborhood of x? It would mean that $f^{(n-2)}$ is differentiable in that neighborhood. This reasoning follows for all $f^{(n-2)}, \ldots, f$. For this reason, the highest order derivative you are able to take of a function f is often associated with how "smooth" it is.

A continuous function is "nice", but it doesn't need to be smooth.[78] A function that is differentiable is smooth. A function that is twice differentiable is even smoother, as it means that f' is differentiable, and therefore continuous! To be formal we say that f is a *smooth function* if it has continuous derivatives up to some desired order. Having a continuous derivative of order $n-1$, is implied by being n times differentiable. In many applied setting this becomes important. For example, you may see theorems assume that a certain function has a continuous second derivative.[79] The smoothest functions are those which are infinitely differentiable.

Example 5.11. Every real polynomial $p(x)$ is infinitely differentiable. Let

$$p(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n = \sum_{k=0}^{n} a_k x^k$$

for $a_k \in \mathbb{R}$ be a polynomial of degree n. We have

$$\begin{aligned}
p'(x) &= a_1 + 2a_2 x + \cdots + (n-1)a_{n-1}x^{n-2} + na_n x^{n-1} \\
p''(x) &= 2a_2 + \cdots + (n-1)(n-2)a_{n-1}x^{n-3} + n(n-1)a_n x^{n-2} \\
&\vdots \\
p^{(n-1)}(x) &= n!a_n x \\
p^{(n)}(x) &= n!a_n \\
p^{(n+1)}(x) &= 0 \\
&\vdots
\end{aligned}$$

The polynomial is still infinitely differentiable even if we keep getting zero, as the zero function is differentiable.

[78] Think of $|x|$ at the points $x = 0$

[79] For example, in statistics and econometrics there is a method of estimation known as maximum likelihood estimation. This method is only "efficient" if we assume that a probability distribution is three times differentiable, i.e it's second derivative is continuous.

Example 5.12. The functions $f(x) = \sin x$ and $g(x) = \cos x$ are infinitely differentiable, and each of their derivatives are given by

$$f^{(n)}(x) = \begin{cases} \cos x \text{ if } n = 1, 5, 9, \ldots \\ -\sin x \text{ if } n = 2, 6, 10, \ldots \\ -\cos x \text{ if } n = 3, 7, 11, \ldots \\ \sin x \text{ if } n = 4, 8, 12, \ldots \end{cases} \qquad g^{(n)}(x) = \begin{cases} -\sin x \text{ if } n = 1, 5, 9, \ldots \\ -\cos x \text{ if } n = 2, 6, 10, \ldots \\ -\sin x \text{ if } n = 3, 7, 11, \ldots \\ \cos x \text{ if } n = 4, 8, 12, \ldots \end{cases}$$

Remark 5.5 (Polynomials and Trig Functions)**.** Polynomials, cos, and sin are not only infinitely differentiable, but they're also reall easy to differentiate. This makes working with them preferable. If we have a complicated function, it would be nice if we could find some way to express it as a polynomial or in terms of sin and cos. This is more of a longterm goal we will return to later.

5.7 Approximation

We now turn to a subject involving differentiation that will return several times – approximation. First we will discuss the derivative in the context of linear approximation of a function, then we will explore how we can get increasingly accurate approximations of a function by using higher order derivatives.

In Remark 5.2, an alternate definition of the derivative was presented.

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

A reformulation of this allows us to write the derivative as the difference quotient (with h) if we introduce some "remainder term".

$$f'(x_0) + hr(h) = \frac{f(x_0 + h) - f(x_0)}{h} \tag{9}$$

As long the ratio $r(h)/h \to 0$ as $h \to 0$, then this is a valid equation, as

$$\begin{aligned} f'(x_0) &= \frac{f(x_0 + h) - f(x_0)}{h} + \frac{r(h)}{h} \\ f'(x_0) &= \lim_{h \to 0} \left[\frac{f(x_0 + h) - f(x_0)}{h} + \frac{r(h)}{h} \right] \\ f'(x_0) &= \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} + 0. \end{aligned}$$

A valid question is, why do we the ratio $r(h)/h$ goes to zero and not just that $r(h) \to 0$? We need the remainder $r(h)$ to "shrink" faster than h, otherwise $r(h)/h \not\to$ as $h \to 0$. In fact, $r(h)/h$ could very well "blow up" if we don't stipulate $r(h)/h \to 0$ as $h \to 0$![80] If this where to happen, then we would not end up with the definition of the derivative. If we multiply by h and rearrange some terms, (9) becomes

$$f(x_0 + h) = f(x_0) + f'(x_0)h + hr(h). \tag{10}$$

Equation (10) may seem random, but it's *very* interesting. We know that $hr(h)$ is going to be small, so we could be informal and write

$$f(x_0 + h) \approx f(x_0) + f'(x_0)h.$$

For a given x_0, this tells us that $f(x_0) + f'(x_0)h$ is an approximation for $f(x_0 + h)$ (f near some point x_0). Furthermore, this approximation is linear, as $f(x_0)$ and $f'(x_0)$ are constants, $f(x_0) + f'(x_0)h$ is a linear function of h. Not only is this a linear approximation, it is the *best* linear approximation we are able to achieve, a fact we can prove.

[80]This would happen if $h \to 0$ "faster" than $r(h) \to 0$.

Notation 5.2 ("Little o")**.** We say that $f(h) = o(g(h))$ (read as *"$f(h)$ is little o of $g(h)$ as $h \to 0$"*) if

$$\lim_{h \to 0} \frac{f(h)}{g(h)} = 0.$$

This means that $f(h)$ tends to zero faster than $g(h)$ as $h \to 0$. For example,

$$\lim_{h \to 0} \frac{h^3}{h} = 0,$$

so we could write $o(h^3) = o(h)$.

When we noted that $r(h)/h \to 0$, what we really were saying was $r(h) = o(h)$.[81] In the case of linear approximation, having a remainder which is $o(h)$ is as good as it gets, because h enters the equation linearly. For example, we cannot end up with some $h^n r(h)$ that vanishes in (10), giving $r(h) = o(h^n)$, because (10) would no longer be a linear function of h. We now can state and prove the result we have been building to.

Theorem 5.5 (Best Linear Approximation)**.** If f is differentiable at x_0, then

$$f(x_0 + h) = f(x_0) + f'(x_0)h + o(h).$$

Conversely, if there exists some linear approximation

$$f(x_0 + h) = B + Ah + o(h),$$

then not only is f differentiable at x_0, but also $B = f(x_0)$ and $A = f'(x_0)$.

Proof. While we did not use the phrase "if and only if" in Theorem 5.5, this theorem is a biconditional statement. A function will have a best linear approximation if and only if it is differentiable.

($\Longrightarrow$) Suppose f is differentiable at x_0. While building up to this theorem, we have already shown that $f(x_0 + h) = f(x_0) + f'(x_0)h + o(h)$, as $r(h) = o(h)$ in (10).

($\Longleftarrow$) Suppose f has a (best) linear approximation of $f(x_0 + h) = B + Ah + o(h)$, for scalers $A, B \in \mathbb{R}$. Setting $h = 0$ gives $f(x_0) = B$. We therefore have

$$f'(x) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0} \frac{f(x_0 + h) - B}{h} = \lim_{h \to 0} \frac{Ah + o(h)}{h} = \lim_{h \to 0} \left[A + \frac{o(h)}{h} \right] = A.$$

□

The first part of Theorem 5.5 tells us that (10) is a good approximation, while the second part tells us that if such a good approximation exists, f is differentiable, and the approximation is given by (10). There is no other way to say it – the second part of this theorem is *****ing* amazing. The simple concept of approximating any function with a line, gives rise to the concept of the derivative.[82] Is this a surprise though?

If you want to approximate a nonlinear function f with a line at a point x_0, you have two degrees of freedom: a point on the line, and the slope of the line.[83] The choice of point is easy. You want to approximate f at x, so you should pick $(x_0, f(x_0))$. Now you have an infinite choice of slopes, which one do you pick? Pick the slope to be the rate of change at that point! This is just $f'(x_0)$ though. You literally[84] cannot make a better choice of point and slope.

[81] One of the weird things about little o notation is we use "=" as a stand in for "is". There is no actual equality when we write $o(h)$ instead of $r(h)$. In general this means when you se $o(h)$ you should think "something here is $o(h)$, the particular form of which is not that important because it is going to zero."

[82] In fact, this is how the derivative is introduced to many people in calculus courses.

[83] Any point and slope uniquely define a line, hence point-slope form from middle school: $y - y_1 = m(x - x_1)$.

[84] And I mean the definition of the word "literally", not "figuratively".

Remark 5.6 (Another Alternate Definition)**.** Theorem 5.5 inspires yet another equivalent definition of differentiation. We can say that f is differentiable at a point x if there exists some scaler $A \in \mathbb{R}$ such that

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - A \cdot h}{h} = 0.$$

In this case we write $f'(x) = A$. You sometimes hear the phrase "locally linear" associated with the existence of the derivative, and it comes from this definition. All we did was subtract $f(x)$ from both sides of the definition given in Remark 5.2. We now are expressing $f'(x)$ as some linear map where $h \mapsto f'(x)h$. This coincides with (10).

Example 5.13. We can show that $f(x) = |x|$ is not differentiable at $x_0 = 0$ using Theorem 5.5 and Remark 5.4. Suppose we had some linear approximation of $|x|$ at the point 0.

$$\begin{aligned} f(x_0 + h) &= B + Ah + r(h) \\ |0 + h| &= B + Ah + r(h) \\ |h| &= B + Ah + \underbrace{(|h| - B - Ah)}_{r(h)} \end{aligned}$$

But do we have $r(h) = o(h)$?

$$\lim_{h \to 0} \frac{r(h)}{h} = \lim_{h \to 0} \frac{(|h| - B - Ah)}{h} = \lim_{h \to 0} \frac{(|h| - B - A)}{h} \neq 0$$

This limit is undefined, as $|h|/h$ has no limit as $h \to 0$. This is the *exact* same limit we attempted to take in Example 5.3, but could not, render $|x|$ not differentiable at 0.

Example 5.14. The equation of the line given in Theorem 5.5 looks a little strange. In calculus, the equation for a tangent line looks a bit like this, but something is off. For a fixed x_0, it's a function of h, where $x_0 + h$ is a point close to x_0. Instead, we can write write this nearby point as $x = x_0 + h$. Equation (10) now becomes

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + r(h) \\ f(x) &\approx f(x_0) + f'(x_0)(x - x_0). \end{aligned}$$

This final equation is what most people see in calculus. Figure 52 shows the approximation for $x_0 + h$ and x.

Our next example not only calculates the best linear approximation for a function at a point, but also serves as motivation for our next theorem.

Example 5.15. Suppose we want to approximate e^x at the point $x_0 = 0$ using Theorem 5.5 (which we can do because e^x is differentiable). Note in this case $x = h + x_0 = h + 0 = h$

$$\begin{aligned} f(x_0 + h) &= f(x_0) + f'(x_0)h + r(h) \\ e^{0+h} &= e^0 + e^0 h + r(h) \\ e^h &= 1 + h + (e^h - 1 - h) \\ e^h &\approx 1 + h \\ e^x &\approx 1 + x. \end{aligned}$$

We won't be able to get a better *linear* approximation than this, but can we get a better approximation?

The linear approximation for f at x_0 is a result of picking the best point and slope to define the approximation. It makes sense that we should pick the point $(x_0, f(x_0))$ and the slope $f'(x_0)$. This is all the information about f we can incorporate into our approximation as far as derivatives go, because the equation

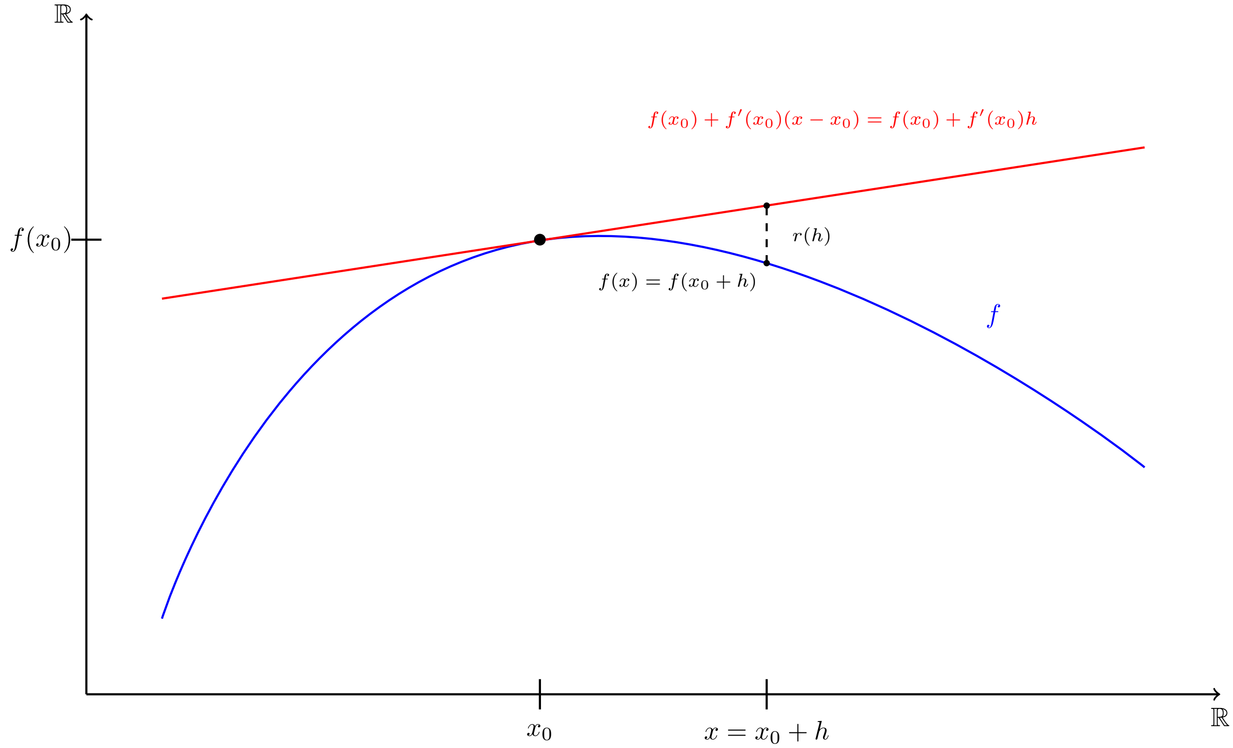

Figure 52: The equation of a tangent line to f at point x_0 is given as $f(x_0)+f'(x_0)(x-x_0)$ or $f(x)+f'(x)h$. The difference between the value of the function and the value of the approximation at a nearby point $x = x_0 + h$ is given by $r(h)$. If we add $r(h)$ to the approximation, then we have an equation for f as given in Equation (10).

of a line *must* have a second derivative of zero.[85] What if we allowed our approximation to have a nonzero second derivative? For e^x, such an approximation would be quadratic. It would look like

$$e^x \approx 1 + x + Cx^2,$$

for a scaler C. We can now incorporate more information about e^x into our approximation, namely we can pick a value of C such that the second derivative of our approximation matches the second derivative of e^x at our point of interest $x_0 = 0$. For $f(x) = e^x$, we have $f''(x) = e^x$, giving $f''(x_0) = f''(0) = 1$. Perhaps we should try $C = 1$.

$$e^x \approx 1 + x + Cx^2$$
$$e^x \approx 1 + x + x^2$$

But if we differentiate our approximation, then we have $(1 + x + x^2)''(0) = 2 \neq 1$. Letting $C = f''(x_0)$ failed to account for the power rule, so we need to scale it by $1/2$ to account for this. Our updated quadratic approximation is

$$e^x \approx 1 + x + \frac{1}{2}x^2.$$

This approximation has the same value, derivative, and second derivative as e^x at $x_0 = 0$. But in what sense is this better than our linear approximation? We can address this with $r(h)$.

$$e^x \approx 1 + x + \frac{1}{2}x^2 + \underbrace{(e^x - 1 - x - \frac{1}{2}x^2)}_{r(h)}$$

[85] All subsequent derivatives will also be 0.

Because $h = x$ in this case, we have $r(h) = e^h - 1 - h - \frac{1}{2}h^2$. Not only do we have $r(h) = o(h)$, but we have $r(h) = o(h^2)$!

$$\lim_{h\to 0} \frac{r(h)}{h^2} = \lim_{h\to 0} \frac{e^h - 1 - h - \frac{1}{2}h^2}{h^2} = 0$$

The remainder is moving to 0 faster than h^2. In a sense, it's shrinking *twice* as fast as h. This makes our quadratic approximation superior to our linear approximation, as our quadratic remainder converges to 0 even faster than that of the linear approximation. But why stop at quadratic, why not cubic, or quartic, *or* quintic? If we repeated this process we would arrive at the following equations:

$$\begin{aligned}
e^x &= 1 + x + o(x) \\
e^x &= 1 + x + \frac{1}{2}x^2 + o(x^2) \\
e^x &= 1 + \frac{1}{2}x^2 + \frac{1}{6}x^3 + o(h^3) \\
e^x &= 1 + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + o(h^4) \\
e^x &= 1 + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5 + o(x^5)
\end{aligned}$$

Each time we increase the degree of our polynomial approximation, we can incorporate the information given by a higher order derivative of f.

The idea of approximating a differentiable function with a polynomial of some fixed degree is formalized in Taylor's Theorem. For this theorem, we will opt to use $(x - x_0)$ instead of h.

Theorem 5.6 (Taylor's Theorem)**.** Suppose f is a real function on $[a, b]$, n is a positive integer, and f is n times differentiable at a point x_0. Then we have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + o((x - x_0)^n).$$

We refer to the polynomial (excluding the remainder term) as the *n-th order Taylor Polynomial of f at x_0*, and write

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}(x - x_0).$$

The proof just amounts to applying L'Hôpital's Rule $n - 1$ tines

Proof. We just need to show that the remainder term is in fact $o((x - x_0)^n)$. If

$$r((x - x_0)) = f(x) - P_n(x)$$

is our remainder term, then we need to verify that

$$\lim_{(x - x_0)\to 0} \frac{f(x) - P_n(x)}{(x - x_0)^n} = 0.$$

By construction, $f^{(k)}(x_0) = P_n^{(k)}(x_0)$ for $k = 0, 1, \ldots n$. This means that our desired limit will give the indeterminate form of $0/0$, so we must use L'Hôpital's Rule.[86] But if we use it once, we still get an

[86] You should verify that we meet the conditions required of Proposition 5.4.

indeterminate form of 0/0. We in fact must apply it $n-1$ times before we do not end up with 0/0:

$$\begin{aligned}
\lim_{(x-x_0)\to 0} \frac{f(x)-P_n(x)}{(x-x_0)^n} &= \lim_{(x-x_0)\to 0} \frac{f'(x)-P_n'(x)}{n(x-x_0)^{n-1}} \\
&= \cdots \\
&= \lim_{(x-x_0)\to 0} \frac{f^{(n-1)}(x)-P_n^{(n-1)}(x)}{n!(x-x_0)} \\
&= \frac{1}{n!} \lim_{x\to x_0} f^{(n)}(x) - P_n^{(n)}(x) \\
&= \frac{1}{n!} f^{(n)}(x_0) - P_n^{(n)}(x_0) \\
&= 0
\end{aligned}$$

Therefore we have $o((x-x_0)^n)$. □

Remark 5.7 (Taylor's Theorem and the MVT)**.** The presentation of Taylor's Theorem Rudin (1976) is slightly different. I suspect his (equally valid) treatment was motivated by the connection between the Mean Value Theorem and Taylor's Theorem. Rudin (1976) explicitly writes "For $n=1$, the [Taylor's Theorem] is just the mean value theorem." If we use Taylor's Theorem to find $P_n(x_0)$ for an $n+1$ differentiable function, we have

$$f(x) = \underbrace{f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n}_{P_n(x)} + \underbrace{\frac{f^{(n+1)}(x_0)}{(n+1)!}(x-x_0)^{n+1}}_{R_n(x)},$$

where $R_n(x)$ is a remainder term. Rudin's version says that under stronger conditions, we can always find some value between $\xi \in (x, x_0)$ which satisfies

$$f(x) = P_n(x) + R_n(\xi).$$

If we do this for $P_0(x)$ and $R_0(x)$, we get

$$f(x) = f(x_0) + f'(\xi)(x-x_0) \implies f'(\xi) = \frac{f(x)-f(x_0)}{x-x_0},$$

which is the Mean Value Theorem.

A natural question that arises with Taylor's Theorem is what happens when we let the $n \to \infty$. If f is infinitely differentiable at x_0, then what is stopping us from writing the Taylor Polynomial as an infinite series? If we do this, will the series have any remainder? If it has no remainder does that mean it simply equals the function at every point it is defined? We will answer all these questions in Section 7!

