# Contents

1. **Introduction**  
   1.1 A little bit of history  
   1.2 Plan for the class and some things we will encounter  
   1.3 Approaching the course and studying  

2. **Metric spaces**  
   2.1 Notions of distance  
   2.2 Definition of metric space and examples  
   2.3 Open Sets  
   2.4 Sequences, Limits, and Closed Sets  
   2.5 Cauchy sequences and Completeness  
   2.6 Compactness  
   2.7 Connectedness  
   2.8 Continuity in metric spaces  
   2.9 Exercises  
   2.10 Metric Spaces Epilogue  

3. **Sets**  
   3.1 Definitions and basic operations  
   3.2 Cartesian products, relations, orders, and functions  
   3.3 Well-ordering and cardinality  
   3.4 Exercises  

4. **The real numbers and topology**  
   4.1 Algebraic properties of the reals  
   4.2 Order properties of the reals  
   4.3 Combining algebra and order  
   4.4 Point-set topology  
   4.5 Exercises
1 Introduction

1.1 A little bit of history

What we call real analysis came from the need to more precisely describe notions of the real numbers and functions on the reals. Zeno of Elea demonstrated a paradox in which it seems necessary to pack an infinite collection of intervals in a finite space, and thus we could conclude that motion is illusory. This has been resolved through a greater understanding of the real numbers by using summation of a geometric series. For example, we can say

\[ \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \left( \frac{1}{2} \right)^n + \cdots = 1, \]

but what do we really mean by the sum on the left hand side? (Zeno’s paradox may still have something to say because due to physical effects we may not be able to carry out this sum in a practical manner in the physical world)

The method of exhaustion was developed in Ancient Greece (and also seems to have been developed independently elsewhere) to find the area of a figure by inscribing it with polygons of increasing numbers of sides. The thought that the polygons approach the shape of the figure contained the ideas of a limit convergence of a sequence of areas to the area of a figure. The method of exhaustion was also used by many people to find extremely accurate (for the time) values of \( \pi \) by inscribing regular polygons of a staggering number of
sides inside circles. This was used by Archimedes to find that \( \frac{227}{71} < \pi < \frac{22}{7} \) and later used by many others to get better approximations. The question still remains, though, under what circumstances does the method of exhaustion work? Does it really work in all the situations it seems to, or is there some right combination of conditions which it fails our intuition? Alternatively, what more general circumstances can this idea be applied?

When Newton and Leibnitz discovered calculus, they were essentially developing a tool to solve differential equations. They had the right intuition, but some of their notions needed further clarification. For example, Newton used the word “fluent,” which means a flowing quantity, and its instantaneous rate of change to be “fluxion.” So in today’s terminology, a fluxion is a derivative and a fluent is a differentiable function. Leibnitz used infinitesimal quantities in his derivations. Both these methods needed some precision. What exactly is a flowing quantity? What are infinitesimals? These notions needed to be defined in terms of previously described notions. It is intuitively obvious what a flowing quantity ought to be - it is a quantity which has no jumps and no sudden changes in speed or direction, and then a fluxion is just the instantaneous rate of change. But then we get to a bit of circular logic. The instantaneous rate of change is measured by a fluxion, and for a fluxion to exist, we must know the function is fluent. But to really know the function is fluent, we have to know it has a fluxion. We really need a definition which relies only on basic properties of functions and the real numbers and not on intuitive terms like “flowing.”

It took close to two centuries through the work of Cauchy, Riemann, Bolzano, Weierstrass, and others for these notions to be developed and made precise enough for these issues to be resolved.

1.2 Plan for the class and some things we will encounter

Our approach will be metric spaces - some undefined and not yet justified notions, then set theory and axioms of the real numbers, then basic topology.

- Treating a function as a single entity, not just as a rule applied to sets
- Finiteness and infinity in surprising places
- Different sizes of infinities

The reasons I picked the texts I did.

1.3 Approaching the course and studying

On reading the book and these notes: when reading, don’t just read each definition as just a definition, each example as just an example, or each theorem as just a theorem. They are all these things, but when reading, keep in mind what each one is saying and how it relates with what you already know. I have also written some discussion which is meant to help in giving context to each concept,
I’ve written a lot of the explanations in a somewhat less formal tone, which I hope will help in your development of intuition of these concepts. Most analysis books give the definitions first and examples, and some (notably Rudin’s book) are pretty terse, but the development of analysis started with intuition, and the intuition of many examples was distilled into more formal definitions.

I will try to mark relevant asides with [Issue! . . .] when we encounter a concept which we haven’t really defined yet but we will later on in a more precise and thorough way.

When attempting the exercises, if you have some trouble starting, a good rule of thumb is to start with the definition. In many situations it can be an invaluable help to draw a picture of the situation. I will try to include some relevant pictures in these typed notes, but it may take a bit of \LaTeX\ wrangling.

Sometimes there are multiple notations for the same concept. I will try to keep the notation we use consistent throughout the course but also make mention of other notations as we go.

Please let me know of any errors or typos in these notes. I would be eternally grateful.