5.8 Exercises

uniform continuity and second derivative
derivative of x^n using h definition
derivative of a^x
taylor approx is best

6 Riemann Integration

The theory of differentiation may at times have seemed painless. It's eerily similar to calculus. Isn't analysis supposed to be hard and confusing? Fear not, because here comes integration to bring us back to reality! The definition of Riemann integration cannot be given immediately like that of the derivative, as we need to do some prep work. Even once it is defined, proofs involving the Riemann integral are a bit more sophisticated than the average proof up until this point. This is a constant theme in math – integration is *way* harder than differentiation. Our goal is to develop a form of integration that works reasonably well for real functions. Fortunately, this simplifies things, as we're comfortable with real functions. There is good news. The Riemann integral mostly achieves our goal![87] Nevertheless, there will be some drawbacks of Riemann integration, and we will only be able to go so far with it. For this reason, we will return to integration again in Sections 12-18. Consider this a first pass at what turns out to be a much more sophisticated problem.

Why are we even interested in integration though? Many problems are concerned with accumulation over time, which is captured by the area underneath a curve. For example, the probability of a certain event can be interpreted as the area under a probability distribution function. Similar problems exist in nearly every field of science, so it's important that we have some tool for measuring area under the curve of some function.

First, we should ask, when is it even possible to calculate the area under the curve of a function? Is it possible to measure the area over some infinite interval, such as all of $\mathbb{R}$? It is not immeadetly clear how to do this, as we have not yet developed a way of measuring the length of $\mathbb{R}$,[88] so we should stick to some bounded interval $[a,b] \subset \mathbb{R}$. We also want f to be bounded on $\mathbb{R}$, otherwise the area under the curve would not be well defined. For these reasons, **we will restrict our attention to bounded real functions on an interval [a,b]** for this whole section.

6.1 Partitions

First we need to develop some notation about partitioning an interval of the real line into smaller intervals.

Definition 6.1. Let $[a,b] \subset \mathbb{R}$. A *partition* P of $[a,b]$ is a finite set of points $P = \{x_0, x_1, \ldots, x_n\}$ such that

$$a = x_0 \leq x_1 \leq \ldots \leq x_{n-1} \leq x_n = b.$$

We will write $\delta x_i = x+i-x_{i-1}$ for $i = 1, \ldots, n$. We will denote the set of all partitions of $[a,b]$ as $\mathbf{P}([a,b])$.[89]

Figure 53 shows a partition of the interval $[a,b]$. Contrary to what this figure shows, the points which comprise a partition need not be evenly spaced out along $[a,b]$. We can refine a partition by adding more points to $P = \{x_0, x_1, \ldots, x_n\}$.

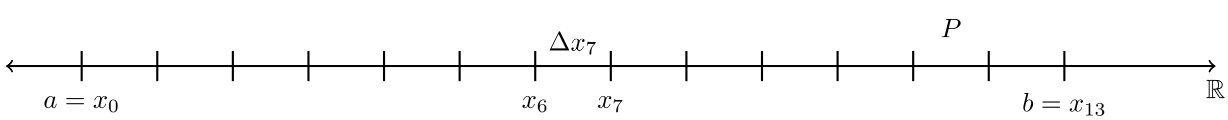

Figure 53: A Partition $P = \{x_0, x_1, \ldots, x_{13}\}$ of the interval $[a,b]$.

Definition 6.2. A partition P^* is *refinement* of P if $P \subset P^*$. Given two partitions, P_1 and P_2, we say P^* is their *common refinement* if $P^* = P_1 \cup P_2$.

[87] There is a reason that it is the only form of integration most people ever need to know and use.
[88] We'll do this in Section 12.
[89] $\mathbf{P}([a,b])$ is *not* standard notation.

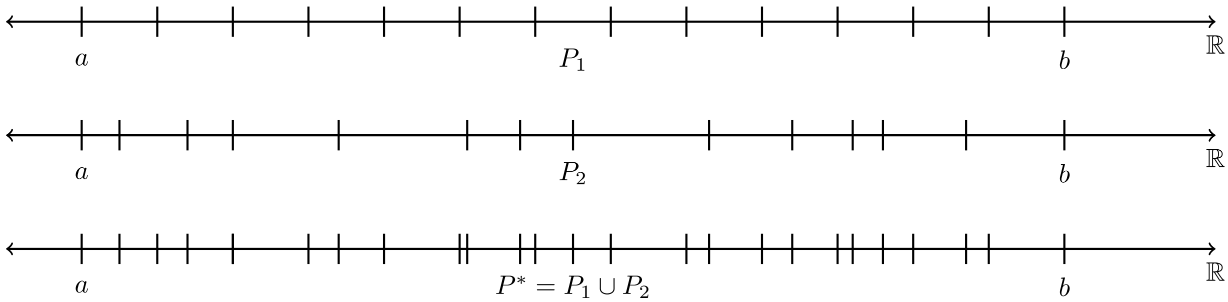

Figure 54: The partition P^* is a common refinement of P_1 and P_2.

6.2 Upper and Lower Riemann Integrals

We will now use partitions to approximate the area under a curve. This will be nearly identical to the method developed in calculus, but with one small difference. Calculus courses normally use left and right Riemann sums, which use the points of a partition to determine the height of the rectangles use to approximate the area under a function. We will instead determine the height of the rectangles using the infimum and supremum of a function.

Definition 6.3. Suppose f is a bounded real function defined on $[a, b]$. Corresponding to each partition P of $[a, b]$, we have

$$\begin{aligned} M_i &= \sup_{x \in [x_{i-1}, x_i]} f(x) \\ m_i &= \inf_{x \in [x_{i-1}, x_i]} f(x) \\ U(P, f) &= \sum_{i=1}^{n} M_i \Delta x_i, \\ L(P, f) &= \sum_{i=1}^{n} m_i \Delta x_i, \end{aligned}$$

We call $U(P, f)$ an *upper Riemann sum* and $L(P, f)$ a *lower Riemann sum*.

Figure 55 shows what $m_i \Delta x_i$ and $M_i \Delta x_i$ may look like for a given function and partition. Figure 56 and Figure 57 show the upper Riemann sum and lower Riemann sum for the partition of $[a, b]$ that was shown in Figure 53, respectively. So far, this is fairly similar to the way integration was developed in calculus, but now we're going to do something a little different. Instead of taking a limit of Riemann sums, we will simply say a function is Riemann integrable if the supremum of all possible upper Riemann sums coincides with the infimum of all the possible lower Riemann sums.

Definition 6.4. Suppose f is a bounded real function on the interval $[a, b]$. Define the values We refer to these as the *lower Riemann integral of f over $[a, b]$* and *upper Riemann integral of f over $[a, b]$* respectively.

The upper and lower Riemann integral will *always* exist for any bounded function on $[a, b]$. The set $\mathbf{P}([a, b]) \neq \emptyset$, as any interval $[a, b]$ has a trivial partition of $\{x_0 = a, x_1 = b\}$. The set is also bounded, as for all $P \in \mathbf{P}([a, b])$,

$$\inf_{x \in [a,b]} f(x) \cdot (b - a) \leq L(P, f) \leq U(P, f) \leq \sup_{x \in [a,b]} f(x) \cdot (b - a).$$

We know the infimum and supremum of f exist on $[a, b]$, because f is bounded. This gives us two nonempty bounded subsets of $\mathbb{R}$ in the form of $\{U(P, f)\}_{P \in \mathbf{P}([a,b])}$ and $\{L(P, f)\}_{P \in \mathbf{P}([a,b])}$. The supremum and infimum of these sets are guaranteed to exist by the completeness of $\mathbb{R}$.

We define the Riemann integral using these upper and lower integrals.

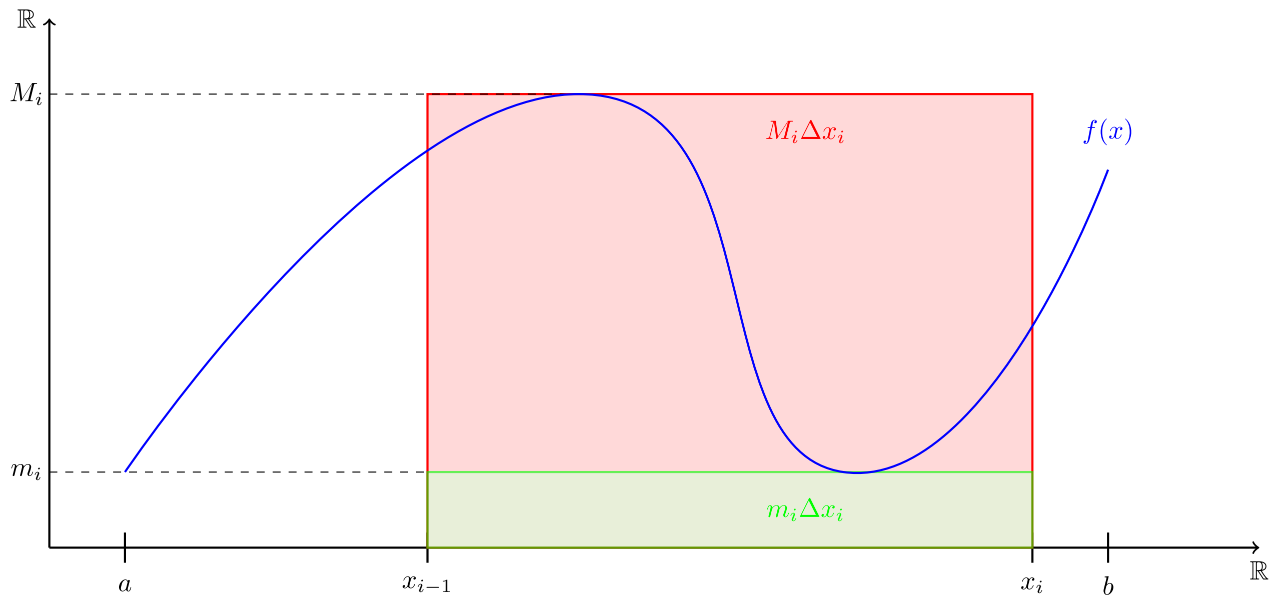

Figure 55: The particular values $m_i\Delta x_i$ and $M_i\Delta x_i$ for some function f and partition of $[a,b]$.

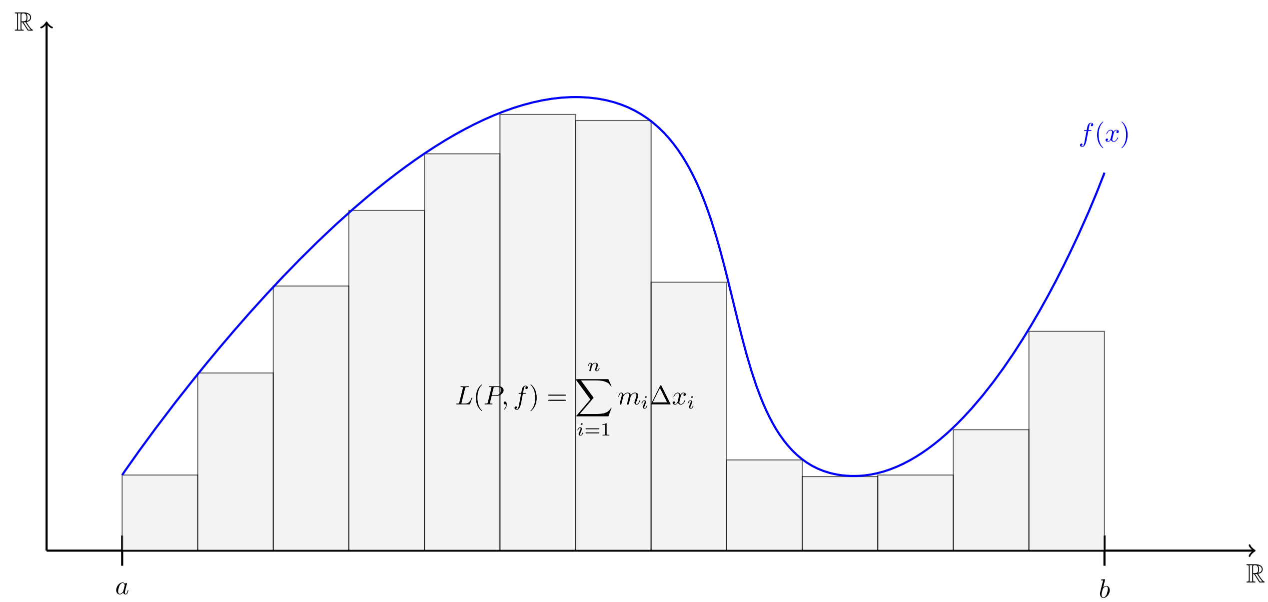

Figure 56: A lower Riemann sum.

Definition 6.5. Suppose f is a bounded real function on the interval $[a,b]$. If

$$\underline{\int_a^b} f(x)\ dx = \overline{\int_a^b} f(x)\ dx,$$

then we say f is *Riemann integrable (on $[a,b]$)* and we write the common value of the upper and lower Riemann integral as

$$\int_a^b f(x)\ dx.$$

We refer to this common value as the *Riemann integral of f on $[a,b]$*.

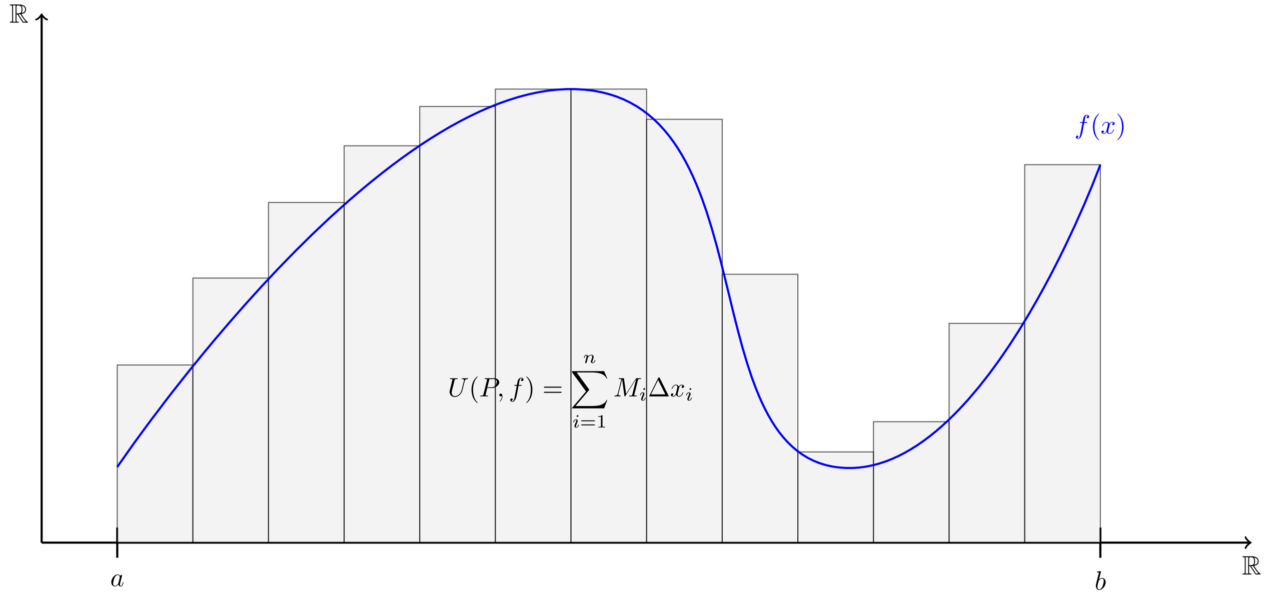

Figure 57: An upper Riemann sum.

Figure 58 shows a Riemann integrable function, and the value of the integral on $[a, b]$.

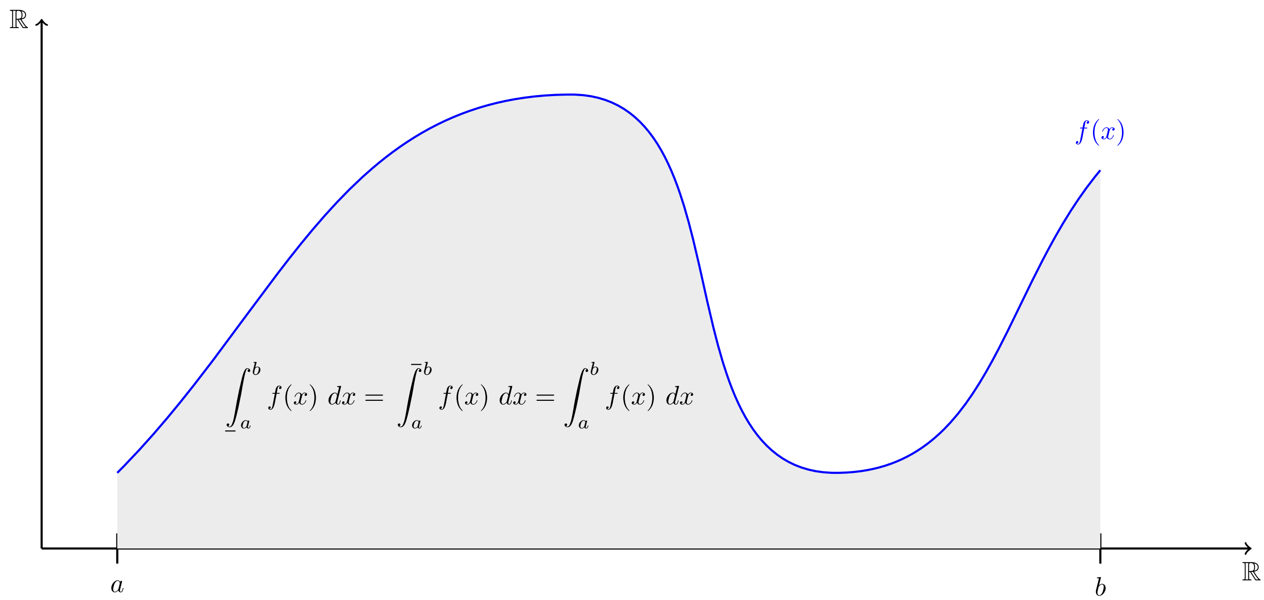

Figure 58: The function f is Riemann integrable on $[a, b]$.

Remark 6.1 (Bounded Functions on a Closed Interval)**.** It cannot be stressed enough that we are only able to define the Riemann integral for real functions that are bounded, and can only do so on a closed interval.

While Definition 6.5 may make sense on a theoretical level, it's not at all practical to use it to verify a function is actually integrable. How are we supposed to possibly find the supremum of all possible lower Riemann sums, and find the infimum of all possible upper Riemann sums?! A simple example will be presented, but it will be the only time we actually use Definition 6.5 to verify integrability. We will therefore need to find some alternate criterion for Riemann integrability.

Example 6.1. Let $f(x) = c$ on $[a, b]$ for some constant $c \in \mathbb{R}$. Suppose $P = \{0, x_1, \dots, x_{n-1}, 1\}$ is a partition of $[a, b]$. We have

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x) = c,$$
$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x) = c,$$

For $i = 1, \dots, n$. The lower and upper Riemann sums will agree, as $M_i = m_i$ for $i = 1, \dots, n$.

$$U(P, f) = L(P, f) = \sum_{i=1}^{n} i \cdot (x_k - x_{k-1}) = (x_n - x_{n-1}) + (x_{n-1} + x_{n-2}) + \dots (x_2 - x_1) + (x_1 - x_0) = x_n - x_0 = b - a.$$

We let P be arbitrary, so these sums will always be 1.

$$\underline{\int_a^b} f(x)\ dx = \sup_{P \in \mathbf{P}([a,b])} L(P, f) = \sup\{c\} = c$$
$$\overline{\int_a^b} f(x)\ dx = \inf_{P \in \mathbf{P}([a,b])} U(P, f) = \inf\{c\} = c.$$

These values agree, so f is Riemann integrable, and

$$\int_a^b c\ dx = c(b - a).$$

6.3 An Alternative Interpretation: Simple Functions (Very Optional)

Before we explore the properties of Riemann integration and develop a way to verify we can integrate a function, we can give an alternate formulation of the upper and lower Riemann integral. Not only is this how Tao (2016a) opts to introduce the Riemann integral, but it is how we will go about defining a superior form of integration in Section 13. The general idea is that we define the Riemann integral for for very "simple" functions, and then relate the integration of these functions to bounded real functions on $[a, b]$. **Disclaimer**: this introduction to simple functions is slightly informal for reasons that will be acknowledged afterwards.

We begin by defining a special type of step function.