2 Metric spaces

2.1 Notions of distance

a. The set of real numbers has both algebraic and geometric properties. The basic relationship on which we build the geometry of the reals is the notion of distance. If $x, y \in \mathbb{R}$, then we define the distance between $x$ and $y$ to be

$$d(x, y) = |x - y|,$$

where the absolute value function $|\cdot| : \mathbb{R} \to \mathbb{R}$ is defined as

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

This definition is perfectly fine if we stay within the real numbers, but there are other sets of things which we want to have a sense of distance.

b. For example, we already know one notion of distance in $\mathbb{R}^2 = \{(a_1, a_2) : a_1, a_2 \in \mathbb{R}\}$. If $b = (b_1, b_2)$ and $c = (c_1, c_2)$ are elements of $\mathbb{R}^2$, then we can define the euclidean distance between the points to be

$$d_2(b, c) = \sqrt{(b_1 - c_1)^2 + (b_2 - c_2)^2}$$

(the subscript 2 indicates we are using the exponent 2 here on individual coordinates and power $\frac{1}{2}$ on the sum). This distance represents the length of the segment in the plane connecting the points $b$ and $c$. One particular good property this notion of distance has is that if you rigidly rotate the plane, then the resulting points will still be the same distance apart.
c. There are other ways we can define a distance, even on the same set. For example in \( \mathbb{R}^2 \) again, we can define another type of distance:

\[
d_1(b, c) = |b_1 - c_1| + |b_2 - c_2|
\]

The distance between two points is given to be the sum of the differences between the first coordinates and the second coordinates. This distance is sometimes called the taxicab metric or the Manhattan metric [citation needed - Fractals book?] because it illustrates very well the distance a car takes in getting between two places in a city with regular rectangular blocks. You can’t go straight there in a single direction; you have to take the streets. So if you are standing at a corner and have to go 3 miles north and 4 miles east, the distance you will have to travel is not 5 miles (if you could fly) but 7 miles (since you have to drive). Oh yeah, and there’s hot lava, so you can’t take a shortcut. (To be proper, Manhattan blocks are rectangular and not square, hence I measured in miles here and not blocks, and if you go 4 miles east from anywhere in Manhattan, you’ll end up in Brooklyn, Queens, or the Bronx, so bring subway fare.)

d. We can have still another distance on \( \mathbb{R}^2 \) by taking

\[
d_{\infty}(b, c) = \max\{|b_1 - c_1|, |b_2 - c_2|\}.
\]

This gives a sort of “what is the maximum distance away out of all the coordinates” sort of distance. We’ll see more of this later.

**Note:** If we translate back these three distances which we defined on \( \mathbb{R}^2 \) back to \( \mathbb{R}^1 \), they all reduce back to the standard distance on \( \mathbb{R} \), since \(|b - c| = |b - c|\), \(\sqrt{(b-c)^2} = |b - c|\), and \(\max\{|b - c|\} = |b - c|\). These three notions can also be naturally extended from \( \mathbb{R}^2 \) to \( \mathbb{R}^n \), just by adding more coordinates.

There are also other sets on which we’d like to define distance which aren’t some version of \( \mathbb{R}^n \). One thing to take note especially, is that in the geometry of \( \mathbb{R}^n \) (especially with a \( d_2 \) kind of distance) we often think of distance as the measure of length of the shortest segment between two points. There are many examples where we don’t necessarily have paths of shortest distance contained in the set in question, but we do have some notion of distance. A quick example would be to suppose we are considering two separate and disjoint disks sitting within the plane and distances from points in one disk to points in the other (it would probably help to draw a picture here). There are no paths between such pairs of points *completely contained within the disks*, so there can be no path of shortest length. Note that if we consider the disks plus the rest of the plane, the distances involved *do* come from paths of shortest length, but these paths have to pass through part of the plane *not* contained within either of the disks.

The following examples are of sets which don’t necessarily fit within \( \mathbb{R}^n \) but are important, anyway, so in studying them it is very useful to have some notion of distance.
e. Let \( n \) be a positive integer, and consider all the strings of 0s and 1s of length \( n \). Then we define the **Hamming distance** between two strings to be the number of individual places in which the strings differ. For example, the Hamming distance between “01000110" and “01001100" is 2, since the two strings differ in two places, namely the emphasized bits.

This notion of distance is very important in data transmission because we can design error checking codes based on the “nearest neighbor principle.” That is, suppose we pick some subcollection of strings, any two of which are a Hamming distance of at least 3 apart. Then if we encode a message using this subcollection, any one bit transmission error can be detected and corrected since a one bit error would result in a Hamming distance of only one away from the intended message.

We can see this through a concrete example. Suppose we encode 4-bit strings into 7-bit strings in the following fashion:

\[
\begin{align*}
0000 & \mapsto 0000000 \\
0001 & \mapsto 0001111 \\
0010 & \mapsto 0010011 \\
0011 & \mapsto 0011100 \\
\vdots & \\
1111 & \mapsto 1111111,
\end{align*}
\]

where \( abcd \mapsto abcd\bar{e}\bar{f}\bar{g} \) with the bits \( \bar{e}, \bar{f}, \) and \( \bar{g} \) defined by

\[
e + a + b + d \equiv f + a + c + d \equiv g + b + c + d \equiv 0 \mod 2.
\]

It’s not too hard to show that any two of these resulting seven bit strings are a Hamming distance of at least 3 away from each other, so any one bit error in transmission can be corrected to the intended bit string.

f. Let \( X \) be a nonempty set, and define a distance between points \( x, y \in X \) to be

\[
d(x, y) = \begin{cases} 
0, & x = y \\
1, & x \neq y.
\end{cases}
\]

This particular distance is known as the **discrete metric** on \( X \). It is called discrete because any two distinct points of \( X \) are completely isolated from each other. There is no way to approach a particular point from other points unless you jump a gulf of distance 1 to actually get there.