Definition 6.6. Let X be a subset of $\mathbb{R}$, and $E \subset X$. We define the *characteristic function on E* as $\chi_E : X \to \{0, 1\}$, where

$$\chi_E(x) = \begin{cases} 1 \text{ if } x \in E \\ 0 \text{ if } x \notin E \end{cases}$$

This function is sometimes called the indicator function, as it indicates whether or not an element of X is in the subset E. We will be interested in characteristic functions in $\mathbb{R}$, and an example of such a function is seen Figure 59. We can use characteristic functions to write *any* step function. Step functions written this way will be the "simple" functions which we are interested in.

Definition 6.7. Let X be a subset of $\mathbb{R}$. A *simple function* $\varphi : X \to \mathbb{R}$ is a function which can be written as a *finite* linear combination of characteristic functions on $E \subset X$. That is there exists a set of finite scalers $\{c_1, \dots, c_n\} \subset \mathbb{R}$ and finite sets $E_1, \dots E_n \subset X$ such that

$$\varphi(x) = \sum_{k=1}^{n} c_k \chi_{E_k}(x).$$

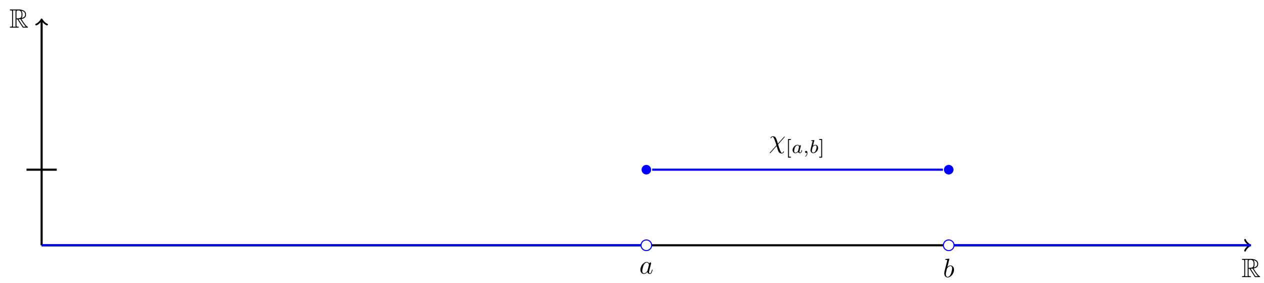

Figure 59: The characteristic function $\chi_{[a,b]} : \mathbb{R} \to \{0, 1\}$.

A simple functions domain will always be X, even if the sets on which the characteristic functions are defined do not cover X, i.e $X \not\subset \bigcup_{k=1}^{n} E_k$. We simply have $\varphi(x) = 0$ on the set $X \backslash \bigcup_{k=1}^{n} E_k$. The codomain of a simple function will *always* be $\mathbb{R}$, as the output of a simple function is determined by the set of real scalers $\{c_1, \ldots, c_n\}$. Where the function attains each value c_k in the range is determined by the set E_k. A concrete example where $X \subset \mathbb{R}$ may make things clearer.

Example 6.2. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be the simple function

$$\varphi(x) = 2\chi_{[1,3.5)} + 5\chi_{[3.5,4)} - 2\chi_{[4,7)} + 4\chi_{[7,8)} - \chi_{[8,10)} + \chi_{[10,14)} + 2\chi_{[14,15]}$$

on $\mathbb{R}$, where each χ_E is defined on a subset $E \subset [1, 15]$. This function is shown in Figure 60. In this case, $E_1 \cup \ldots \cup E_8 \neq \mathbb{R}$, but $\varphi(x)$ is still defined on all of $\mathbb{R}$. The function takes on the value 0 outside of $E_1 \cup \ldots \cup E_7$, but is still defined there. We could also define $\varphi : [1, 15] \to \mathbb{R}$ in a similar fashion. The only difference would be that φ would not be defined as zero on $(-\infty, 1) \cup (15, \infty)$.

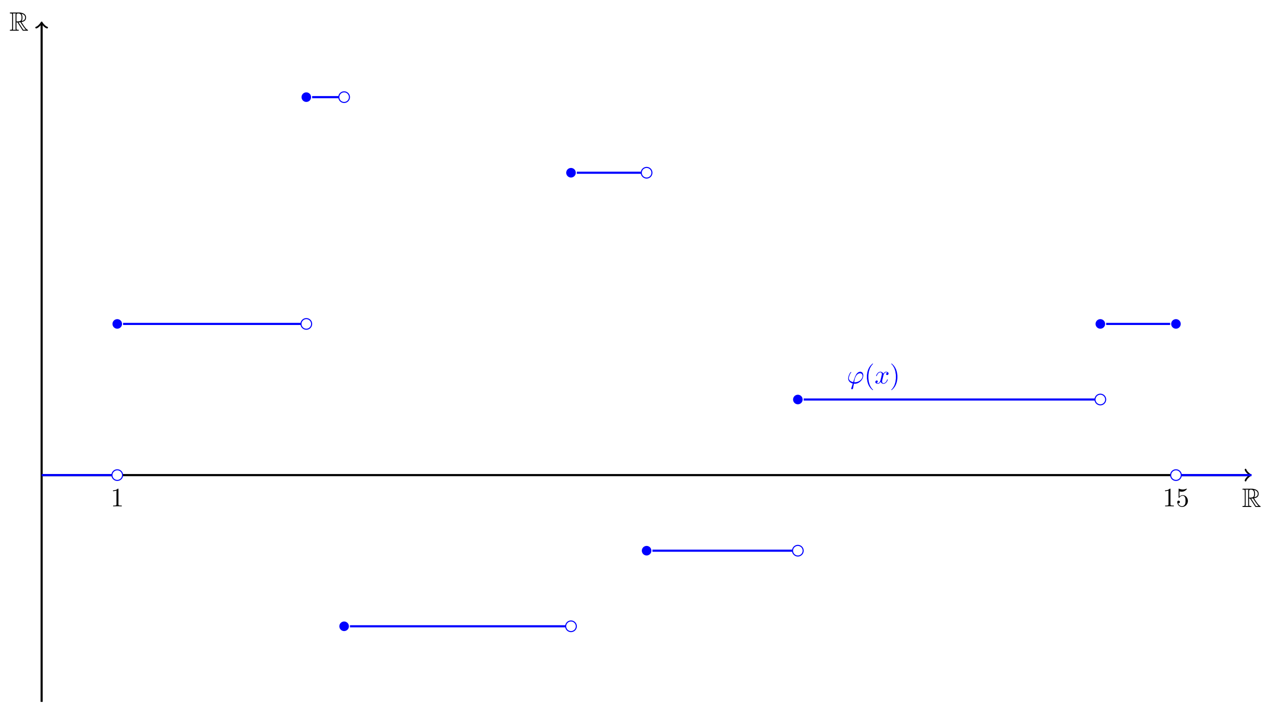

Figure 60: A simple function $\varphi : \mathbb{R} \to \mathbb{R}$.

In this particular case the set $\{E_k\}$ are piecewise disjoint. This need not be the case. Just know that if the intervals are not piecewise disjoint, we can always rewrite $\varphi(x)$ so they are. We will prove this later on when we treat simple functions with a bit more care.

Simple functions get their name, as in the context of integration they are simple. Because a simple function is written as a finite linear combination, it is bounded on the entirety of its domain.[90] If φ is only defined on some interval $[a, b]$, then every set $E_k \subset [a, b]$ is bounded for all k. These facts allow us to appeal directly to geometry to calculate the Riemann integral of a simple function.

Definition 6.8. Suppose $\varphi : [a, b] \to \mathbb{R}$ is a simple function which can be expressed as $\sum_{k=1}^{n} c_k \chi_{E_k}(x)$ for $E_k = (a_k, b_k) \subset [a, b]$ and $c_k \in \mathbb{R}$. Then we define the *Riemann integral of a simple function (on $[a, b]$)* as

$$\int_a^b \varphi(x)\ dx = \sum_{k=1}^{n} c_k \cdot (b_k - a_k).$$

For each "step" of the simple function, we multiply the width of the interval, $(b_k - a_k)$, by the height the function achieves on that interval, c_k. Adding the area of these rectangles up gives the integral. In Definition 6.8, E_k is open, but it really does not matter. It could be closed, or half-open. We will freely interchange them in this section, as it won't make a difference. The length of the interval would still be $b_k - a_k$.

Example 6.3. Define $\varphi : [1, 15] \to \mathbb{R}$ as

$$\varphi(x) = 2\chi_{[1,3.5)} + 5\chi_{[3.5,4)} - 2\chi_{[4,7)} + 4\chi_{[7,8)} - \chi_{[8,10)} + \chi_{[10,14)} + 2\chi_{[14,15]}.$$

Definition 6.8 gives

$$\begin{aligned}
\int_1^{15} \varphi(x)\ dx &= \sum_{k=1}^{7} c_k \cdot (b_k - a_k) \\
&= 2(3.5 - 1) + 5(4 - 3.5) - 2(7 - 4) + 4(8 - 7) - (10 - 8) + (14 - 10) + 2(15 - 14) \\
&= 9.5.
\end{aligned}$$

This integral can be seen in Figure 61.

Example 6.4. Let $\varphi : [a, b] \to \mathbb{R}$ be the trivial simple function of $\varphi(x) = \chi_{[a,b]}$. We have

$$\int_a^b \varphi(x)\ dx = b - a.$$

Now we that we can integrate simple functions, how do we extend this to any bounded real function on $[a, b]$?

Definition 6.9. Let X be a set, Y be an ordered set, $f : X \to Y$, and $g : X \to Y$. We say *f is greater than or equal to g (on X)* if $f(x) \geq g(x)$ for all $x \in X$, and write $f \geq g$. Similarly, we say *f is less than or equal to g (on X)* if $f(x) \leq g(x)$ for all $x \in X$, and write $f \leq g$.

Like many of the definitions that have been introduced, it will be important to specify a domain when we say a function is greater than or less than another. It should always be specified if it is unclear.

Let there be some real valued function f that is bounded on the interval $[a, b]$. Suppose for two simple functions defined on $[a, b]$, call them φ and ψ, we have $\varphi \leq f \leq \psi$ on $[a, b]$, as seen in Figure 62. Can we somehow use the well defined integrals of φ and ψ to approximate that of f? If you compare Figure 62 to Figure 56 and Figure 57, then things become clearer. The integrals of φ and ψ on $[a, b]$ coincide with the lower and upper Riemann sums shown in Figure 56 and Figure 57! In fact, we can think of a Riemann sum as the integral of a simple function. For a partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$ we may write

$$\varphi(x) = \sum_{i=1}^{n-1} m_i \chi_{[x_{i-1},x_i)} + m_n \chi_{[x_{n-1},x_n]},$$

$$\psi(x) = \sum_{i=1}^{n-1} M_i \chi_{[x_{i-1},x_i)} + M_n \chi_{[x_{n-1},x_n]},$$

[90] The function is bounded by $\max\{|c_1|, \ldots |c_n|\}$.

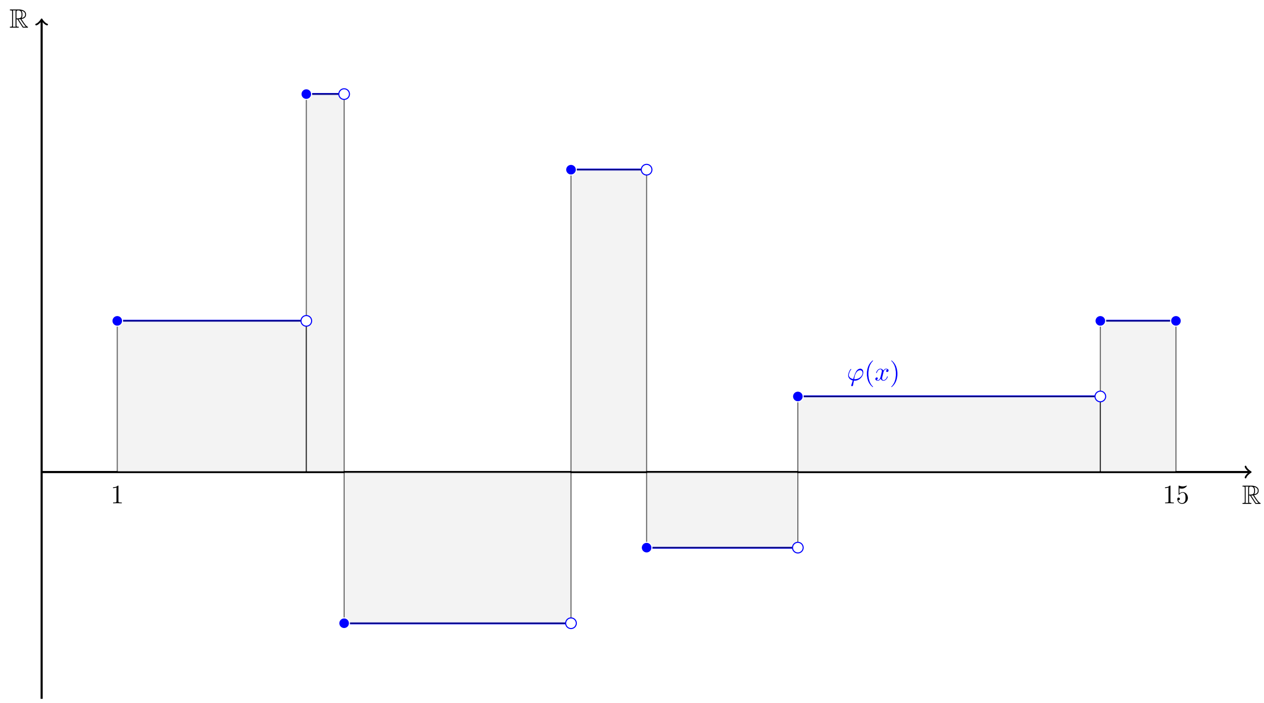

Figure 61: The integral of a simple function $\varphi : [1, 15] \to \mathbb{R}$

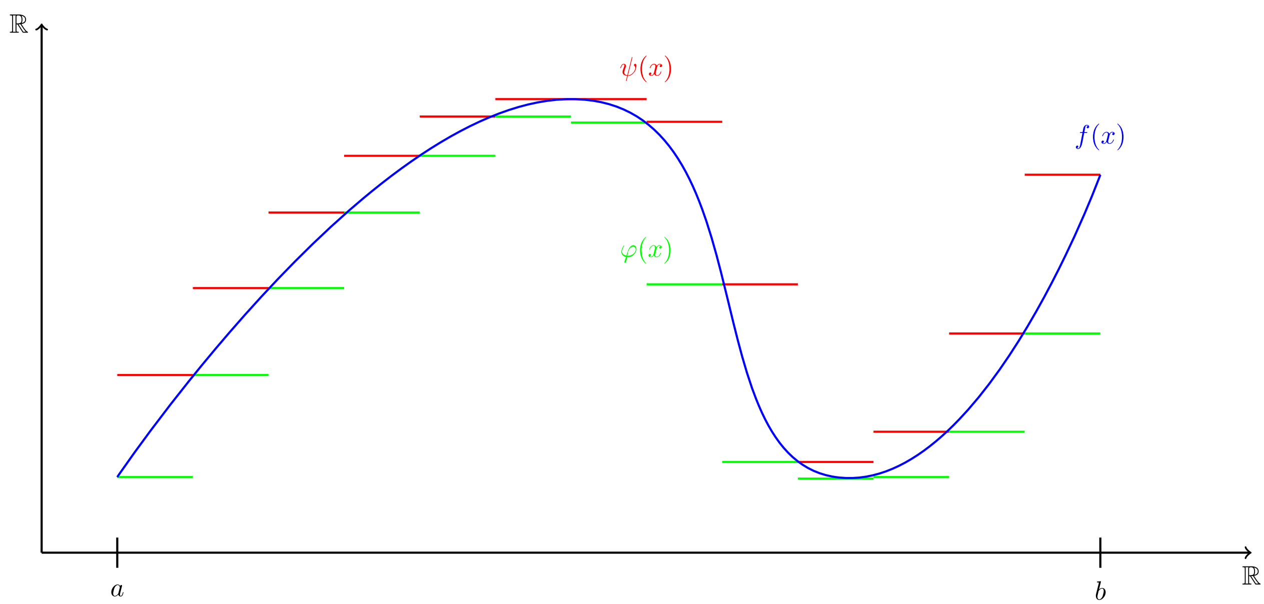

Figure 62: The function f is real and bounded on the interval $[a, b]$. We have $\varphi \leq f \leq \psi$ on $[a, b]$ for two simple functions defined on that interval.

which gives

$$\int_a^b \varphi(x)\ dx = \sum_{i=1}^{n} m_i \Delta x_i,$$

$$\int_a^b \psi(x)\ dx = \sum_{i=1}^{n} M_i \Delta x_i.$$

Thus we have established the following set inclusions:

$$\{L(P,f)\}_{P\in\mathbf{P}([a,b])} \subset \left\{ \int_a^b \varphi(x)\ dx \mid \varphi \le f \text{ on } [a,b],\ \varphi \text{ simple} \right\}, \tag{11}$$

$$\{U(P,f)\}_{P\in\mathbf{P}([a,b])} \subset \left\{ \int_a^b \varphi(x)\ dx \mid \varphi \ge f \text{ on } [a,b],\ \varphi \text{ simple} \right\}. \tag{12}$$

I'm now going to make what may seem like a bold claim, but is certainly corroborated by Figure 62 and our geometric understanding of the area under a curve:

$$\underline{\int_a^b} f(x)\ dx = \sup\{L(P,f)\}_{P\in\mathbf{P}([a,b])} = \sup\left\{ \int_a^b \varphi(x)\ dx \mid \varphi \le f \text{ on } [a,b],\ \varphi \text{ simple} \right\},$$

$$\overline{\int_a^b} f(x)\ dx = \inf\{U(P,f)\}_{P\in\mathbf{P}([a,b])} = \inf\left\{ \int_a^b \varphi(x)\ dx \mid \varphi \ge f \text{ on } [a,b],\ \varphi \text{ simple} \right\}.$$

Let's provide an informal proof sketch as to why this works for the lower Riemann integral.[91]

We know that $\sup A \le \sup B$, for $A \subset B$, so (11) gives us the first half of the result. We just need to show that

$$\sup\{L(P,f)\}_{P\in\mathbf{P}([a,b])} \ge \sup\left\{ \int_a^b \varphi(x)\ dx \mid \varphi \le f \text{ on } [a,b],\ \varphi \text{ simple} \right\}.$$

We can show that any integral of such a simple function is less than or equal to a lower Riemann sum, implying that the supremum of the set must take the form of a lower Riemann sum. Suppose $\varphi \le f$ on $[a,b]$ and is writen as

$$\varphi(x) = \sum_{i=k}^{n} c_k \chi_{E_k} = \sum_{k=1}^{n} c_k \chi_{[a_k,b_k)}$$

for $c_k \in \mathbb{R}$ and $a_k, b_k \in [a,b]$. Assume that $a_{k+1} = b_k$, so $\cup_{k=1}^n [a_k,b_k) = [a,b]$. We have $\varphi \le f$, so $c_k \le \inf_{x\in[a_k,b_k)} f(x)$. Therefore

$$\int_a^b \varphi(x)\ dx = \sum_{k=1}^{n} c_k(b_k - a_k) \le \sum_{k=1}^{n} \inf_{x\in[a_k,b_k)} f(x)(b_k - a_k).$$

But this final sum is just the integral of a simple function corresponding to a lower Riemann sum. Therefore for any integral of a simple function $\varphi \le f$, there exists lower Riemann sum that is greater than or equal to it.

This all justifies the following alternate definition of the upper and lower Riemann integrals:

$$\underline{\int_a^b} f(x)\ dx = \sup\left\{ \int_a^b \varphi(x)\ dx \mid \varphi \le f \text{ on } [a,b],\ \varphi \text{ simple} \right\},$$

$$\overline{\int_a^b} f(x)\ dx = \inf\left\{ \int_a^b \varphi(x)\ dx \mid \varphi \ge f \text{ on } [a,b],\ \varphi \text{ simple} \right\}.$$

If this makes no sense, that is fine. For the rest of this section, we will use the definitions presented in Subsection 6.1. In fact, we will never formally use any proofs involving Riemann integrals. We will however use simple functions in certain examples. This was only presented in an effort to make Section 13 easier. When we return to the idea of simple functions then, it will actually be even clearer than this case, as Section 12 will develop a formal concept of measuring an interval.

Remark 6.2 (About That Disclaimer)**.** Before the introduction of simple functions, I warned that it would not be totally formal. I made a very strong assumption that the length of an interval $[a,b]$ is $b-a$, and I made many assumptions about the properties of this length, and the length of the analogous open interval (a,b). Once again, we haven't defined any notion of length on $\mathbb{R}$, so this was careless. Tao (2016a) does a very good job of introducing the basic idea of length on $\mathbb{R}$ when discussing partitions and simple functions.

[91] The case for the upper Riemann integral is similar.

6.4 Verifying Riemann Integrability

Okay, that detour is now over. We return to the more pressing problem of us not having any easy way of verifying a bounded real function on $[a, b]$ is Riemann integrable. After intorudcing a series of lemmas, we will arrive at a familiar criterion that we will use to verify Riemann integrability.

Lemma 6.1. Let P be a partition of the interval $[a, b]$. If P^* is a refinement of P, then

$$L(P, f) \leq L(P^*, f)$$
$$U(P^*, f) \leq U(P, f)$$

Proof. We will show the result for the first inequality, as the proof for the second is analogous. Suppose that P^* contains exactly one more point than P^*, call it x^*. There are two consecutive points of P, x_{i-1} and x_i, such that $x_{i-1} < x^* < x_i$. Define

$$w_1 = \inf_{x \in [x_{i-1}, x^*]} f(x),$$
$$w_2 = \inf_{x \in [x^*, x_i]} f(x).$$

For $m_i = \inf_{x \in [x_{i-1}, x_i]} f(x)$, we have $w_1 \geq m_i$ and $w_2 \geq m_i$.[92] We therefore have

$$\begin{aligned} L(P^*, f) - L(P, f) &= w_1(x^* - x_{i-1}) + w_2(x_i - x^*) - m_i[x_i - x_{i-1}] \\ &= (w_1 - m_1)(x^* - x_{i-1}) + (w_2 - m_2)(x_i - x^*) \end{aligned}$$

This is the desired result if P^* only has one additional point.[93] If instead P^* contains k more points that P, we simply repeat this process k times to arrive at our result. □

Example 6.5. Let f be the bounded real function on $[0, 1]$ defined as $f(x) = x^2$. Suppose $P = 0, 1$, and $P^* = 0, 1/2, 1$. We have

$$L(P, f) = \inf_{x \in [0,1]} x^2 = 0$$
$$L(P^*, f) = \inf_{x \in [0,1/2]} x^2 + \inf_{x \in [1/2,1]} x^2 = 0 + \frac{1}{4} = \frac{1}{4}$$

This agrees with Lemma 6.1 (as it should).

Our next lemma provides an inequality we would hope holds for the upper and lower Riemann integral.

Lemma 6.2. Let f be a bounded real function on $[a, b]$. We have

$$\underline{\int_a^b} f(x)\ dx \leq \overline{\int_a^b} f(x)\ dx.$$

Proof. Let P_1 and P_2 be partitions of $[a, b]$, and P^* be a common refinement of P_1 and P_2. By Lemma 6.1,

$$L(P_1, f) \leq L(P^*, f) \leq U(P^*, f) \leq U(P_2, f).$$

If we fix P_1 and P_2, we can take the supremum and infimum over all of $\mathbf{P}([a, b])$.

$$\sup_{P_1 \in \mathbf{P}([a,b])} L(P_1, f) \leq \inf_{P_2 \in \mathbf{P}([a,b])} U(P_2, f)$$
$$\underline{\int_a^b} f(x)\ dx \leq \overline{\int_a^b} f(x)\ dx$$

□

[92] The bound m_i is the original infimum on the interval $[x_{i-1}, x_i]$. When we add x^* to our partition, this interval gets split into two intervals: $[x_{i-1}, x^*]$, and $[x^*, x_i]$. We're saying that the infima of these two new subintervals are weakly greater than that of the original interval. Drawing a picture of this can be helpful!

[93] How did we end up with such a simple expression for $L(P^*, f) - L(P, f)$? Well the only difference between P^* and P is that P^* has the intervals $[x_{i-1}, x^*]$, and $[x^*, x_i]$ *instead of* $[x_{i-1}, x_i]$. All the other intervals are the same and cancel out. We're left with the two additional intervals, and need to subtract the interval which they replaced.

When does Lemma 6.2 hold with equality? By Definition 6.5, this occurs precisely when f is Riemann integrable, because this is how we defined Riemann integrability. This would imply that we have

$$\underline{\int_a^b} f(x)\ dx < \overline{\int_a^b} f(x)\ dx$$

if and only if f is *not* Riemann integrable. At the end of this section, we will see some examples of this.

We are able to present the first of several results that will allow us to determine if a function is Riemann integrable. This theorem not only gives us sufficient conditions for integration, but it also gives a necessary condition for integrability. This gives it much more bite.

Theorem 6.1 (Riemann's Criterion)**.** Suppose f is a bounded real function on $[a, b]$. The function f is Riemann integrable *if and only if* for all $\varepsilon > 0$ there exists a partition such that

$$U(P, f) - L(P, f) < \varepsilon.$$

The proof of this crucial theorem amounts to just shuffling around inequalities. It only looks so long due to liberal typesetting.

Proof.

($\Longrightarrow$) Suppose f is Riemann integrable, and let $\varepsilon > 0$. By the completeness of the real numbers and properties of the supremum, there exists a partition P_1 such that

$$\begin{aligned}
U(P_1, f) - \sup_{P \in \mathbf{P}([a,b])} U(P, f) &< \frac{\varepsilon}{2}, \\
U(P_1, f) - \overline{\int_a^b} f(x)\ dx &< \frac{\varepsilon}{2}, \\
U(P_1, f) - \int_a^b f(x)\ dx &< \frac{\varepsilon}{2}.
\end{aligned}$$

Similarly, there exists a partition P_2 such that

$$\int_a^b f(x)\ dx - L(P_2, f) < \frac{\varepsilon}{2}.$$

If we choose P^* to be the common refinement of P_1 and P_2,[94] then by Lemma 6.1,

$$\begin{aligned}
U(P^*, f) - \int_a^b f(x)\ dx < U(P_1, f) - \int_a^b f(x)\ dx < \frac{\varepsilon}{2}, \\
\int_a^b f(x)\ dx - L(P^*, f) < \int_a^b f(x)\ dx - L(P_2, f) < \frac{\varepsilon}{2}.
\end{aligned}$$

Combining these two inequalities for P^* yields

$$\begin{aligned}
U(P^*, f) - L(P^*, f) &= U(P^*, f) - L(P^*, f) + 0 \\
&= U(P^*, f) - L(P^*, f) + \left[-\int_a^b f(x) + \int_a^b f(x)\ dx\right] \\
&= \left[U(P^*, f) - \int_a^b f(x)\right] + \left[\int_a^b f(x)\ dx - L(P^*, f)\right] \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

[94]Remember all the proofs with convergent sequences where we had two values N_1 and N_2, and took $N = \max\{N_1, N_2\}$ (see the proof of Proposition 3.2 for the first example of this)? This is essentially what we're doing by taking the common refinement of P_1 and P_2. We know that the all the ε inequalities will hold simultaneously by Lemma 6.1, allowing us to combine them.

($\Longleftarrow$) Suppose for all $\varepsilon > 0$ there exists a partition such that $U(P, f) - L(P, f) < \varepsilon$. By the definition of supremum and infimum we have

$$L(P, f) \leq \sup_{P \in \mathbf{P}([a,b])} U(P, f) = \underline{\int_a^b} f(x)\ dx,$$

$$\overline{\int_a^b} f(x)\ dx \leq \inf_{P \in \mathbf{P}([a,b])} U(P, f) \leq U(P, f),$$

for all $P \in \mathbf{P}([a, b])$. Combining this inequalities with Lemma 6.2 gives

$$L(P, f) \leq \underline{\int_a^b} f(x)\ dx \leq \overline{\int_a^b} f(x)\ dx \leq U(P, f).$$

This implies that

$$0 \leq \underline{\int_a^b} f(x)\ dx - \overline{\int_a^b} f(x)\ dx \leq \varepsilon,$$

but if this holds for all $\varepsilon > 0$,

$$\underline{\int_a^b} f(x)\ dx = \overline{\int_a^b} f(x)\ dx.$$

Therefore f is Riemann integrable.

$\square$

Riemann's Criterion should not catch us off guard. If f is integrable then the lower and upper Riemann integrals are equal. By the definition of those values, we need that the upper and lower Riemann sums become arbitrarily close to each other. We've seen this type of behavior with sequences, limits, and continuity. This time instead of some n or δ term which correspond to ε, we have a partition P that corresponds to ε. As we take ε to be smaller and smaller, we will need to find finer and finer partitions P, but we will always be able to do this in order to satisfy $U(P, f) - L(P, f) < \varepsilon$.

Remark 6.3 (Riemann Criterion and Cauchy Criterion)**.** The Riemann Criterion may have nothing to do with sequences, but it shares one important similarity with the Cauchy Criterion. Recall that the Cauchy Criterion is so useful because it does not require us to know the limit of a sequence. The same is true with the Riemann Criterion. We don't need to know the integral of f on $[a, b]$ to prove it exists!

Example 6.6. Let $f : [0, 1] \to \mathbb{R}$ be

$$f(x) = \begin{cases} 1/n \text{ if } x = 1/n \\ 0 \text{ otherwise} \end{cases}.$$

We can use the Riemann Criterion to show this function is integrable on $[0, 1]$. Let $\varepsilon > 0$. The sequence $\{1/n\}$ converges to 0, so for all $\varepsilon/2 > 0$ there exists an N such that $1/n \in [0, \varepsilon/2]$ for all $n \geq N$. Only a finite number of values of the form $1/n$ are in the interval $[\varepsilon/2, 1]$ (Proposition 3.1). We can cover these finite values by intervals $[x_1, x_2], \ldots, [x_{m-1}, x_m]$ such that $x_i \in [\varepsilon/2, 1]$ for all $i = 1, \ldots, m$. The finite length of these m intervals is less than $\varepsilon/2$. Let $P = \{0, \varepsilon, x_1, \ldots, x_m\}$. We have

$$U(P, f) - L(P, f) = (\varepsilon/2 - 0) + \underbrace{(x_2 - x_1) + \ldots + (x_m - x_{m-1})}_{<\varepsilon/2} < \varepsilon.$$

The function is Riemann integrable.[95]

The Riemann Criterion gives way to several related results that put integrability in conversation with limitings processes.

[95] If you keep going with this example, you can verify that $\int_0^1 f(x)\ dx = 0$.

Proposition 6.1. Suppose f is a bounded real function on the interval $[a, b]$.

1. If $U(P, f) - L(P, f) < \varepsilon$ for some P, then it holds for every refinement of P.
2. If $U(P, f) - L(P, f) < \varepsilon$ for some $P = \{x_0, \ldots, x_n\}$, and if $s_i, t_i \in [x_{i-1}, x_i]$, then
$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta x_i < \varepsilon.$$
3. If f is Riemann integrable, $U(P, f) - L(P, f) < \varepsilon$ for some $P = \{x_0, \ldots, x_n\}$, and $t_i \in [x_{i-1} - x_i]$, then
$$\left| \sum_{i=1}^{n} f(t_i) \Delta x_i - \int_a^b f(x)\ dx \right| < \varepsilon.$$

Proof.

1. Let P^* be a refinement of P. By Lemma 6.1,
$$U(P^*, f) - L(P^*, f) < U(P, f) - L(P, f) < \varepsilon.$$
2. We have $f(s_i), f(t_i) \in [m_i, M_i]$, as f is bounded on any interval by the supremum and infimum on that interval. This means that
$$|f(s_i) - f(t_i)| \leq M_i - m_i.$$
Therefore
$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta x_i \leq \sum_{i=1}^{n} (M_i - m_i) \Delta x_i = \sum_{i=1}^{n} M_i \Delta x_i - \sum_{i=1}^{n} m_i \Delta x_i =< U(P, f) - L(P, f) < \varepsilon.$$
3. We will always have the two following inequalities:
$$L(P, f) \leq \sum_{i=1}^{n} f(t_i) \Delta x_i \leq U(P, f),$$
$$L(P, f) \leq \int_a^b f(x)\ dx \leq U(P, f).$$
These combine to give
$$\left| \sum_{i=1}^{n} f(t_i) \Delta x_i - \int_a^b f(x)\ dx \right| \leq U(P, f) - L(P, f) < \varepsilon.$$

□

This proposition seems like a random string of inequalities,[96] but we will use them to prove many important results.

6.5 Properties of Riemann Integration

Now we can present some familiar properties of the integral. Just like when we did this with limits and derivatives, most of these may be familiar, so examples won't be furnished for every single possible property.

Theorem 6.2. Suppose f and g are both bounded real functions on $[a, b]$ which are integrable. Then

[96] To be honest, most of real analysis seems like a random string of inequalities sometimes.

1. $f+g$ is integrable, and
$$\int_a^b f(x)+g(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx;$$
2. cf is integrable for $c \in \mathbb{R}$, and
$$\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx;$$
3. (Monotonicity) if $f \leq g$ on $[a,b]$ we have
$$\int_a^b f(x)\ dx \leq \int_a^b g(x)\ dx.$$
4. (Additivity) for $c \in [a,b]$, f is integrable on $[a,c]$ and $[c,b]$, and
$$\int_a^c f(x)\ dx + \int_c^b f(x)\ dx = \int_a^b f(x)\ dx$$
5. if $|f(x)| \leq M$ on $[a,b]$ we have
$$\left|\int_a^b f(x)\ dx\right| \leq M(b-a)$$

The proofs of these are all really similar and monotonous, so fair warning.[97]

Proof.

1. Let P be a partition of $[a,b]$. We have
$$L(P,f)+L(P,g) \leq L(P,f+g) \leq U(P,f+g) \leq U(P,f)+U(P,g). \tag{13}$$
For $\varepsilon > 0$ there are partitions P_1 and P_2 such that
$$U(P_1,f)-L(P_1,f) < \frac{\varepsilon}{2},$$
$$U(P_2,g)-L(P_2,g) < \frac{\varepsilon}{2}.$$
If we take P^* to be the common refinement of P_1 and P_2, then
$$U(P^*,f)-L(P^*,f) < \frac{\varepsilon}{2},$$
$$U(P^*,g)-L(P^*,g) < \frac{\varepsilon}{2}.$$
These inequalities combined with (13) gives $U(P^*,f+g)-L(P^*,f+g) < \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon$, so $f+g$ is Riemann integrable. For this same P^*,
$$U(P^*,f) < \int_a^b f(x)\ dx + \frac{\varepsilon}{2},$$
$$U(P^*,g) < \int_a^b g(x)\ dx + \frac{\varepsilon}{2}.$$
Again by (13),
$$\int_a^b f(x)+g(x)\ dx \leq U(P^*,f+g) \leq U(P^*,f)+U(P^*,g) < \int_a^b f(x)\ dx + \int_a^b g(x)\ dx + \varepsilon.$$

[97]There is a reason that Rudin (1976) omits most of them.

This holds for all $\varepsilon > 0$, so

$$\int_a^b f(x) + g(x)\ dx \leq \int_a^b f(x)\ dx + \int_a^b g(x)\ dx.$$

Now for the reverse direction all at once:

$$\begin{aligned}
L(P^*, f) &> \int_a^b f(x)\ dx - \frac{\varepsilon}{2}, \\
L(P^*, g) &> \int_a^b g(x)\ dx - \frac{\varepsilon}{2}, \\
\int_a^b f(x) + g(x)\ dx \geq L(P^*, f+g) \geq L(P^*, f) + L(P^*, g) &> \int_a^b f(x)\ dx + \int_a^b g(x)\ dx - \varepsilon \\
\int_a^b f(x) + g(x)\ dx &\geq \int_a^b f(x)\ dx + \int_a^b g(x)\ dx.
\end{aligned}$$

Therefore we have

$$\int_a^b f(x) + g(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx.$$

2. First suppose $c \geq 0$. For any set $A \subset [a, b]$,

$$\begin{aligned}
\sup_{x \in A} cf(x) &= c \sup_{x \in A} f(x), \\
\inf_{x \in A} cf(x) &= c \inf_{x \in A} f(x).
\end{aligned}$$

This gives

$$\begin{aligned}
\overline{\int_a^b} f(x)\ dx &= \inf_{P \in \mathbf{P}([a,b])} U(P, cf) = c \inf_{P \in \mathbf{P}([a,b])} U(P, f) = c \overline{\int_a^b} f(x)\ dx, \\
\underline{\int_a^b} f(x)\ dx &= \sup_{P \in \mathbf{P}([a,b])} L(P, cf) = c \inf_{P \in \mathbf{P}([a,b])} L(P, f) = c \underline{\int_a^b} f(x)\ dx.
\end{aligned}$$

Since f is Riemann integrable, its upper and lower Riemann integrals are equal. This establishes the integrability of f, as

$$c \underline{\int_a^b} f(x)\ dx = c \overline{\int_a^b} f(x)\ dx = c \int_a^b f(x)\ dx.$$

Now let $c = -1$. In this case, we can't "factor" out a constant from a supremum and infimum. Instead, for any $A \subset [a, b]$, we have

$$\begin{aligned}
\sup_{x \in A} -f(x) &= -\inf_{x \in A} f(x), \\
\inf_{x \in A} -f(x) &= -\sup_{x \in A} f(x).
\end{aligned}$$

We will have $U(P, -f) = -L(P, f)$ and $L(P, -f) = -U(P, f)$ for any partition. This gives

$$\begin{aligned}
\overline{\int_a^b} -f(x)\ dx &= \inf_{P \in \mathbf{P}([a,b])} U(P, -f) = \inf_{P \in \mathbf{P}([a,b])} -L(P, f) = -\sup_{P \in \mathbf{P}([a,b])} L(P, f) = -\underline{\int_a^b} f(x)\ dx, \\
\underline{\int_a^b} -f(x)\ dx &= \sup_{P \in \mathbf{P}([a,b])} L(P, -f) = \sup_{P \in \mathbf{P}([a,b])} -U(P, f) = -\inf_{P \in \mathbf{P}([a,b])} U(P, f) = -\overline{\int_a^b} f(x)\ dx.
\end{aligned}$$

Since f is Riemann integrable, its negative upper and lower Riemann integrals are equal, so

$$\underline{\int_a^b} -f(x)\ dx = c \overline{\int_a^b} -f(x)\ dx = c \int_a^b -f(x)\ dx.$$

In general, if $c < 0$, we can write it as $-1 \cdot |c|$ and apply the first two cases.

3. By the two previous parts, we can show that the integral of $g-f$ is not negative, as

$$\int_a^b g(x)\ dx + \int_a^b -f(x)\ dx = \int_a^b g(x) - f(x)\ dx.$$

We know $g - f \geq 0$, so $\inf_{x\in A}(f(x) - g(x)) \geq 0$ for any $A \subset [a,b]$. This means that for any P, $L(P,f) \geq 0$, so

$$\int_a^b g(x) - f(x)\ dx \geq L(P,f) \geq 0.$$

4. For all $\varepsilon > 0$, there exists a partition P such that $U(P,f) - L(P,f) < \varepsilon$. Define $P^* = P \cup \{c\}$ to be the refinement of P which results from adding c.[98] Let $Q = P^* \cap [a,c]$ and $R = P^* \cap [c,b]$ be partitions of $[a,b]$ and $[c,b]$ respectively. For these partitions

$$\begin{aligned} U(P^*,f) &= U(Q,f) + U(R,f), \\ L(P^*,f) &= L(Q,f) + L(R,f). \end{aligned}$$

We can conclude

$$\begin{aligned} U(Q,f) - L(Q,f) &= U(P^*,f) - L(P^*,f) - [U(R,f) - L(R,f)] \leq U(P,f) - L(P,f) < \varepsilon, \\ U(R,f) - L(R,f) &= U(P^*,f) - L(P^*,f) - [U(Q,f) - L(Q,f)] \leq U(P,f) - L(P,f) < \varepsilon. \end{aligned}$$

This shows that f is integrable on $[a,c]$ and $[c,b]$ by Riemann's Criterion.

We have

$$\begin{aligned} \int_a^b f(x)\ dx &\leq U(P,f) = U(Q,f) + U(R,f) < L(Q,f) + L(R,f) + \varepsilon < \int_a^c f(x)\ dx + \int_c^b f(x)\ dx + \varepsilon, \\ \int_a^b f(x)\ dx &\geq U(P,f) = U(Q,f) + U(R,f) > U(Q,f) + U(R,f) - \varepsilon > \int_a^c f(x)\ dx + \int_c^b f(x)\ dx - \varepsilon, \end{aligned}$$

which combine to give

$$\int_a^c f(x)\ dx + \int_c^b f(x)\ dx - \varepsilon < \int_a^b f(x)\ dx < \int_a^c f(x)\ dx + \int_c^b f(x)\ dx + \varepsilon.$$

If this for all $\varepsilon > 0$, then

$$\int_a^b f(x)\ dx = \int_a^c f(x)\ dx + \int_c^b f(x)\ dx$$

5. Treat M as a constant function on $[a,b]$. Example 6.1 showed that

$$\int_a^b M\ dx = M(b-a).$$

If $|f(x)| \leq M$, then $-M \leq f(x) \leq M$. By monotonicity,

$$-M(b-a) = \int_a^b -M\ dx \leq \int_a^b f(x)\ dx \leq \int_a^b M\ dx = M(b-a),$$

which can be written as

$$\left| \int_a^b f(x)\ dx \right| \leq M(b-a).$$

□

[98] It could be the case that $c \in P$. This would just mean that $P^* = P$.