If \( X \) consists of \( n < \infty \) points, then we can model this within real euclidean space, that is we can consider the points to be the vertices of an equilateral triangle, a regular tetrahedron, or a regular corresponding polytope in \( \mathbb{R}^{n-1} \). However, if \( X \) is an infinite set, then it can’t be represented by a subset of any finite dimensional euclidean space. In that case, often it may be best to treat it abstractly. (Under further study, maybe a little beyond what we will cover in this course, we can show that any space with the discrete metric can be embedded in a suitable Hilbert space)
g. We will revisit this example many times throughout the course, so be sure to become familiar with $C[a, b]$.

Let $a, b \in \mathbb{R}$ with $a < b$. Then let $C[a, b]$ be the collection of functions $f : [a, b] \to \mathbb{R}$ which are continuous. [Issue! As an aside here, what is continuity? We will address this a little later in this chapter and fully after we’ve studied the axioms of $\mathbb{R}$.]

So the elements of $C[a, b]$ are functions. This may be a little surprising concept to some because in many lower division math classes functions are treated as rules, as means to an end, the end of transforming elements of one set into elements of another. But now we want to treat the functions themselves as objects worthy of study. So we can regard a continuous function $f : [a, b] \to \mathbb{R}$ as a “point” of $C[a, b]$. Two functions $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$ are the same if and only if for every $c \in [a, b]$, $f(c) = g(c)$. So in this way we tell our points apart; two points of $C[a, b]$ are the same if and only if the functions are the same function.

One way we have to tell how ‘far apart’ two functions are is the total area between the two functions. We can define a notion of distance by

$$d_1(f, g) = \int_a^b |f(x) - g(x)| \, dx.$$  

One thing to think about is that we are using the same terminology here as the $d_1$ notion of distance we used on $\mathbb{R}^2$. What are the similarities in the two situations? Is the use of the same notation justified?

h. Just as in the case of $\mathbb{R}^n$, we can have other notions of distance on $C[a, b]$.

We can define  

$$d_\infty(f, g) = \max_{x \in [a, b]} |f(x) - g(x)|.$$  

[Issue! How do we know the maximum always exists? In principle, the maximum could not exist, or there could be two functions whose difference goes to infinity. We will resolve these things after we examine the axioms of $\mathbb{R}$ and some topology.] Similar to the last case, is the $d_\infty$ notation similar enough to the $\mathbb{R}^2$ case to be justified?

So what are the properties the we want a notion of distance to have? What basic properties do all these examples possess? Or maybe do some of these examples have properties we want and the rest we’d rather not deal with? Ideally we’d like to discern some properties which are general enough so to encompass all these spaces since these are definitely of interest, but can we really do it?

Think about this, and use the rest of the page to write your ideas. If reading this on a computer, just write directly on the screen (This joke probably won’t be so funny in five years).
2.2 Definition of metric space and examples

Well, first we'd like the distance between two points to be a nonnegative real number. This just seems to make since we always measure a nonnegative linear distance. (The notions of negative and imaginary distance, however, have been developed, and the ramifications are the subject of a lot of research in differential geometry and relativity.) Also, the distance from a point to itself should be zero, that is, there shouldn't be a journey to stay in the same place (Carl Jung might have something to say about that).

Another thing we’d like to have is that distance can tell distinct points apart, so if \( x \) and \( y \) are distinct elements of a space \( X \), then a reasonable distance should have \( d(x, y) > 0 \). Of course, we can have functions \( d \) for which \( d(x, y) = 0 \) even though \( x \neq y \), and they are certainly worthy of study.

Also, one thing to notice is that the distance as measured from \( x \) to \( y \) is the same as the distance as measured from \( y \) to \( x \). It would be a funny kind of distance to be able to walk from one point to another with the trip back being many light years long. Of course, there are examples of the distances between points not being symmetric, such as driving in a city with many one way streets, but we would like symmetry in however we measure the shortest distance.

We seem to have covered how a point relates with itself and two points relate to each other, but what about three or more points? Well, we’d like to keep distances reasonable and well-behaved, and in particular we wouldn’t like to have a situation similar to this: to go from point A to point B takes only a few steps, and to go from point B to point C takes only a few steps, but to get from A to C, it’s a three day road trip. Certainly there is a way, if we use B as a waystation, to get from A to C in only several steps, so our notion that the distance from point to point represents some sort of minimal path between the points is violated by this example.

So that brings us to the definition of a metric space.

**Definition 2.1.** A set \( X \) together with a function \( d : X \times X \to \mathbb{R} \) is called a **metric space** if the following hold:

a. For every \( x, y \in X \), \( d(x, y) \geq 0 \), and \( d(x, y) = 0 \) iff \( x = y \). (positive definiteness)

b. For every \( x, y \in X \), \( d(x, y) = d(y, x) \). (symmetry)

c. For every \( x, y, z \in X \), \( d(x, y) + d(y, z) \geq d(x, z) \). (triangle inequality)

We often will call \((X, d)\) a metric space or just call it \( X \), if the metric is understood. **Note:** We can have many possible metrics on the same set, so even though the basic sets are the same, if the spaces have two different metrics, then they are two *different* metric spaces. The previous sentence is very important; please read it again.