Example 6.7 (Linearity)**.** The first 2 parts of Theorem 6.2 give that Riemann integration is linear. For any $c, d \in \mathbb{R}$, we will have

$$\int_a^b cf(x) + dg(x)\ dx = c\int_a^b f(x) + d\int_a^b g(x)\ dx.$$

This will turn out to be *very very* important in later sections. We will not be working with the Riemann integral than, but we'll see that integration in general is intrinsically linked to linearity.

While *the* Mean Value Theorem deals with derivatives, a similar result holds for integrals. It asserts that a continuous function must take on its average value on an interval.

Proposition 6.2 (Mean Value Theorem for Integrals I)**.** Let f be a bounded real function on $[a, b]$. If f is continuous, then there exists a $c \in [a, b]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\ dx.$$

Proof. Since f is continuous, we can use the Extreme Value Theorem. The function f must attain a maximum M and minimum m on $[a, b]$. By monotonicity,

$$\begin{aligned}
\int_a^b f(m)\ dx \leq \int_a^b f(x)\ dx \leq \int_a^b f(M)\ dx,\\
f(m)(b-a) \leq \int_a^b f(x)\ dx \leq f(M)(b-a),\\
f(m) \leq \frac{1}{b-a}\int_a^b f(x)\ dx \leq f(M).
\end{aligned}$$

By the Intermediate Value Theorem, f takes on every value in $[m, M]$, so there must be some c such that

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\ dx.$$

□

The integral in Proposition 6.2 corresponds to the average value attained by the function on $[a, b]$. This proposition has a nice geometric interpretation if the equality is written as

$$f(c)(b-a) = \int_a^b f(d)\ dx.$$

As shown in Figure 63, we can find some rectangle of height $f(c)$ on the interval $[a, b]$ that equals the area under f. A second form of the Mean Value Theorem for Integrals is even more useful, as it allows us to find a value in $[a, b]$ that allows us to "factor" a function out of the integral of a product.

Proposition 6.3 (Mean Value Theorem for Integrals II)**.** Let f be a bounded real function on $[a, b]$. If f is continuous, and g is integrable function $[a, b]$ and does not change signs, then there exists a $c \in [a, b]$ such that

$$\int_a^b f(x)g(x)\ dx = f(c)\int_a^b g(x)\ dx.$$

Proof. If $g(x) = 0$ for all $x \in [a, b]$, then the result holds trivially. Assume instead that $g(x) > 0$. Since f is continuous, we can use the Extreme Value Theorem. The function f must attain a maximum M and minimum m on $[a, b]$.

$$\begin{aligned}
f(m) \leq& f(x) \leq f(M),\\
f(m)g(x) \leq& f(x)g(x) \leq f(M)g(x).
\end{aligned}$$

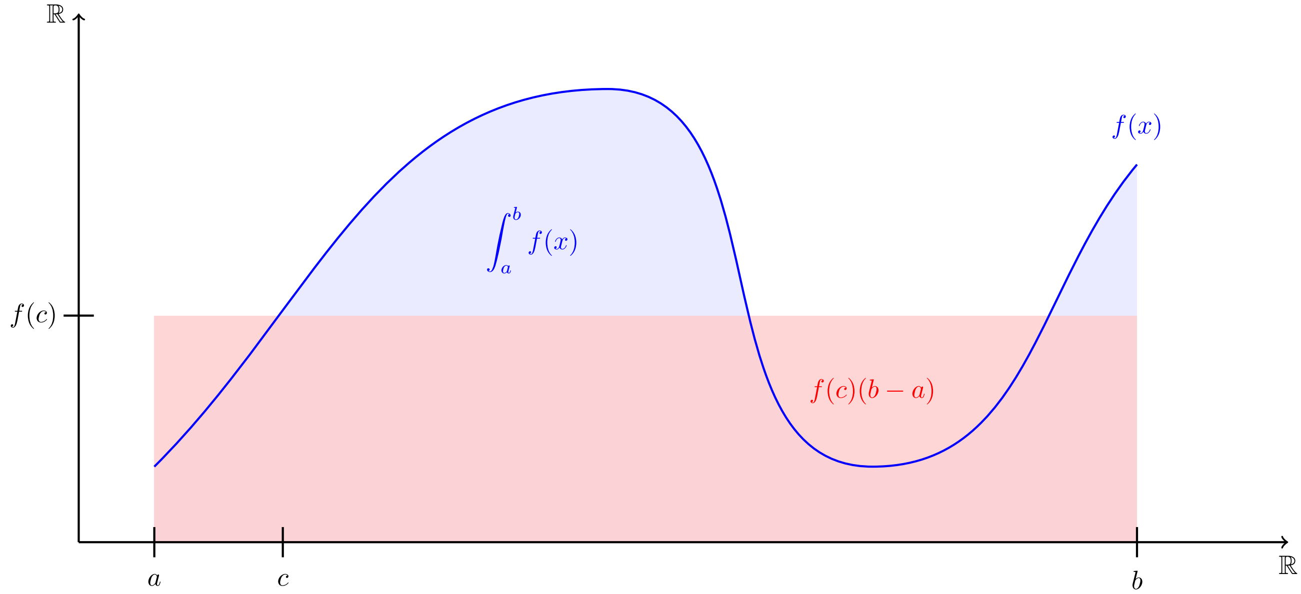

Figure 63: The Mean Value Theorem for Integrals says that we can find a value $c \in [a, b]$ such that the red rectangle is equal in area to the integral shown in blue.

Multiply by $g(x)$ will not change the inequality, as $g(x) > 0$ for all $x \in [a, b]$.[99] By monotonicity and the linearity of integration,

$$\begin{aligned} f(m)g(x) \leq & f(x)g(x) \leq f(M)g(x), \\ \int_a^b f(m)g(x)\ dx \leq & \int_a^b f(x)g(x)\ dx \leq \int_a^b f(M)g(x)\ dx, \\ f(m)\int_a^b g(x)\ dx \leq & \int_a^b f(x)g(x)\ dx \leq f(M)\int_a^b g(x)\ dx. \end{aligned}$$

If we divide by the integral of g, then

$$f(m) \leq \frac{\int_a^b f(x)g(x)\ dx}{\int_a^b g(x)\ dx} \leq f(M).$$

By the continuity of f and the Intermediate Value Theorem, there exists a $c \in \mathbb{R}$ such that

$$f(c) = \frac{\int_a^b f(x)g(x)\ dx}{\int_a^b g(x)\ dx},$$

which can be expressed as

$$\int_a^b f(x)g(x)\ dx = f(c)\int_a^b g(x)\ dx.$$

As mentioned in Footnote 99, the case where $g < 0$ is virtually the same. The only difference is that all the inequalities are reversed, but we're still able to apply the Intermediate Value Theorem. $\square$

[99] The other case where $g(x) < 0$ would reverse the order of the inequality, but that is fine. What's important is that we have an upper and lower bound on $f(x)g(x)$, not what those bounds happen to be.

Example 6.8. For Proposition 6.3, let $g(x) = 1$ on $[a, b]$. In this case we recover Proposition 6.2.

$$\int_a^b f(x)g(x)\ dx = f(c)\int_a^b g(x)\ dx,$$
$$\int_a^b f(x)\cdot 1\ dx = f(c)\int_a^b 1\ dx,$$
$$\int_a^b f(x)g(x)\ dx = f(c)(b-a).$$

This observation shows that Proposition 6.3 is a generalized version of its predecessor.

Example 6.9 (What if g Changes Sign?)**.** Proposition 6.3 stipulates that the function g cannot change sign on $[a, b]$. This is essential. If this is not the case, then the inequality

$$f(m)g(x) \le f(x)g(x) \le f(M)g(x)$$

will not hold for all $x \in [a, b]$. Suppose $g(x) = x$ on $[-1, 1]$, and $f(x) = x$. In this case

$$f(x)\int_{-1}^1 g(x)\ dx = 0$$

for all x, but $f(x)g(x) = x^2$, so

$$\int_{-1}^1 f(x)g(x)\ dx \neq 0.$$

Therefore Proposition 6.3 does not hold.

6.6 Riemann Integration and Continuity

With the Riemann Criterion in hand, we're able to prove that a function is integrable. We will now use it to show that two very general classes of functions are integrable, the first being continuous functions.

The fact that all continuous functions (on $[a, b]$) are integrable should not come as a surprise. Not only is nearly every function integrated in a calculus course continuous, but continuity and integrability are both the results of limiting behavior with an arbitrary $\varepsilon > 0$ (the latter being the case as a result of Riemann's Criterion).

Theorem 6.3 (Continuity implies Integrability)**.** If f is a real continuous function on $[a, b]$, then f is Riemann integrable on $[a, b]$.

Proof. First note that a continuous function on $[a, b]$ is in fact bounded on $[a, b]$ by the Extreme Value Theorem.[100] Furthermore, $[a, b]$ is compact (Heine-Borel), so f is uniformly continuous on $[a, b]$ (Theorem 4.8). For all $\varepsilon > 0$, there exists a δ such that

$$|f(x) - f(y)| < \frac{\varepsilon}{n(b-a)},$$

for all $x, y \in [a, b]$ which satisfy $|x - y| < \delta$. Choose a partition P of $[a, b]$ such that $|x_i - x_{i-1}| < \delta$ for $i = 1, \ldots, n$.[101]

Our function is continuous, so it achieves its maximum and minimum on each $[x_{i-1}, x_i] \subset P$, but in this case those are M_i and m_i. We therefore have

$$|f(s_i) - f(t_i)| < |M_i - m_i| < \frac{\varepsilon}{b-a}$$

[100] This is important to point out because we only can define the Riemann integral for bounded real functions.
[101] For example, you could take form P with n intervals of length $(b-a)/n$ for $n > (b-a)/\delta$

for $|s_i - t_i| < \delta$, where $f(s_i) = M_i$ and $f(t_i) = m_i$.[102] This allows us to verify that the Riemann Criterion holds for the partitions f:

$$U(P,f) - L(P,f) = \sum_{i=1}^{n} M_i \Delta x_i - \sum_{i=1}^{n} m_i \Delta x_i = \sum_{i=1}^{n} (M_i - m_i)\Delta x_i < n \cdot \frac{\varepsilon}{n(b-a)} \cdot (b-a) < \varepsilon.$$

□

Example 6.10 (Integrable But Not Continuous)**.** The converse of Theorem 6.3 *is not true*. Example 6.6 showed a function that failed to be continuous on $[0,1]$ but is still integrable on $[0,1]$.

Proposition 6.4 (Continuous Composition Preserves Integrability)**.** Suppose f is a bounded real function on $[a,b]$, $m \leq f \leq M$, ϕ is continuous on $[m,M]$, and $h(x) = \phi \circ f(x) = \phi(f(x))$ on $[a,b]$. Then h is integrable on $[a,b]$.

For the proof of this result, it's important to remember that the domain of ϕ is the range of f. It can be easy to get bogged down with the notation in this one. This also will be the most intense proof involving ε yet.

Proof. Let $\varepsilon > 0$. The function ϕ is continuous on the compact set $[a,b]$, so it is uniformly continuous. There exists a $\varepsilon > 0$ such that $\delta < \varepsilon$ and

$$|\phi(s) - \phi(t)| < \frac{\varepsilon}{(b-a) + 2\sup_{v \in [m,M]} |\phi(v)|}$$

whenever $|s - t| < \delta$ for $s,t \in [m,M]$.[103]

Since f is integrable, there is a partition $P = \{x_0, \ldots, x_n\}$ on $[a,b]$ for which

$$U(P,f) - L(P,f) < \delta^2 \tag{14}$$

by Riemann's Criterion. This partition gives a partition of $[m,M]$ in $P^* = \{m = f(x_0), f(x_1), \ldots, f(x_n) = M\}$.[104] Let

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x), \qquad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \qquad M_i^* = \sup_{t \in [f(x_{i-1}), f(x_i)]} \phi(t), \qquad m_i^* = \inf_{t \in [f(x_{i-1}), f(x_i)]} \phi(t).$$

Figure 64 shows what a partition M_i^* and m_i^* will look like. We can divide the indices for the partition P into two sets

$$A = \{i \mid M_i - m_i < \delta\},$$
$$B = \{i \mid M_i - m_i \geq \delta\}.$$

If $i \in A$, then by the uniform continuity of ϕ,

$$|\phi(M_i) - \phi(m_i)| = |M_i^* - m_i^*| < \varepsilon.$$

If $i \in B$ then

$$|M_i^* - m_i^*| \leq 2K,$$

[102]That is, s_i and t_i are where the function acheives its maximum M_i and minimum m_i respectively.

[103]Woah, what is up with us saying $\delta < \varepsilon$!? Well if the inequalities hold for δ, then it should hold for any number less than $\delta' = \min\{\delta, \varepsilon\}$. If we have $\delta' < \varepsilon$, then $\delta' = \delta$. Therefore we lost no generality in saying $\delta' < \varepsilon$. We simply introduce it as δ instead of explicitly going through this though process with δ'.

[104]This is the partition that we will use with $\phi : [m,M] \to \mathbb{R}$.

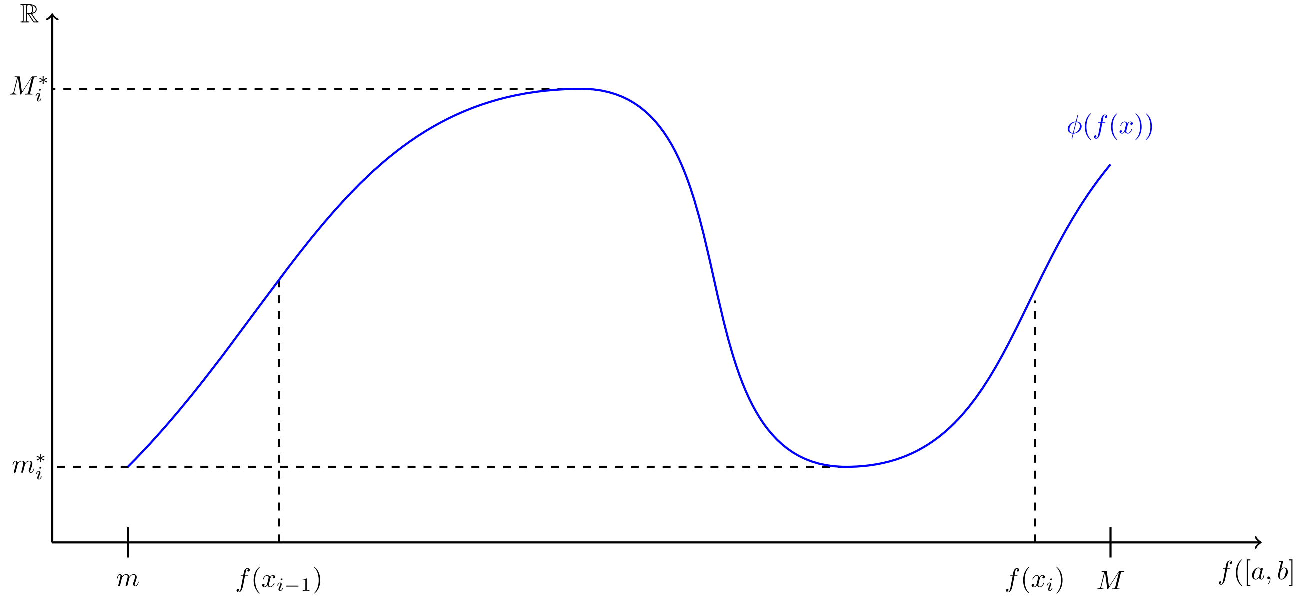

Figure 64: An interval $[f(x_{i-1}), f(x_i)] \subset P^*$, and the values M_i^* and m_i^*.

for $K = \sup |\phi(t)|$ and $m \leq t \leq M$.[105] By (14),

$$\begin{aligned}
& U(P, f) - L(P, f) = \sum_{i=1}^{n} M_i \Delta x_i - \sum_{i=1}^{n} m_i \Delta x_i = \sum_{i=1}^{n} (M_i - m_i) \Delta x_i < \delta^2 \\
\implies & \sum_{i \in A} (M_i - m_i) \Delta x_i + \sum_{i \in B} (M_i - m_i) \Delta x_i < \delta^2 \\
\implies & \sum_{i \in B} \underbrace{(M_i - m_i)}_{\geq \delta} \Delta x_i < \delta^2 \\
\implies & \delta \sum_{i \in B} \Delta x_i \leq \sum_{i \in B} (M_i - m_i) \Delta x_i < \delta^2 \\
\implies & \sum_{i \in B} \Delta x_i < \delta.
\end{aligned}$$

This final inequality allows us to show the Riemann condition holds for $\phi(f(x))$, as required:

$$\begin{aligned}
U(P, \phi(f)) - L(P, \phi(f)) &= \sum_{i \in A} \underbrace{(M_i^* - m_i^*)}_{< \varepsilon} \Delta x_i + \sum_{i \in B} \underbrace{(M_i^* - m_i^*)}_{2K} \underbrace{\Delta x_i}_{< \delta} \\
&< \frac{\varepsilon}{(b - a) + \sup |\phi(t)|} (b - a) + 2K \underbrace{\delta}_{< \frac{\varepsilon}{(b-a) + 2 \sup |\phi(t)|}} \\
&< \frac{\varepsilon}{(b - a) + 2K} [(b - a) + 2K] \\
&= \varepsilon.
\end{aligned}$$

$\square$

[105] In the context of Figure 64, this inequality makes a bit more sense. The distance between the sup and inf of $\phi(t)$ on the interval $[f(x_{i-1}), f(x_i)]$ must be less than twice the absolute value of the sup and of ϕ on the whole domain being partitioned $[m, M]$. If $m_i^* = -M_i^*$, then $|M_i^* - m_i^*| = 2M_i^*$. If M_i^* happens to be the supremum on all of $[m, M]$, that is $K = M_i^*$, then we have the equality $|M_i^* - m_i^*| = 2K$. Understanding this inequality may be the toughest part of this proof, and it's worth drawing some pictures if it is not clear.

Along with keeping track of if we're working in $[a, b]$ or $[m, M]$, I think another part of this proof that is not clear is how to handle the case where $M_i - m_i \geq \delta$. Splitting the partition of $[a, b]$ into two parts, one corresponding to the case where $M_i - m_i < \delta$ (indexed by A), the other where $M_i - m_i \geq \delta$ (indexed by B), is a pretty clear answer, but it goes against a lot on the instincts developed when doing analysis. Many proofs require you to show some inequality holds in every possible case, so if you were trying to develop this proof on your own, when you realize $M_i - m_i \geq \delta$ for some i, it may seem like you're doing something wrong.

Proposition 6.4 allows us to prove two additional properties of integration.

Proposition 6.5. Suppose f and g are both bounded real functions on $[a, b]$ which are integrable. Then fg is integrable on $[a, b]$.

Proof. By Theorem 6.2, $f \pm g$ is integrable. Proposition 6.4 tells us that $\phi(f \pm g)$ and will be integrable as long as ϕ is continuous. If we let $\phi(t) = t^2$, then

$$\phi(f - g) = (f - g)^2,$$
$$\phi(f + g) = (f + g)^2,$$

are integrable. Again by Theorem 6.2, their difference scaled should be integrable.

$$fg = \frac{(f + g)^2 + (f - g)^2}{4}$$

This gives that fg is integrable. $\square$

The converse of Proposition 6.5 is not true. Being able to factor an integrable function does not guarantee the factors are integrable as the next example shows.

Example 6.11. Define $f : \mathbb{R} \to \mathbb{R}$ as

$$f(x) = \begin{cases} 1 \text{ if } x \in \mathbb{Q} \\ -1 \text{ if } x \notin \mathbb{Q} \end{cases}.$$

This function is not integrable on $[0, 1]$ (see Subsection 6.11), but f^2 is, as $f^2(x) = 1$ for all $x \in \mathbb{R}$.

Proposition 6.6. Suppose f is bounded real functions on $[a, b]$ which is integrable. Then $|f|$ is integrable on $[a, b]$, and

$$\left| \int_a^b f(x)\ dx \right| \leq \int_a^b |f|\ dx.$$

Proof. Proposition 6.4 tells us that $\phi(f)$ and will be integrable as long as ϕ is continuous, so let $\phi(t) = |t|$. This shows that $\phi(f) = |f|$ is integrable. Chose $c \in \{-1, 1\}$ such that

$$c \int_a^b f(x)\ dx \geq 0.$$

We have $cf \leq |f|$, so by the monotonicity and linearty of integration,

$$\left| \int_a^b f(x)\ dx \right| = c \int_a^b\ f\ dx = \int_a^b\ cf\ dx \leq \int_a^b |f|\ dx.$$

$\square$

Example 6.12 (Which Direction is the Inequality?)**.** If you need to cram for an exam, you're not too concerned with having a deep understanding of propositions and theorems, so remembering the direction of the inequality of Proposition 6.6 may come down to memorization.[106] In this case just remember the simple examples of $f(x) = x$ on the interval $[-1, 1]$. In this case we have

$$\left| \int_{-1}^1 x\ dx \right| = 0 \leq 1 = \int_{-1}^1 |x|\ dx = 0.$$

[106]This *may* be slightly autobiographical.

6.7 Riemann Integration, Monotonicity, and Discontinuities

The second class of functions that are integrable are monotonic functions. The integrability on continuous functions was not a surprise, but this is. Proposition 4.2 limited the number of discontinuities of a monotonic function to a countably infinite number, but this is still an infinite number. The fact that such a function could feasibly be integrated is not something we should have expected.

Proposition 6.7. If f is monotonic on $[a, b]$, then f is integrable.

Proof. Let $\varepsilon > 0$. For any $n \in \mathbb{N}$, define a partition of $[a, b]$ such that

$$\Delta x_i = \frac{b-a}{n}.$$

We will show the result for a monotonically increasing function. Because the function is monotonically increasing, $M_i = f(x_i)$, and $m_i = f(x_i)$.[107] We have

$$\begin{aligned} U(P,f) - L(P,f) &= \sum_{i=1}^{n} (M_i - m_i)\Delta x_i \\ &= \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\frac{b-a}{n} \\ &= \frac{b-a}{n}\sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) \end{aligned}$$

By the Archimedean Property of $\mathbb{R}$, we can find an n large enough such that

$$U(P,f) - L(P,f) = \frac{b-a}{n}\sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) < \varepsilon$$

for all $\varepsilon > 0$. Because this n corresponds to a choice of a partition, we have found a partition that gives $U(P, f) - L(P, f) < \varepsilon$, so Riemann's Criterion is satisfied. □

Example 6.13.

Proposition 6.8. Suppose f is bounded on $[a, b]$ and has only finitely many points of discontinuity on $[a, b]$. Then f is Riemann integrable on $[a, b]$.

Proof. Let $\varepsilon > 0$. Set $M = \sup |f(x)|$, and $E = \{x \in [a, b] \mid f \text{ discontinuous}\}$. The set E is finite by assumption, so we can cover E by finitely many disjoint intervals $[u_j, v_j] \subset [a, b]$.[108] We can pick these intervals such that the sum of their lengths is ε:

$$\sum_{j=1}^{m} v_j - u_j < \frac{\varepsilon}{(b-a)+2M}.$$

Now we will remove the open segments (u_j, v_j) from $[a, b]$, giving a set K.

$$K = [a,b] \setminus \bigcup_{j=1}^{m} (u_j, v_j)$$

Because we removed open segments, K is still closed, so it is compact.[109] On the compact set K, f is continuous,[110] giving uniform continuity. There exists a $\delta > 0$ such that

$$|f(s) = f(t)| < \frac{\varepsilon}{(b-a)+2M}$$

[107]If f is always increasing, then the infimum on $[x_{i-1}, x_i]$ will be achieved at smallest value in the interval, namely x_i. Similarly, the supremum on the interval will be achieved at largest value in the interval, namely x_{i-1}.

[108]So if f is discontinuous as x_0, we form a little interval of length ε around it such that no other point in the small interval is a point of discontinuity.

[109]Removing segments never put the boundedness of $[a, b]$ in jeopardy, so clearly K is bounded.

[110]Remember, we just removed all the point pf discontinuity by removing (u_j, v_j).

for any $s, t \in K$ satisfying $|s - t| < \delta$.

Now construct a partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$ such that: $u_j \in P$ for all j, $v_j \in P$ for all j, $(u_j, v_j) \cap P = \emptyset$. We will have $\Delta x_i < \delta$, as $x_i, x_{i-1} \in K$...FINISH

□

Remark 6.4 (Integrability is Weaker than Continuity)**.** Integrability is often seen as equivalent to continuity in calculus courses, but this section goes to show this is not at all correct. Any monotonic function is integrable, even though it may have countably infinite discontinuities. Any bounded function with finite points of discontinuity is bounded. These facts make integrability a far weaker condition than continuity.

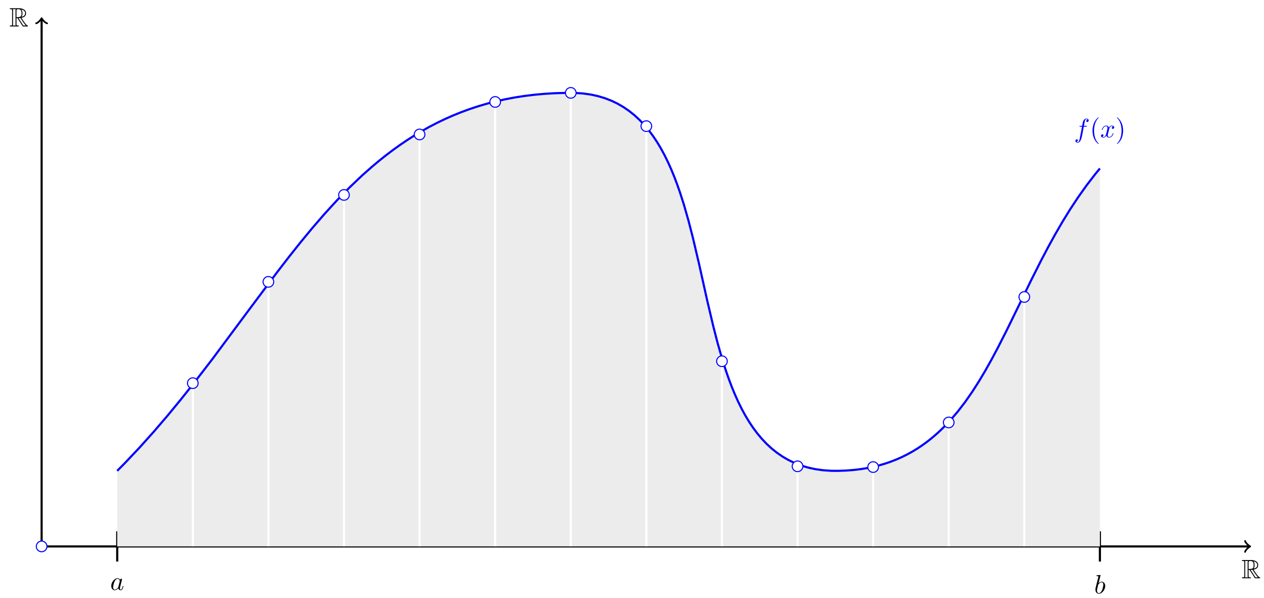

Figure 65: Despite having 12 points of discontinuity on $[a, b]$, f is still integrable.

Example 6.14 (Simple Functions)**.** Any simple function (Definition 6.6) is Riemann integrable, despiting them often being discontinuous. Example 6.3 and Figure 61 show an example of this type of discontinuous, yet integrable, function.

6.8 The Riemann-Stieltjes Integral (Optional)

Disclaimer: Again, we need to be a bit informal here. The purpose of this section is to build intuition for Sections 12-14, so I'm not as worried about using proper verbiage. Again, I'll be assuming that the "actual length", or simply "length", of an interval $[x_{i-1}, x_i]$ is $|x_i - x_{i-1}|$.

If you think back to the definition of the upper and lower Riemann sums, we always treated the width of the rectangle at $[x_{i-1}, x_i] \subset P$ as $|x_{i-1} - x_i|$. Is there anything stopping us from defining the width of the corresponding rectangle as $|2x_{i-1} - 2x_i|$? Or what about $|x_{i-1}^2 - x_i^2|$? It turns out, we can still perform Riemann integration if we decide to assign different widths to the rectangles of a Riemann sum.

Let $\alpha : [a, b] \to \mathbb{R}$ be any monotonic function. We can redefine the upper and lower Riemann sums as

$$U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha(x_i),$$

$$L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha(x_i),$$

where $\Delta\alpha(x_i) = |\alpha(x_i) - \alpha(x_{i-1})|$. The function α is determining how we assign length to different intervals in P. We can think of it as a *weighted* version of the original upper and lower sums, where α determine the weight assigned to different parts of $[a, b]$, and each interval is assigned an "artificial length" of $|\alpha(x_i) - \alpha(x_{i-1})|$. By "weight", I'm talking about how much "artificial" length we assign to an interval relative to it's "actual" length. It is meant in the same sense as a "weighted" average, where you take some observations to be more significant, and assign them more weight when calculating the mean. This clearly merits several examples.

Example 6.15 ($\alpha(x) = x$)**.** Let $\alpha(x) = x$. In this case we have

$$\alpha(\Delta x_i) = |\alpha(x_i) - \alpha(x_{i-1})| = |x_i - x_{i-1}|,$$

so each interval is assigned it's "true" length (Figure 66). This is a special case for two reasons. Firstly, each

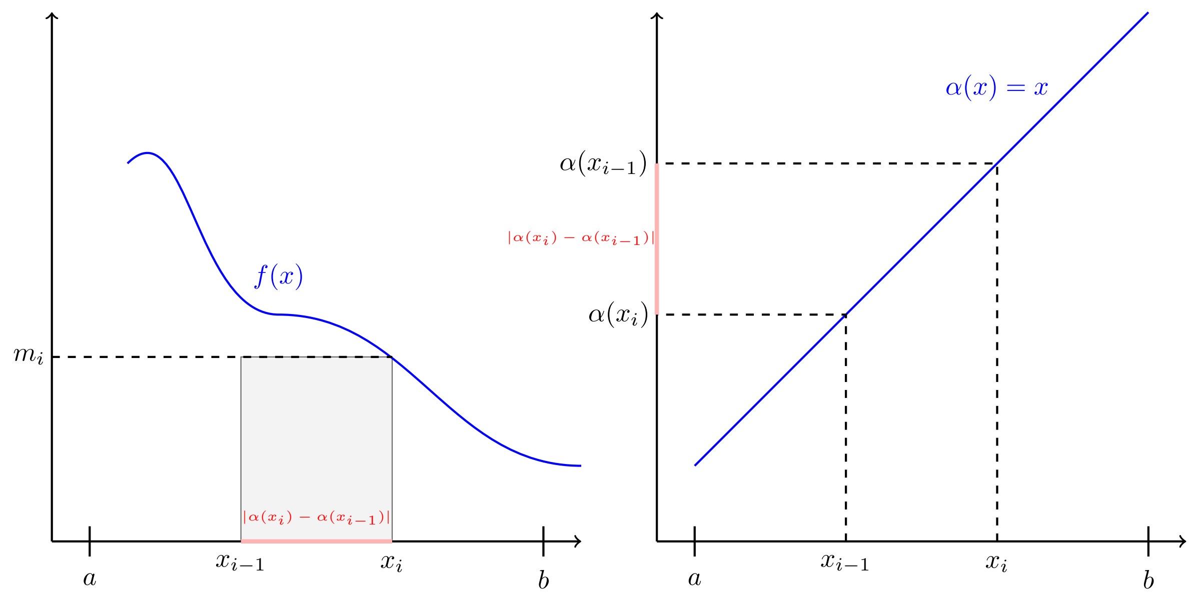

Figure 66: To determine how we weight the interval $[x_{i-1}, x_i]$, we calculate $|\alpha(x_i) - \alpha(x_{i-1})|$, and then assign it to the interval.

interval is weighted equally.[111] Secondly, each interval is weighted such that $|\alpha(x_i) - \alpha(x_{i-1})| = |x_i - x_{i-1}|$, so we get back the original Riemann sums. The weighted area assigned to the gray rectangle is

$$m_i \Delta\alpha(x_i) = m_i \Delta x_i.$$

Example 6.16 ($\alpha(x) = 2x$)**.** Let $\alpha(x) = 2x$. In this case we have

$$\alpha(\Delta x_i) = |\alpha(x_i) - \alpha(x_{i-1})| = |2x_i - 2x_{i-1}|,$$

so each interval is assigned twice its "actual" length (Figure 67). Now we have

$$m_i \Delta\alpha(x_i) = 2m_i \Delta x_i.$$

This is still a somewhat special case though, because we are assigning "artificial" lengths that are proportional to the "actual" length of an interval. For example, this choice of α gives $|\alpha(1) - \alpha(0)| = 2$ and $|\alpha(2) - \alpha(0)| = 4$. In both cases, the "artificial" length is double the "actual" length. In fact, if $\alpha(x)$ is linear, then each interval is weighted equally. For $\alpha(x) = \beta x + \gamma$,

$$\Delta\alpha(x_i) = |(\beta\Delta x_{i-1} + \gamma)| - |(\beta\Delta x_i + \gamma)| = \beta|x_{i-1}, x_i| = \beta\Delta x_i \propto \Delta x_i,$$

so all we're doing is scaling the "actual" length.

[111] Meaning that $|\alpha(x_i) - \alpha(x_{i-1})| \propto |x_i - x_{i-1}|$ for all i.

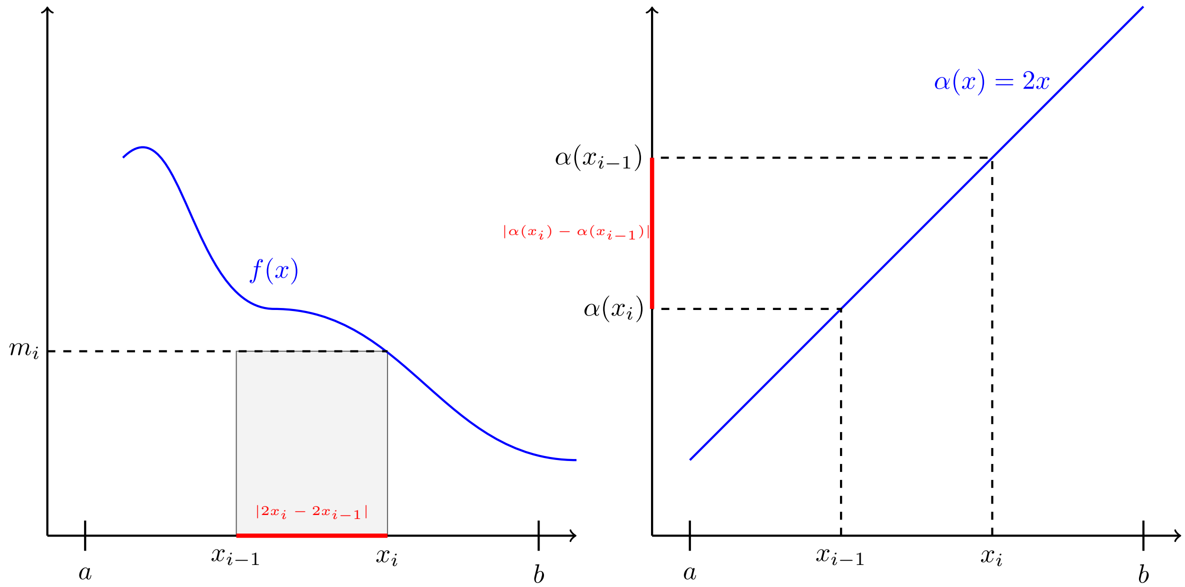

Figure 67: We now assign more weight to $[x_{i-1}, x_i]$ than we did with $\alpha(x) = x$.

Example 6.17. $\alpha(x) = x^2$ Now let's see what happens for a nonlinear choice of $\alpha(x)$. Suppose we have two intervals in a partition, $[0, 0.1]$ and $[1, 10]$. With this choice of α we have

$$\begin{aligned} |\alpha(0.1) - \alpha(0)| &= 0.01, \\ |\alpha(10) - \alpha(1)| &= 81. \end{aligned}$$

This value of α assigns a lot of weight to large intervals, and a small amount of weight to small intervals. This is the first example we've seen where α truly assigns different weights to different intervals, and not just scaling up every length by a constant.

Using α, we can redefine the Riemann integral as developed in Subsection 6.2.

$$\underline{\int_a^b} f(x)\ d\alpha = \sup_{P \in \mathbf{P}([a,b])} L(P, f, \alpha), \qquad \overline{\int_a^b} f(x)\ d\alpha = \inf_{P \in \mathbf{P}([a,b])} U(P, f, \alpha).$$

Definition 6.10. Suppose f is a bounded real function on the interval $[a, b]$ and α is a monotonic function on $[a, b]$. If

$$\underline{\int_a^b} f(x)\ d\alpha = \overline{\int_a^b} f(x)\ d\alpha,$$

then we say f is *Riemann–Stieltjes integrable (on $[a, b]$)* and we write the common value of the upper and lower Riemann integral as

$$\int_a^b f(x)\ d\alpha.$$

We refer to this common value as the *Riemann–Stieltjes integral of f on [a,b]*.

The Riemann–Stieltjes (RS) is simply a weighted integral. Nearly all the results developed up until now can be established for the RS integral. In fact many standard texts such as Rudin (1976) opt to work exclusively with the RS integral for this reason.

Remark 6.5 (Why Monotonic?)**.** If α is not monotonic, then it could assign negative "artificial length" to an interval $[x_{i-1}, x_i]$ (which requires $x_i > x_{i-1}$).

Example 6.18 (Probability Distributions)**.** A random variable X has a cumulative distribution function (CDF) $F : [a, b] \to [0, 1]$ such that

$$\Pr(c \le X \le d) = F(d) - F(c).$$

The CDF is assigning a probability to $[c, d]$. Depending on our choice of F, the weight assigned to each interval in probability will differ. If F is a uniform distribution, every interval is assigned equal probability, i.e equal weight. But this sounds *just like* RS integration! If we let $\alpha(x) = F(x)$, then we have

$$\int_c^d dF = \int_c^d 1\ dF = 1(F(d) - F(c)) = \Pr(c \le X \le d).$$

Now suppose that the random variable X only takes on values $x \in [a, b]$. If we want to find the expected value of X over, then we calculate a the RS integral of x over all of $[a, b]$.

$$E[x] = \int_a^b x\ dF$$

This is akin to a weighted average, because we are averaging every single possible realization of X, but we first assign weight to the interval $[a, b]$ in accordance with the probability of each event occurring.

Example 6.19 (Inertia)**.** Suppose there is a rod of unit length where the mass contained in $[0, x]$ is given by $m(x)$. The inertia I is given as

$$I = \int_0^1 x^2\ dm.$$

If you have taken a probability of physics course, these last two examples may look somewhat familiar. Using integration to calculate the expected value of a random variable, or the moment of inertia are standard applications, but it is normally introduced as a Riemann integral of a different quantity, not a RS integral. What is the connection between the two? Is it possible to calculate a RS integral as a Riemann integral? The next theorem allows us to do just this.

Theorem 6.4 (RS Integral as Riemann Integral)**.** Assume α increase monotonically and α' is integrable on $[a, b]$. Let f be a bounded real function on $[a, b]$. The function f is RS integrable on $[a, b]$ *if and only if* $f\alpha'$ is Riemann integrable on $[a, b]$. In this case we have

$$\int_a^b f(x)\ d\alpha = \int_a^b f(x)\alpha'(x)\ dx.$$

Before we prove this, we should rationalize why this equation holds. We want to integrate $f(x)$ as if the interval $[a, b]$ is not weighted. To do this we need to convert between $\alpha(x)$ and x somehow. For the RS sums, $\alpha(x)$ is manipulating the width of a rectangle, thereby manipulating its area. But couldn't we instead manipulate the height of the rectangle to achieve the same change in area? If we were to do this, we would need to multiply the height of the rectangle by the same factor that we change the width by. But what quantity captures the change in width? That would be α'. What determines the height of the rectangle? That is f (vicariously through its supremum or infimum). So if we let the height of the rectangle be $f(x)\alpha'(x)$, we acheive the same area that resulted from reweighting the width using α.

Let's accompany this argument with a quick example. Using a constant function would be best, as the integral in this case just amounts to a rectangle. Let $f(x) = 10$ on $[5, 10]$, and $\alpha(x) = 2x$. If we calculate the area using a RS integral, we have

$$\int_5^{10} f(x)\ d\alpha = \int_5^{10} 10\ d\alpha = 10[2(10) - 2(5)] = 100.$$

Alternatively, we could acheive the same area under $f(x)$ by multiply $f(x)$ by 2, which just happens to be α' (which we are about to prove is no coincidence).

$$\int_5^{10} f(x)a'(x)\ d\alpha = \int_5^{10} 10 \cdot 2\ dx = 20(10 - 5) = 100.$$

We get the same exact value. The actual proof, which is on the harder side, is not nearly as important as understanding the intuition behind the result.