The first question we may have about this definition is, ”Does it make sense?” Does this coincide with the intuition we have and describe it precisely? Well, positive definiteness allows us to assign a distance between any two points, and
since it’s definite, we can tell distinct points apart. The symmetry axiom tells us that the distance between two points is the same regardless of where we start and where we end. The triangle inequality makes sure that the metric stays relatively well-behaved, as in the road trip idea. So the definition fits all that we want. One question remains: is this all that we need, or is there some other property which is essential but that we haven’t covered? With the benefit of lots of examples and hindsight, I can assure you that when talking about general spaces, these three axioms are exactly what we need. When considering specific examples, there are more properties involved, so we can gain results particular to these examples.

One of the interesting problems in mathematics is when given a collection of examples which possess a particular property, how do we distill these examples to arrive at a precise formulation of the property in question? Also, how do we cull the examples which don’t exactly fit the idea? Often this process involves many people developing many examples and testing things over time until someone or some people arrive at the proper definition. This is one of the important parts of research.

So we should start with a collection of examples of metric spaces. Fortunately, each of the examples we had in looking at notions of distance is indeed a metric space!

**Example 2.2.**

a. The real numbers $\mathbb{R}$ under the standard notion of distance $d(x, y) = |x - y|$ is a metric space.

*Proof. First, since $|a| \geq 0$ for all real $a$ and $|a| = 0$ if and only if $a = 0$, this shows that the function $d$ is positive definite. For symmetry, observe that for any real numbers $x$ and $y$, $|x - y| = |y - x|$. To show the triangle inequality, by relabeling the points assume $x \leq y \leq z$. Then $|x - z|$ is larger than each of $|x - y|$ and $|y - z|$, so this shows $d(x, y) \leq d(y, z) + d(x, z)$ and $d(y, z) \leq d(x, y) + d(x, z)$. Since $y$ is somewhere in between $x$ and $z$ (and may be equal to one or both), we may state $d(x, z) = d(x, y) + d(y, z)$ from which follows the inequality $d(x, z) \leq d(x, y) + d(y, z)$. Thus the inequality holds for any three points in any configuration. ☐

Note that we proved the triangle inequality three times in this situation. This comes from our assumption about labeling the points. For any three points we will have a lowest point and a highest point, and for simplicity we have called the lowest $x$ and the highest $z$. If we assumed an arbitrary labeling of our three arbitrary points then we could have proved the inequality once, but it would have been more difficult.

b. The real plane $\mathbb{R}$ under the standard euclidean metric

$$d_2(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

is a metric space.
Proof. Exercise.

(c) The plane \( \mathbb{R}^2 \) with the distance \( d_1((b_1, b_2), (c_1, c_2)) = |b_1 - c_1| + |b_2 - c_2| \) is a metric space.

Proof. By definition \( d_1((b_1, b_2), (c_1, c_2)) \) is a sum of two nonnegative things which makes it nonnegative. If \( b = c \), then the distance between them is 0, and if the distance is zero then \( |b_1 - c_1| = |b_2 - c_2| = 0 \). Thus the function is positive definite.

We can establish symmetry in the same way as we did for the standard metric on \( \mathbb{R} \). The triangle inequality follows from the triangle inequality in \( \mathbb{R}^1 \), we just observe the inequality in each coordinate then add the inequalities.

(d) The plane \( \mathbb{R}^2 \) with the distance \( d_\infty((b_1, b_2), (c_1, c_2)) = \max\{|b_1 - c_1|, |b_2 - c_2|\} \) is a metric space.

Proof. Positive definiteness and symmetry are straightforward (you should check these). For the triangle inequality, let \( b, c, d \in \mathbb{R}^2 \) and let \( i \) be the index so that \( |b_i - c_i| = \max\{|b_1 - c_1|, |b_2 - c_2|\} \). Then

\[
    d_\infty(b, c) = |b_i - c_i| \leq |b_i - d_i| + |d_i - c_i| \leq d_\infty(b, d) + d_\infty(d, c),
\]

which is the triangle inequality. The first inequality comes from the triangle inequality in \( \mathbb{R}^1 \), and the second one comes from the definition of \( d_\infty \).

(e) The Hamming distance on strings is a metric space. For positive definiteness, observe that the function counts the number of bit differences, so it's positive, and two strings have zero differences if and only if they are the same string. Symmetry is fairly clear, so we show the triangle inequality since the number of differences between strings \( x \) and \( y \) cannot be more than the sum of the number of differences between strings \( x \) and \( z \) and between \( y \) and \( z \).

(f) Let \( X \) be a nonempty set, and define a distance by

\[
    d(x, y) = \begin{cases} 
0, & x = y \\
1, & x \neq y 
\end{cases}
\]

for \( x, y \in X \). Then \( (X, d) \) is a metric space.