Proof. The function α is integrable, so it satisfies the Riemann Criterion. That is for $\varepsilon > 0$ there exists a partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$ such that

$$U(P, \alpha') - L(P, \alpha') < \frac{\varepsilon}{\sup |f(x)|}.$$

By the differentiability of α', we can apply the Mean Value Theorem on each interval $[x_{i-1}, x_i]$. There exists a $t_i \in [x_{i-1}, x_i]$ such that

$$\begin{aligned} \alpha_i(x_{i-1}) - \alpha_i(x_i) &= \alpha'(t_i)(x_{i-1} - x_i), \\ \Delta\alpha_i &= \alpha'(t_i)\Delta x_i, \end{aligned} \tag{15}$$

for $i = i, \ldots, n$. By Part 2 of Proposition 6.1 and the integrability of α', for $s_i \in [x_{i-1}, x_i]$,[112]

$$\sum_{i=1}^{n} |\alpha'(s_i) - \alpha'(t_i)|\Delta x_i < \frac{\varepsilon}{|\sup f(x)|}. \tag{16}$$

By (15),

$$\sum_{i=1}^{n} f(s_i)\Delta\alpha_i = \sum_{i=1}^{n} f(s_i)\alpha'(t_i)\Delta x_i. \tag{17}$$

If we set $M = \sup |f(x)|$, then (17) and (16) give

$$\left| \underbrace{\sum_{i=1}^{n} f(s_i)\Delta\alpha_i}_{U(P,f,\alpha)} - \underbrace{\sum_{i=1}^{n} f(s_i)\alpha'(t_i)\,\Delta x_i}_{U(P,f\alpha')} \right| \leq M\varepsilon.$$

This will hold for all $s_i \in [x_{i-1}, x_i]$, so

$$|U(P, f, \alpha) - U(P, f\alpha')| < M\frac{\varepsilon}{|f(x)|} = \varepsilon.$$

This all will hold for any refinement of P, so we can conclude

$$\left| \overline{\int_a^b} f(x)\ d\alpha - \overline{\int_a^b} f(x)\alpha'(x)\ dx \right| \leq \varepsilon.$$

This holds for all $\varepsilon > 0$, so

$$\overline{\int_a^b} f(x)\ d\alpha = \overline{\int_a^b} f(x)\alpha'(x)\ dx.$$

This *exact* reasoning will give

$$\underline{\int_a^b} f(x)\ d\alpha = \underline{\int_a^b} f(x)\alpha'(x)\ dx,$$

so we can conclude

$$\int_a^b f(x)\ d\alpha = \int_a^b f(x)\alpha'(x)\ dx.$$

□

[112]Part 2 of Proposition 6.1 requires two points in$s_i, t_i \in [x_{i-1}, x_i]$. We will use the t_i that we found using the Mean Value Theorem.

Example 6.20 (Probability Distributions)**.** A random variable X with CDF F, has a probability density function (PDF) defined as $F' = f$. If we apply Theorem 6.4 to Example 6.16 we get

$$\Pr(c \le X \le d) = \int_c^d dF = \int_c^d F'(x)\ dx = \int_c^d f(x)\ dx,$$
$$E[x] = \int_a^b x\ dF = \int_a^b xF'(x)\ dx = \int_a^b xf(x)\ dx.$$

This is the last time we'll see the RS integral, but the ideas which motivate it will return.

6.9 The Fundamental Theorem Of Calculus and Consequences

Wait...how do we actually calculate these things? So far we've only been able to integrate constant functions (Example 6.1).[113] This section will present *the* theorem of calculus. The Fundamental Theorem of Calculus allows us to easily calculate integrals by the means of derivatives. But is this what we really care about? From the standpoint of analysis, who cares about calculating the area under a curve? Mathematicians have been able to do this quite accurately for millennial. What is *amazing* is that the area under a curve is somehow related to differentiation at all. Despite being able to calculate areas and rates for centuries, mathematicians thought these two tasks were totally unrelated. This is to say, *everyone* takes this theorem for granted. When introduced to integrals in calculus, most are taught that integration is the opposite of differentiation only minutes after the integral is defined, but this is not a very obvious of a relationship.[114] Math students are so conditioned to think of integration hand-in-hand with differentiation, that it makes it hard to appreciate the Fundamental Theorem from the standpoint of analysis. **So do not think of this section as a means to calculation, think of it as a beautiful insight into two seemingly disparate topics**. Pontification over.

For such an important pair of results, The Fundamental Theorem of Calculus is not overwhelmingly difficult to prove. The second part is just an application of the Mean Value Theorem,[115]

Definition 6.11. Let F be a real differentiable function. If $F' = f$, then we say F is the antiderivative of f.

Theorem 6.5 (Fundamental Theorem of Calculus I)**.** Let f be integrable on $[a, b]$. For $a \le x \le b$, let

$$F(x) = \int_a^x f(t)\ dt.$$

Then F is continuous on $[a, b]$; furthermore, if f is continuous at a point $x_0 \in [a, b]$, then F is differentiable, and $F'(x_0) = f(x_0)$.

Proof. We will first show that F is continuous Because f is integrable, it is bounded, so there exists an M for which $|f(t)| \le M$ on $[a, b]$. If $a \le x < y \le b$, Theorem 6.2 Parts 3 and 5 give

$$|F(y) - F(x)| = \left| \int_x^y f(t)\ dt \right| \le M(y - x). \tag{18}$$

If we have $|y - x| < \varepsilon/M$, then

$$|F(y) - F(x)| < \varepsilon,$$

so F is continuous.[116]

Now suppose f is continuous at x_0. For $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|f(t) - f(x_0)| < \varepsilon$$

[113] If we think of the integral in terms of simple functions, we can integrate those too, but only because we defined the integral explicitly in that case.

[114] You're probably thinking "yes it is", but keep in mind that you have known this for years, and have most likely seen several interpretations of the relationship.

[115] This is one reason that the Mean Value Theorem is so important.

[116] It actually ended up being uniformly continuous, so happy day.

for $t \in [a,b]$ which satisfy $|t - x_0| < \delta$. Therefore if we have

$$x_0 - \delta < s \leq x_0 \leq t < x_0 + \delta,$$

and $a \leq s < t \leq b$, by Theorem 6.2 Part 5 we have

$$\left| \int_s^t \underbrace{f(u) - f(x_0)}_{<\varepsilon} \ du \right| < \varepsilon(t-s)$$

Not only does this inequality take the form of (18),[117] but we satisfy the conditions under which (18) holds, so we have

$$\begin{aligned} |F(t) - F(s) - f(x_0)| &= \left| \int_s^t f(u) \ du - f(x_0) \right| < \varepsilon(t-s) \\ \left| \frac{F(t)-F(s)}{s-t} - f(x_0) \right| &= \left| \frac{1}{s-t} \int_s^t f(u) - f(x_0) \ du \right| < \varepsilon \\ \left| \frac{F(t)-F(s)}{s-t} - f(x_0) \right| &< \varepsilon. \end{aligned}$$

That is, for all $\varepsilon > 0$ this final inequality holds for any $s \in (x_0 - \delta, x_0 + \delta)$, so if we take $t = x_0$ this becomes the definition of the following limit:

$$F'(x_0) = \lim_{s \to x} \frac{F(t) - F(x_0)}{s - x_0} = f(x_0).$$

□

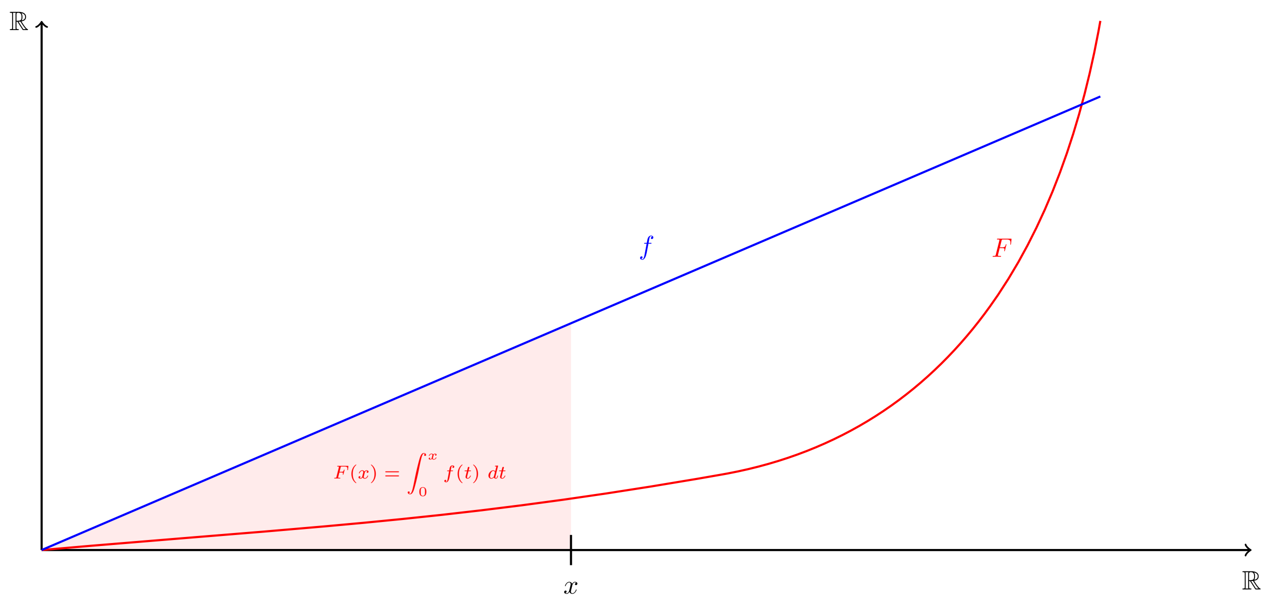

Figure 68: The first part of The Fundamental Theorem of Calculus says that $F' = f$.

Theorem 6.6 (Fundamental Theorem of Calculus II)**.** Let f be integrable on $[a,b]$. If there exists a differentiable function F on $[a,b]$ such that $F' = f$, then

$$\int_a^b f(x) \ dx = F(b) - F(a).$$

[117] Let $M = \varepsilon$, $y = t$, $x = s$, and $f(t) = f(u)$.

Proof. Let $\varepsilon > 0$. By Riemann's Criterion, there exists some $P + \{x_0, \dots, x_n\}$ of $[a, b]$ such that $U(P, f) - L(P, f) < \varepsilon$. By the Mean Value Theorem, for each i we have a $t_i \in [x_{t-1}, x_t]$ such that

$$F(x_i) - F(x_{i-1}) = F'(t_i)(x_i - x_{i-1}),$$
$$F(x_i) - F(x_{i-1}) = f(t_i)\Delta x_i.$$

Taking the sum over all i gives

$$\sum_{i=1}^{n} f(t_i)\Delta x_i = F(b) - F(a).$$

By Proposition 6.1 Part 3,

$$\left| F(b) - F(a) - \int_a^b f(x)\ dx \right| = \left| \sum_{i=1}^{n} f(t_i) - \int_a^b f(x)\ dx \right| < \varepsilon.$$

This holds for all ε, so

$$\int_a^b f(x)\ dx = F(b) - F(a).$$

□

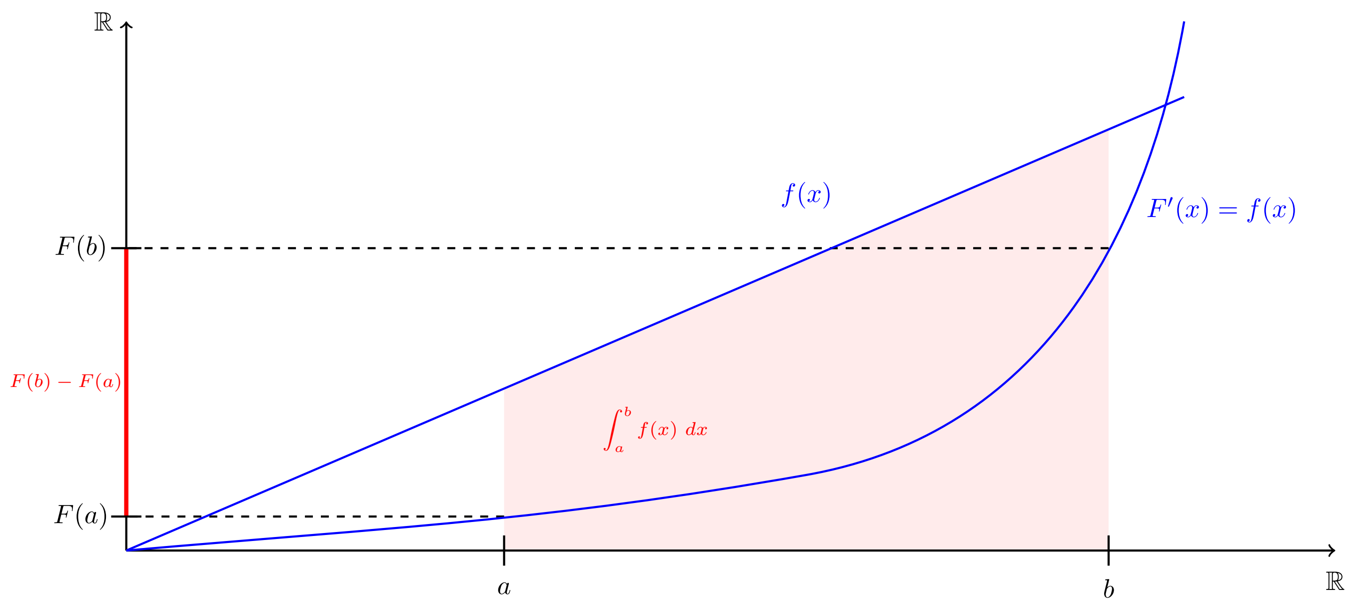

Figure 69: The second part of The Fundamental Theorem of Calculus says that if $F' = f$, then we have $\int_a^b f(x)\ dt = F(b) - F(a)$.

In essence, the Fundamental Theorem of Calculus (FTC) tells us that integration and differentiation are inverse operations.

A great application of the FTC is to find some integration technique that is analogous to the Product Rule of differentiation.

Proposition 6.9 (Integration by Parts)**.** Suppose F and G are differentiable functions on $[a, b]$, and $F' = f$ and $G' = g$ are integrable as well. Then

$$\int_a^b F(x)g(x)\ dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)\ dx.$$

Proof. Let $H(x) = F(x)G(x)$. We have

$$H'(x) = F'(x)G(x) + G'(x)F(x) = f(x)G(x) + g(x)F(x),$$

so H' is integrable by Proposition 6.5. The FTC gives

$$\begin{aligned}
\int_a^b H'(x)\ dx &= H(b) - H(a) \\
\int_a^b f(x)G(x) + g(x)F(x)\ dx &= F(b)G(b) - F(a)G(a) \\
\int_a^b f(x)G(x)\ dx + \int_a^b g(x)F(x)\ dx &= F(b)G(b) - F(a)G(a) \\
\int_a^b F(x)g(x)\ dx &= F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)\ dx
\end{aligned}$$

Remark 6.6 (But What About Discontinuous Functions?)**.** Subsection 6.7 showed that the set of Riemann integrable functions includes many more functions than just those are which continuous. Unfortunately, the FTC requires continuity. We still have no clear way of treating discontinuities when integrating. Example 6.20 and Example 6.23 will discuss this more.

□

6.10 Change of Variables

Change of variables may be something you first saw explicitly in multivariable calculus when working with polar or spherical coordinates, but you actually learned a basic form of it earlier. In the case functions of a single variable it is often called u-substitution. For the sake of generalization later on, we'll refer to it as change of variables, but just keep in mind that all it is, is u-substitution.

The general idea of change of variables is that it is the opposite of the Chain Rule.

$$\int_{\phi(a)}^{\phi(b)} f(x)\ dx = \int_a^b f(\phi(y))\phi'(y)\ dy$$

We can justify this equality with reasoning similar to that which accompanied Theorem 6.6. This is best done with an example, as we need to deal with two sets of Riemann sums.

Let $f(x) = 5$, $\phi(y) = 2y$,and $[a, b] = [0, 10]$. We have

$$\int_{\phi(a)}^{\phi(b)} f(x)\ dx = \int_0^{20} 5\ dx = 100, \tag{19}$$

$$\int_a^b \phi(y)\ dy = \int_0^{10} 2y\ dy = 25. \tag{20}$$

Equation (19) corresponds to a square of area 100, while Equation (20) corresponds to a square of area 25. The goal of change of variables is to manipulate the area of the triangle formed by ϕ so it has an area equal to the square formed by f. This allows us to integrate f without even considering its domain. We instead work in ϕ's domain. First, we want to change the height of the Riemann sums which form the triangle so they correspond to the height of the sums corresponding to f and the rectangle (see Figure 68). This is achieved by composing ϕ with f. Our updated area is the integral of $f(\phi(y)) = 5$ over $[0, 10]$.

$$\int_0^{10} 5\ dx = 50. \tag{21}$$

But we aren't done yet. The difference between (21) and (19) is the width of the corresponding rectangles. We cannot change the width of the prior though, because the whole point of this is to integrate in the domain of φ. The bounds of integration should not be touched. We can instead scale the height of this rectangle

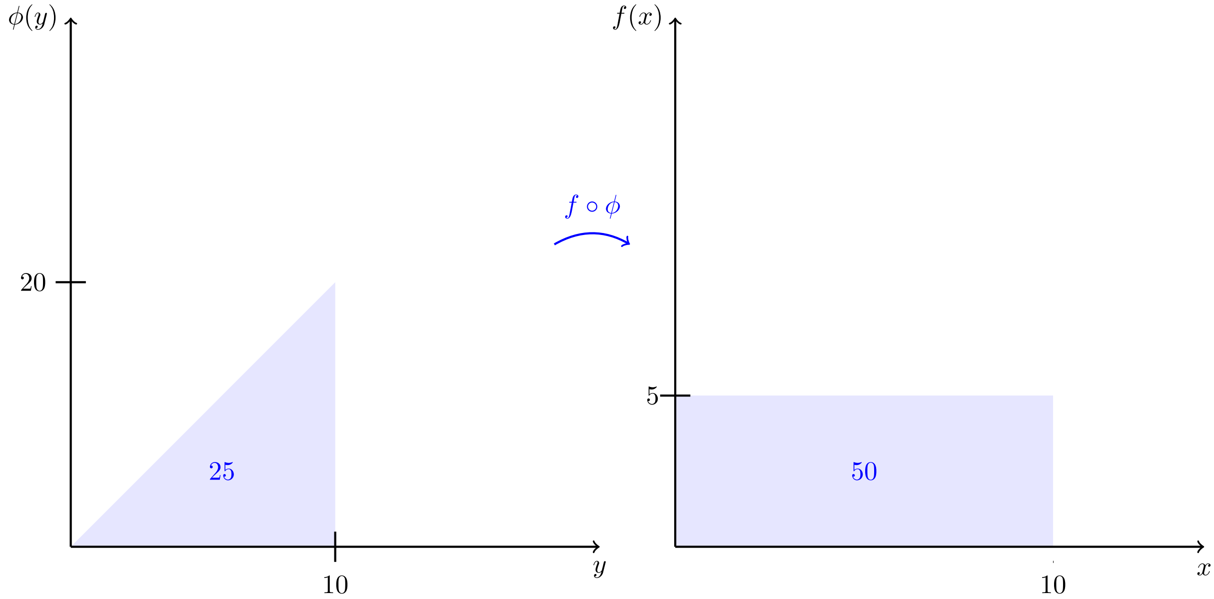

Figure 70: By composing ϕ with f, we modify the height of the integral of ϕ.

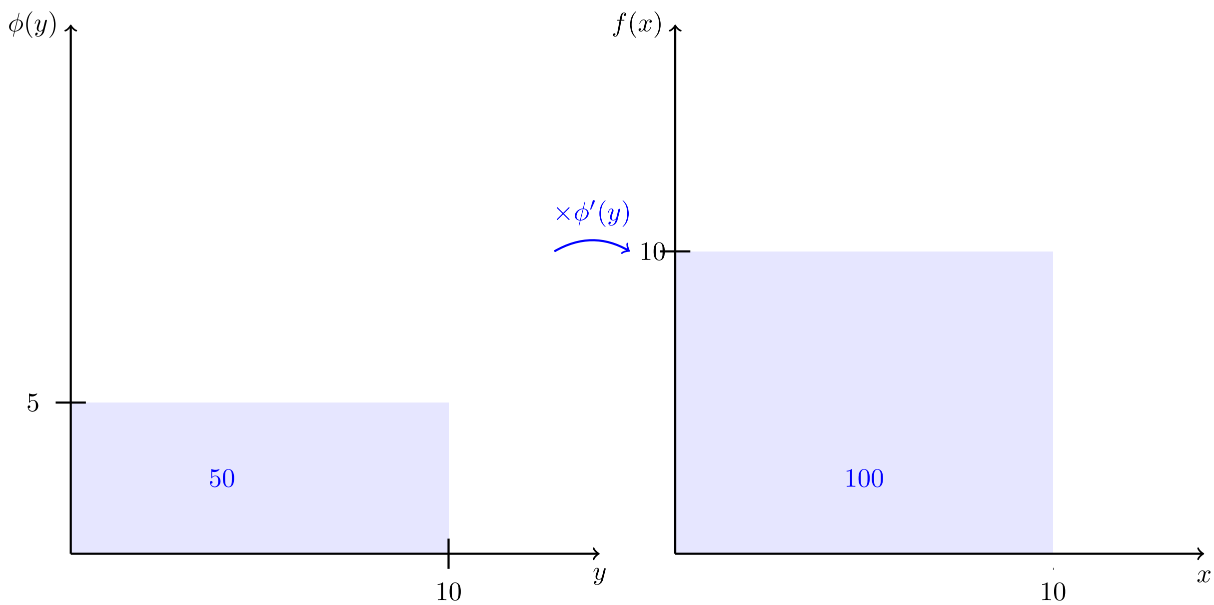

Figure 71: By scaling the height of the square from Figure 67 by a factor of ϕ', we get the desired area.

by $\varphi'(y)$, which is the rate at which the width would be scaled if we decided to move to the domain of f (Figure 68). This gives us

$$\int_0^{10} 5\phi'(y)\ dy = \int_0^{10} 5 \cdot 2\ dy = 100.$$

Therefore we have scaled the area in (16) to that in (15) by geometrically transforming the function over which we are integrating. In one step this is:

$$\int_{\phi(a)}^{\phi(b)} f(x)\ dx = \int_0^{20} 5\ dx = 100 = \int_0^{10} 5 \times 2\ dy = \int_a^b f(\phi(y))\phi'(y)\ dy$$

To convince yourself this works, it may be helpful to draw the pictures for the reverse direction. That is, how would we manipulate the square of area 100, to end up with a triangle of area 25. This particular direction is actually the useful direction, as you start with the simpler integral! It is also the direction you proceed in when performing u-substitution.

Remark 6.7 (Why Did I Go to All This Trouble?)**.** This very simple example may seem a little redundant, but change of variables becomes very complicated when working with mulitvariable functions. It will be really important to have a strong intution about this process when we move to the general case. It also helps in this simplified case because change of variables doesn't have this nice geometric representation when working with nontrivial functions. Even picking f to be linear would have prevented us from using well known formulas for the area of shapes.

Theorem 6.7 (Change of Variables)**.** Suppose f is a bounded real function on $[a, b]$ which is integrable. If $\varphi : [a, b] \to [\varphi(a), \varphi(b)]$ is a continuous and strictly monotonic function, then $f(\varphi)$ is integrable on $[a, b]$, and

$$\int_{\phi(a)}^{\phi(b)} f(x)\ dx = \int_a^b f(\phi(y))\phi'(y)\ dy.$$

The proof will follow immediately from the Chain Rule and the Fundamental Theorem of Calculus.

Proof. Let F be the function such that $F' = f$. We have

$$F(\phi(y))' = F'(\phi(y))\phi'(y) = f(\phi(y))\phi'(y),$$

so $F(\phi(u))$ is the antiderivative of $f(\phi(y))\phi'(y)$. By the Fundamental Theorem of Calculus,

$$\int_{\phi(a)}^{\phi(b)} f(x)\ dx = F(\phi(b)) - F(\phi(a)) = \int_a^b f(\phi(y))\phi'(y)\ dy.$$

$\square$

Example 6.21. Suppose $f(x) = (2x^3 + 1)^7 x^2$. We would like to integrate

$$\int_0^1 (2x^3 + 1)^7 x^2\ dx,$$

but it is not clear how to integrate with respect to x in this case. We will instead change the variable we are integrating with respect to and appeal to Theorem 6.5. Let $\phi(x) = 2x^3 + 1$, and $f(u) = u^7$. Integration with respect to u gives

$$\int_0^1 (2x^3+1)^7 x^2\ dx = \frac{1}{6}\int_0^{\phi(1)} f(\phi(x))6x^2\ dx = \frac{1}{6}\int_{\phi(0)}^{\phi(1)} f(\phi(x))\phi'(x)\ dx = \frac{1}{6}\int_1^3 f(u)\ du = \frac{1}{6}\int_1^3 u^7\ du = \frac{410}{3}.$$

6.11 Shortcomings of Riemann Integration

We conclude this section by giving several examples of functions that are either not integrable, or we have no means of evaluating the integral of. From now until section 13, **when we write "integral", we will mean "Riemann integral"**, but you should keep these examples in the back of your head. They allude to the fact that we have only scratched the surface of integration.

Example 6.22 (Dirichlet Function)**.** Define $f : \mathbb{R} \to \mathbb{R}$ as

$$f(x) = \begin{cases} 1 \text{ if } x \in \mathbb{Q} \\ 0 \text{ if } x \notin \mathbb{Q} \end{cases}.$$

As we say in Example 4.26, this function is nowhere continuous. This function also fails to be integrable on any interval. Without loss of generality, we can show that f is not integrable on $[0, 1]$. By the density of

$\mathbb{Q}$ in $\mathbb{R}$, any $[x_{i-1}, x_i]$ will contain both a rational and irrational number for all $[x_{i-1}, x_i] \subset P$. This gives $M_i = 1$ and $m_i = 0$ for all i. Therefore

$$U(P, f) - L(P, f) = \sum_{i=1}^{n}(M_i - m_i)\Delta x_i = 1 - 0 = 1,$$

for any P. The Riemann Criterion does not hold, so f is not integrable on $[0, 1]$. While it is hard to argue one would ever need to integrate a nowhere continuous function, this is still dissatisfying. We know that $\mathbb{Q}$ is countable, where as $\mathbb{R}$ is not. This means that an uncountably infinite number of points of $f(x) = 0$ an uncountable number of times, where as $f(x) = 1$ a countable number of times. Is it crazy to think that the integral on $[0, 1]$ should just be zero then? Regardless of what we think it should be, it still is not Riemann integrable.

Example 6.23 (Unbounded Function)**.** Let $f(x) = 1/x$. This function cannot be integrated on $(0, 1]$, as it is unbounded on this interval, because $f(x-) = \infty$. If we instead tried to use the interval $[0, 1]$, then f would not be defined as $0 \in [0, 1]$.

Example 6.24 (Integral Over All of $\mathbb{R}$)**.** Suppose we want to integrate some function over all of $\mathbb{R}$. The whole real line is unbounded, so how do we measure the length of an infinite set and find upper and lower Riemann sums? Integration this way is not possible in this case.

Remark 6.8 (But What About Improper Integrals?)**.** You've likely seen improper integrals that allow us to evaluate integrals of functions on unbounded sets, but these technically are not Riemann integrals. They are limits of Riemann integrals. This is another topic we will take up in Section 13.

Example 6.25 (Discontinuities)**.** While the Fundamental Theorem of Calculus allows us to evaluate integrals for continuous functions, but how do we evaluate discontinuous functions? We know a function with a finite number of discontinuous can be integrated, but we never specified how those discontinuities contribute to the area. If we follow the reasoning from Example 6.11, we would hope that they contribute nothing to area, as there are an uncountably infinite number of continuous points on $[a, b]$ that greatly out number the discontinuous ones.

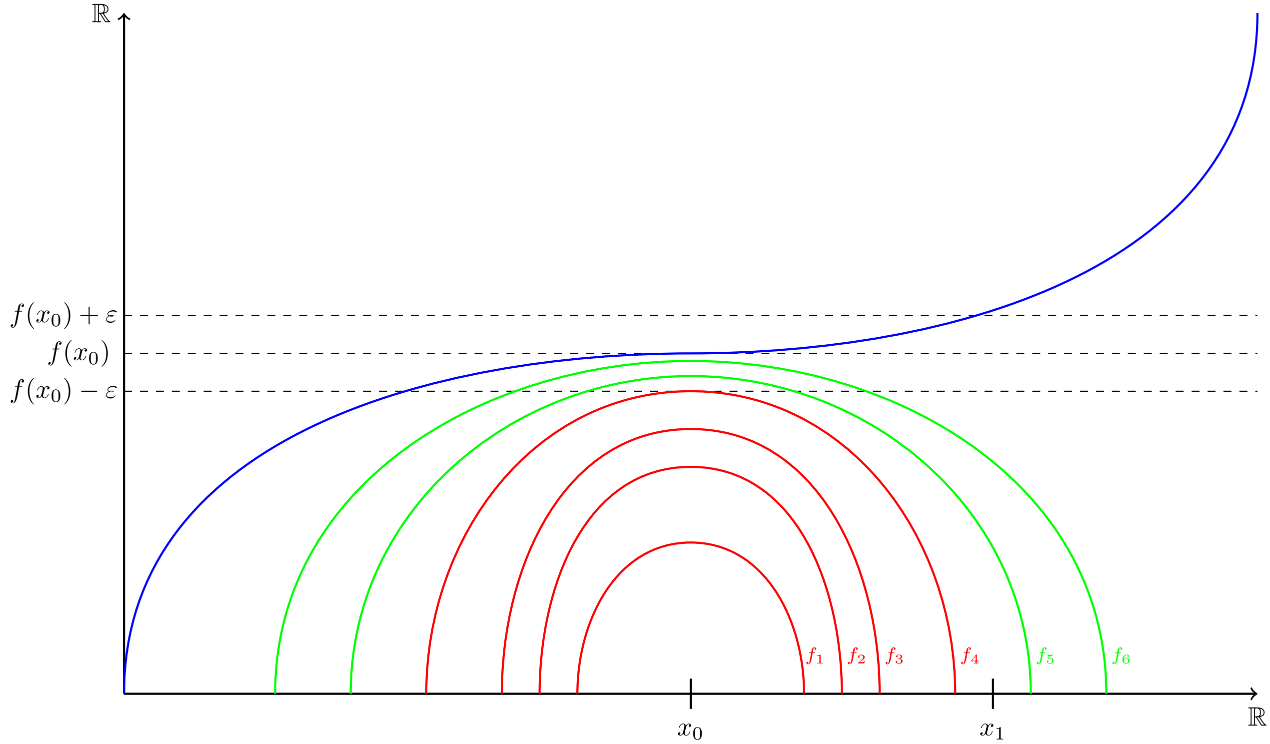

Figure 72: pointwise.

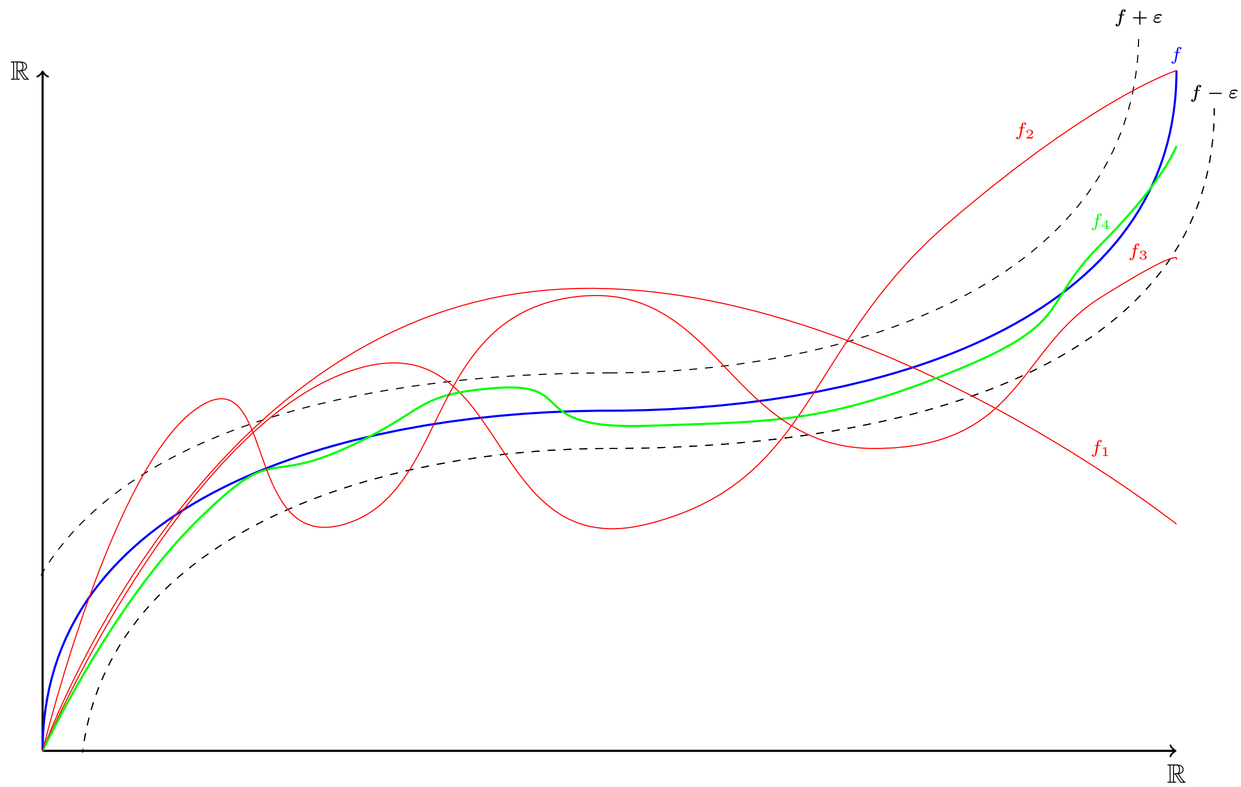

Figure 73: The sequence $\{f_n\}$ converges uniformly to f.

7 Sequences and Series of Functions

7.1 Metric Spaces of Functions

7.2 Sequences

7.3 Pointwise Convergence

7.4 Uniform Convergence

7.5 Properties of Uniform Convergence

7.6 Series

7.7 Power Series

7.8 Taylor Series

8 Real Functions of Several Variables

8.1 Linear Transformations

9 Differentiation with Several Variables

9.1 The Derivative as a Linear Map

9.2 The Chain Rule

9.3 The Inverse Function Theorem

9.4 The Implicit Function Theorem

10 Riemann Integration with Several Variables

10.1 Integration over a Rectangle

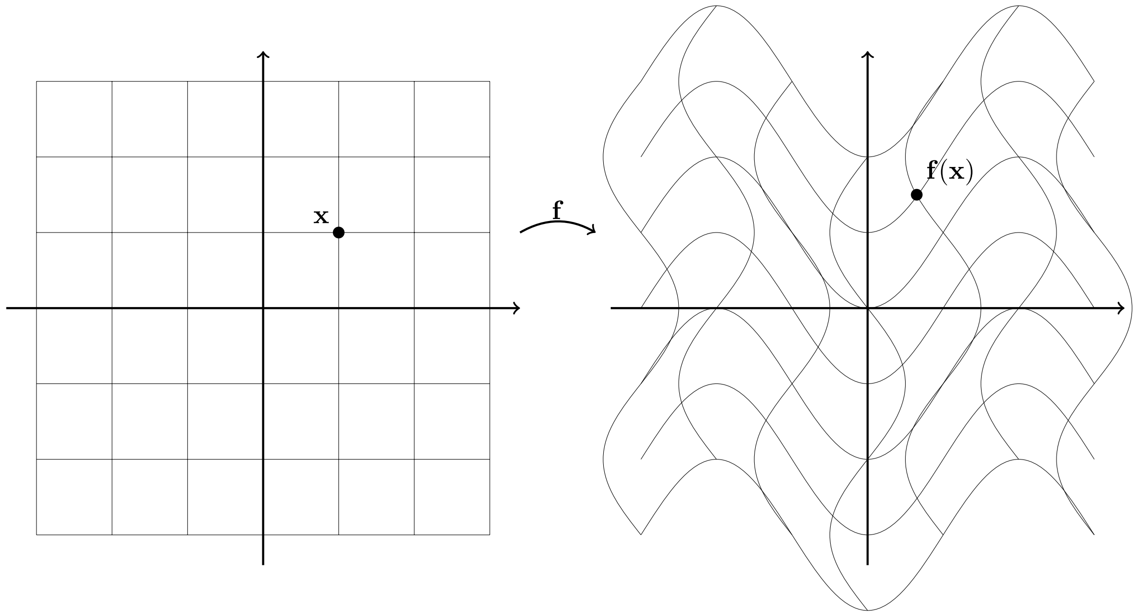

Figure 74: derivative linear map.

How do discontinuities play with integration

12.1 Motivation

12.2 σ-Algebras

12.3 Measures

dirac, counting,

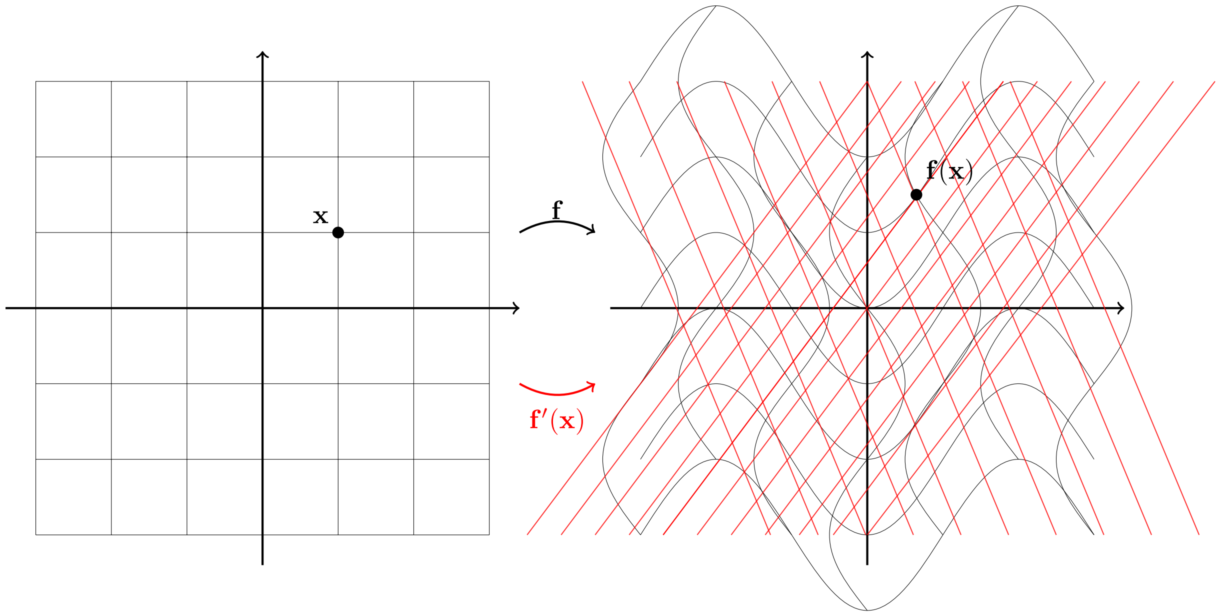

Figure 75: derivative linear map.

12.4 Measures on $\mathbb{R}$

13 Integration Revisited

13.1 Measurable Functions

13.2 Itegration of Simple Functions

13.3 Integration of Nonnegative Functions

13.4 Integration of Real Functions

13.5 Product Measures

13.6 Lebesgue Integration in $n-$Dimensions

14 Differentiation with Measures

14.1 Motivation in $\mathbb{R}$

14.2 Signed Measures

14.3 Radon-Nikodym Derivative

15 Point-Set Topology Revisited

15.1 Topological Spaces Revisited

15.2 Nets

15.3 Filters

15.4 Various Notions of Compactness

15.5 Continuity

15.6 The Product Topology

15.7 Pointwise and Uniform Convergence

16 Foundations of Functional Analysis

16.1 Normed Vector Spaces

ℓ^p, L^p

References

Aliprantis, C. D. and K. Border (2006). *Infinite Dimmensional Analysis: A Hitchhiker's Guide* (3 ed.). Springer.

Dummit, D. S. and R. M. Foote (2004). *Abstract Algebra* (3 ed.). John Wiley & Sons.

Grafakos, L. (2008). *Classical Fourier Analysis* (2 ed.). Springer.

Rudin, W. (1976). *Principles of Mathematical Analysis* (3 ed.). McGraw-Hill.

Rudin, W. (1991). *Functional Analysis* (2 ed.).

Tao, T. (2016a). *Analysis I* (1 ed.). Hindustan Book Agency.

Tao, T. (2016b). *Analysis II* (1 ed.). Hindustan Book Agency.