Proof. Positive definiteness is clear from the definition as well as symmetry. For the triangle inequality, observe that \( d(x, z) \leq 1 \) for any \( x, z \in X \), so for any three points \( x, y, z \in X \), if either \( x \neq y \) or \( y \neq z \), then the inequality \( d(x, z) \leq d(x, y) + d(y, z) \) is satisfied. If it is the case that \( x = y = z \), then \( x = z \) and the inequality reduces to \( 0 \leq 0 \) which is still true. So the triangle inequality holds, and \( d \) is a metric on \( X \).
g. Consider $C[0,1]$ with the distance
\[ d_1(f, g) = \int_0^1 |f(x) - g(x)| \, dx. \]
Then this is a metric space.

Proof. Since the integrand is nonnegative, we see that this integral is 0 if and only if $f(x) = g(x)$ pretty much everywhere, and since the functions are continuous, this implies that $f(x) = g(x)$ on $[0,1]$. \[\textbf{Issues:} \text{How do we know that this integral is finite? Also can it happen that } f(x) = g(x) \text{ except at say, } x = \frac{1}{2} \text{ so that } f \neq g \text{ for instance?}\]
Symmetry is achieved by noting that $|f(x) - g(x)| = |g(x) - f(x)|$ for every $x \in [0,1]$.

Now let $f, g, h \in C[0,1]$. From the triangle inequality in $\mathbb{R}$, we have the inequality
\[ |f(x) - h(x)| \leq |f(x) - g(x)| + |g(x) - h(x)|. \]
We then integrate this inequality from 0 to 1 to get what we want:
\[ \int_0^1 |f(x) - h(x)| \, dx \leq \int_0^1 |f(x) - g(x)| \, dx + \int_0^1 |g(x) - h(x)| \, dx. \]

h. The set $C[0,1]$ is a metric space with metric
\[ d_\infty(f, g) = \max_{0 \leq x \leq 1} |f(x) - g(x)|. \]

Proof. First we state that the maximum is both finite and that it is achieved at some $c \in [0,1]$, that is, there is some $c \in [0,1]$ so that
\[ |f(c) - g(c)| = \max_{0 \leq x \leq 1} |f(x) - g(x)|. \]
\[\textbf{Issue:} \text{We will prove later that the maximum exists, is finite, and is achieved at some point in the interval.}\]
The function $d_\infty$ is nonnegative, and it’s clear that $d_\infty(f, f) = 0$ for any function $f$. If $d_\infty(f, g) > 0$ then $f$ and $g$ must differ at some point, so then $f \neq g$. Also $d_\infty$ is symmetric because $|f - g| = |g - f|$.

Let $f, g, h \in C[0,1]$. We establish the triangle inequality by using the $c$ described above. Then the triangle inequality in $\mathbb{R}^1$ implies
\[ |f(c) - g(c)| \leq |f(c) - h(c)| + |h(c) - g(c)|. \]
The values of $|f(x) - h(x)|$ and $|h(x) - g(x)|$ may not achieve their maxima at $c$, so this means that

$$|f(c) - h(c)| \leq \max_{x \in [0, 1]} |f(x) - h(x)|$$
and

$$|h(c) - g(c)| \leq \max_{y \in [0, 1]} |h(y) - g(y)|,$$

and the right halves of these inequalities are $d_\infty(f, h)$ and $d_\infty(h, g)$ respectively. Thus

$$d_\infty(f, g) \leq d_\infty(f, h) + d_\infty(h, g).$$

\[\Box\]

y. We can also create new metric spaces from old ones. Let $(X, d)$ and $(Y, \rho)$ be two metric spaces. If we define the function $\phi$ so that for any two points $(x_1, y_1), (x_2, y_2) \in X \times Y$,

$$\phi((x_1, y_1), (x_2, y_2)) = d(x_1, x_2) + \rho(y_1, y_2),$$

then the resulting space $(X \times Y, \phi)$ becomes a metric space.

**Idea of proof.** You should check this and provide some more details, but here’s a sketch. The properties of a metric space fit $\phi$ and follow directly from the properties of $d$ and $\rho$. To check this, start by writing the expression you want to manipulate in terms of $\phi$ and points of $X \times Y$. Then break it down into an expression involving $d$ and $\rho$ and points of $X$ and $Y$ respectively; apply the properties as they are given for $d$ and $\rho$, then restate the resulting expression in terms of $\phi$ and points of $X \times Y$.

z. Let $(X, d)$ be any metric space. Define $d' : X \times X \to \mathbb{R}$ by

$$d'(x, y) = \frac{d(x, y)}{1 + d(x, y)}. \quad (1)$$

Then $(X, d')$ is also a metric space! The proof for this is left to the exercises.

Sometimes having an example of an instance which doesn’t possess the properties desired can be as instructive or even more so than examples of things which follow the definition. These objects which follow are all not metric spaces. Often they fail in more than one of the three properties of a metric space, sometimes in all three.

**Example 2.3.** a. Consider $\mathbb{R}^2$ with a function $f$ so that

$$f((x_1, y_1), (x_2, y_2)) = \min\{|x_1 - x_2|, |y_1 - y_2|\}.$$  

Then $(\mathbb{R}^2, f)$ is not a metric space.
Proof. This function fails to be positive definite because the points $(3, 4) \neq (3, 7)$, but $f((3, 4), (3, 7)) = 0$. Thus this is not a metric space, even though $f$ is symmetric. It also fails in the triangle inequality: consider the three points $(0, 0), (0, 1)$, and $(1, 0)$.

b.

2.3 Open Sets

For the rest of this chapter, we will assume that $(X, d)$ is a metric space.

When we study functions $\mathbb{R} \rightarrow \mathbb{R}$, we like to consider functions which are “nice” in some way. That is, if we consider just an arbitrary function, it may not be as useful or significant as some other functions. For example, if we have some function $f : \mathbb{R} \rightarrow \mathbb{R}$, one condition we may like to have is that the metric on the domain meshes with the metric on the range. One way to express this is that for an $x$ in the domain, if $y$ is close to $x$, then $f(y)$ is close to $f(x)$. Another way is to say that for $y$ near $x$, $f(y) - f(x)$ is approximately linear with respect to $y - x$. In the first case we have that $f$ is continuous, and in the second that $f$ is differentiable.

So we have to say what it means for two points to be “close” or what “approximately” means to be able to really define the concepts of continuity or differentiability. We start with balls.

Definition 2.4. Let $x \in X$ and $r > 0$. Then the open ball of radius $r$ centered at $x$ is the set

$$B_r(x) = \{y \in X | d(x, y) < r\}.$$  

The closed ball of radius $r$ centered at $x$ is the set

$$\{y \in X | d(x, y) \leq r\}.$$

You can compare this with definition 8.9 in [1] or section 6.1 in [2].

So given some point $x \in X$ we can think that if $r < s$, then the aggregate of points in $B_r(x)$ is closer to $x$ than the points in $B_s(x)$, or that $B_r(x)$ is a tighter “neighborhood” of points around $x$ than $B_s(x)$.

Certainly if $r < s$ then $B_r(x) \subseteq B_s(x)$, but one should be cautious about assuming strict containment. In $\mathbb{R}$ with the usual metric, if $r < s$, then $B_r(x)$ contains some points outside of $B_r(x)$. However if we consider the infinite set $X$ with the discrete metric, then $B_\frac{1}{2}(x) = B_\frac{1}{2}(x) = \{x\}$. So we see that depending on the metric space, if two balls are defined with different radii, they still may be the same set.

Definition 2.5. Let $U \subseteq X$. Then we say $U$ is an open set if around any point $x \in U$ there is an open ball contained entirely within $U$. Precisely, for every $x \in U$, there is some $r > 0$ so that $B_r(x) \subseteq U$.

First, any open ball is an open set:
Proof. Let \( y \in B_r(x) \) and \( s = r - d(x, y) \). We claim that \( B_s(y) \subseteq B_r(x) \). Let \( z \in B_s(y) \), so then the triangle inequality implies
\[
d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + r - d(x, y) = r.
\]
Thus \( z \) is within a distance \( r \) of \( x \), and thus is contained within \( B_r(x) \).

Since there are open sets, the notation suggests that we may have closed sets. Indeed we do, and we will define them after we have encountered sequences and convergence.

Example 2.6. Here are some examples of open balls and sets.

a. Any open interval in \( \mathbb{R} \) under the usual metric is an open ball (and thus an open set): If \( U = (a, b) \), then \( U = B_r(\frac{a+b}{2}) \), where \( r = \frac{b-a}{2} \).

b. Any arbitrary union of open sets in a metric space is also open; the proof is left to an exercise.

c. Also, the empty set \( \emptyset \) is open vacuously since for every \( x \in \emptyset \) we can indeed find an open ball around \( x \) contained within \( \emptyset \). The issue is that we will never find some \( x \in \emptyset \). [Recall from logic that \( \sim P \Rightarrow (P \Rightarrow Q) \).]

d. The whole space \( X \) is open because any open ball is automatically contained within \( X \) by definition.

e. In the space \( C[0,1] \) with the \( d_\infty \) metric,
\[
B_1(0) = \{ f(x) \in C[0,1] : |f(x)| \leq 1 \text{ for } 0 \leq x < 1 \}.
\]

f. Let \( Y \) be a discrete metric space. Every subset of \( Y \) is open.

y. Any finite intersection of open sets is open.

Proof. Let \( U_1, U_2, \ldots, U_n \) be open sets in \( X \), and let \( U = \bigcap_{j=1}^n U_j \). Now consider \( x \in U \). Therefore \( x \in U_k \) for \( 1 \leq k \leq n \), so there are open balls \( B_{r_1}(x), B_{r_2}(x), \ldots, B_{r_n}(x) \) such that \( B_{r_k}(x) \subseteq U_k \). Now define
\[
r = \min\{r_1, r_2, \ldots, r_n\}.
\]
Then \( r > 0 \) since it is the minimum of a finite number of positive numbers. Therefore \( B_r(x) \subseteq B_{r_k}(x) \) for every \( k \), and thus \( B_r(x) \subseteq U_k \) for each \( k \). Hence, it is also contained in their intersection \( U \). Since \( x \) was an arbitrary element of \( U \), we can conclude that \( U \) is open.
z. An infinite intersection of open sets is not necessarily open. In $\mathbb{R}$ consider the open intervals $U_n = (-\frac{1}{n}, 1 + \frac{1}{n})$. Then $\bigcap_n U_n = [0, 1]$ since if $x < 0$ or $x > 1$, then it will be outside of some $U_n$ provided we make $n$ large enough. The interval $[0, 1]$ is not open since around either of the points 0 or 1 there are no open balls contained within $[0, 1]$.

We can’t use the reasoning of the previous example because in taking the minimum of the radii of an infinite number of balls, we can’t automatically conclude that the minimum $r$ is positive. [Issue: We actually need to use infimum here instead of minimum, because there may not be a minimum within the set. More on this later.]

**Definition 2.7.** Expanding on the idea of an open ball about a point $x$, we define a **neighborhood** of $x$ to be an open set containing the point $x$.

Note that this definition is of an open set, but we are thinking of an open set with respect to a particular point. We are building a language of what it means to “approach” a point. Also, in some places, neighborhood is defined differently, so that a neighborhood of a point $x$ is a set $U$ which contains an open ball which contains $x$. While this definition works similarly to our definition, our definition is simpler to use in many proofs. The common idea is that a neighborhood of a point contains everything in the metric space within some distance to the selected point.

Another concept we use is **interior** of a set. Loosely speaking, the interior of a set is all the points inside the set which are located away from the outside of the set; basically $x$ is on the interior of a set $A$ if the distance from the complement of $K$ is positive, i.e. all the points in some neighborhood of $x$ are contained within $A$.

**Definition 2.8.** Let $A$ be a subset of a metric space $X$. The point $x \in A$ is called an **interior point** of $A$ if there is some open ball $B_r(x) \subseteq A$. The **interior** of $A$ is the collection of all interior points of $A$ and is denoted by $A^\circ$ or $\text{int}(A)$.

The point $y$ is called an **exterior point** of $A$ if there is some open ball around $y$ so that $B_r(y) \subseteq X \setminus A$. The collection of exterior points of $A$ is called the **exterior** of $A$ and is denoted by $\text{ext}(A)$.

A point $z$ is called a **boundary point** of $A$ if every neighborhood of $z$ contains points of $A$ and of $X \setminus A$. The collection of boundary points of $A$ is its **boundary** and is sometimes denoted by $\text{bd}(A)$ or $\partial A$.

For any set $A$, $A^\circ$ is an open set, and for some sets, $A^\circ$ may even be empty.

**Proof that $A^\circ$ is open.** Let $x \in A^\circ$. Then there is some $B_r(x) \subseteq A$. Now we show that $B_r(x)$ is also contained in $A^\circ$, which will prove that $A^\circ$ is open.

Consider $y \in B_r(x)$, and let $s = r - d(x, y)$. Then $B_s(y) \subseteq B_r(x) \subseteq A$, which shows that $y \in A^\circ$ as well. Since $y$ was an arbitrary element of $B_r(x)$, this implies that $B_r(x) \subseteq A^\circ$. Since for an arbitrary element $x \in A^\circ$ we have found an open ball around $x$ contained within $A^\circ$, this proves that $A^\circ$ is open. □
Note: The interior of a set depends on the metric space it lives in. In particular, it depends on how the open balls are defined.

**Example 2.9.** These are some examples of how interiors, exteriors, and boundaries behave.

a. Let \( A = [a, b] \) be a real interval. Then \( A^\circ = (a, b) \).

b. Let \( C \) be a finite subset of \( \mathbb{R}^n \) with the euclidean metric. Then \( C^\circ = \emptyset \).

c. Let \( V = [\pi - 1, \pi + 1] \cap \mathbb{Q} \) be a subset of the metric space \( \mathbb{Q} \). Then \( V^\circ = V \), namely because \( [\pi - 1, \pi + 1] \cap \mathbb{Q} = (\pi - 1, \pi + 1) \cap \mathbb{Q} \). Note that \( \partial V = \emptyset \).

d. Let \( K \) be any subset of a discrete metric space. Then \( K^\circ = K \).

e. The exterior of \( S \) is the interior of \( X \setminus S \). The proof is left to the exercises.

f. Given any \( B \subseteq X \), \( B^\circ \), \( \partial B \), and \( \text{ext}(B) \) are disjoint sets, and \( X = B^\circ \cup \partial B \cup \text{ext}(B) \).

g. Let \( U = \{ x \in \mathbb{R}^2 : 0 < d(0, x) < 1 \} \) be the punctured unit disk in \( \mathbb{R}^2 \). Then \( \partial U = \{ x \in \mathbb{R}^2 : d(0, x) = 1 \} \cup \{ 0 \} \), i.e. the boundary of \( U \) is the unit circle along with the origin.

h. If \( U \) is open, then \( U^\circ = U \).

i. Let \( A \) be a set. Then \( A^\circ \) is the union of all open sets contained in \( A \). Notationally,

\[
A^\circ = \bigcup_{U \text{ is open}} U.
\]

We leave the proof to the exercises.

j. Let \( U \) be any subset of a discrete metric space \( X \), then \( U^\circ = U \) and \( \partial U = \emptyset \). This occurs since any subset of a discrete space is open.

k. Let \( A = \{ 1, 1/2, 1/3, 1/4, \ldots \} \cup \{ 0 \} \subseteq \mathbb{R} \). Then \( A^\circ = \emptyset \) and \( \partial A = A \).

y. Even if a set is infinite, it may have empty interior. Consider \( \mathbb{Q} \subseteq \mathbb{R} \) the set of rational numbers; \( \mathbb{Q}^\circ = \emptyset \). We can show this because there is an irrational number between any two rational numbers, so there can be no open ball in \( \mathbb{R} \) contained within \( \mathbb{Q} \) \[\textbf{Issue: We need to prove the fact that between any two rational numbers is an irrational. Similarly, between any two irrationals is a rational.}\].

z. If we consider \( \mathbb{Q} \) as the metric space, then open balls are defined to be within \( \mathbb{Q} \), and in this situation, perhaps surprisingly, \( Q^\circ = \mathbb{Q} \).
These last two examples illustrate the need to be specific about what metric space we’re working in and a set’s relationship to the metric space. Whenever we speak about a set’s interior, either we will specify what space we’re in, or the space will be understood from context. The set \( \mathbb{Q} \) is a subset of \( \mathbb{R} \), and the distances between rational points in \( \mathbb{Q} \) is the same as the distances between the corresponding points in \( \mathbb{R} \). We can think of \( \mathbb{Q} \) as a metric subspace of \( \mathbb{R} \), then.

**Definition 2.10.** Let \((X, d)\) be a metric space and \(Y \subseteq X\). Let \(d'\) be the function \(d\) restricted to \(Y \times Y\) so that for \(x, y \in Y\) means \(d'(x, y) = d(x, y)\). Then we say that \((Y, d')\) is a **metric subspace** or **subspace** of \((X, d)\).

A set \(V \subseteq Y\) is open in \(Y\) if and only if there is an open set \(U \subseteq X\) so that \(V = U \cap Y\). We sometimes say that \(V\) is relatively open in \(Y\).

If the containment \(Y \subseteq X\) is proper, that is \(Y \subset X\), then we say \(Y\) is a **proper subspace** of \(X\).

**Remark 2.11.** In many of the definitions in this section, we mention only metric spaces even though we apply the definition to subsets of metric spaces. Fortunately, we consider subsets of metric spaces as metric spaces in their own right (via above), so we are ok in the applications.

**Example 2.12.**

a. We have already seen that \(\mathbb{Q}\) is a subspace of \(\mathbb{R}\) using the standard metrics.

b. If we consider \(\mathbb{R}^2\) under any of the metrics we have already described and identify \(\mathbb{R}\) with the x-axis, then \(\mathbb{R}\) is a metric subspace of \(\mathbb{R}^2\).

c. Consider \(\mathbb{R}^n\) with the standard euclidean \(d_2\) metric. Let

\[
A = \left\{ \left( \frac{1}{\sqrt{2}}, 0, 0, \ldots, 0 \right), \left( 0, \frac{1}{\sqrt{2}}, 0, \ldots, 0 \right), \ldots, \left( 0, 0, \ldots, 0, \frac{1}{\sqrt{2}} \right) \right\}.
\]

Then \(A\) is a finite subset of \(\mathbb{R}^n\) consisting of \(n\) points, and the metric on \(A\) inherited from \(\mathbb{R}^n\) is the same as the discrete metric. (you should check this)

d. Consider \(A = [-1, 1] \subseteq \mathbb{R}\), under the usual metric. Then the set \([-1, 0)\) is not open in \(\mathbb{R}\), but it is relatively open in \(A\) since \([-1, 0) = (-2, 0) \cap A\). By the same token \([-1, 0)\) is open in \(A\) in the subspace metric.

If you have trouble seeing this, consider that in \(A\), \(B_2^A(-1) = [-1, -\frac{1}{2})\).

**Definition 2.13.** Let \(A\) be a subset of a metric space \(X\). Then we say \(A\) is **bounded** if it is contained within some open ball in \(X\). Precisely, \(A\) is bounded if there is an \(x \in X\) and \(r > 0\) so that \(A \subseteq B_r(x)\). We say that \(A\) is **unbounded** if there is no such ball.

If a metric space is contained within an open ball, then we say it is a **bounded metric space**.

**Example 2.14.**

a. Let \(a, b \in \mathbb{R}\) with \(a < b\). Then \([a, b]\) and \((a, b)\) are both bounded, \((-\infty, a)\) and \((a, \infty)\) are unbounded, and \(\mathbb{R}\) is unbounded. The set \(\mathbb{Q}\) is also unbounded.
b. If \( M \) is bounded, then for any set \( K \), \( M \cap K \) is bounded, since a ball which contains \( M \) also contains \( M \cap K \).

c. The intersection of two unbounded sets could be bounded. If \( a < b \) in \( \mathbb{R} \), then \( (-\infty, b] \) and \([a, \infty) \) are both unbounded, but their intersection \([a, b] \) is bounded.

d. Any ball is bounded, whether it is open or closed. The empty set is bounded because it is a subset of any set; in particular it is a subset of a ball.

e. A finite union of bounded sets is bounded.

Proof. Let \( A_1, A_2, \ldots, A_k \) be a finite collection of bounded sets in metric space \( X \). We proceed by induction. A union of one bounded set is bounded, since \( \bigcup_{n=1}^{1} A_n = A_1 \). Now if \( A_1 \) and \( A_2 \) are bounded, then they are contained in some \( B_{r_1}(x_1) \) and \( B_{r_2}(x_2) \) respectively. Let \( r = \max\{r_1, d(x_1, x_2) + r_2\} \). Let \( x \in A_1 \cup A_2 \). If \( x \in A_1 \), then \( d(x_1, x) < r_1 \leq r \). If \( x \in A_2 \), then \( d(x_1, x) \leq d(x_1, x_2) + d(x_2, x) < d(x_1, x_2) + r_2 \leq r \). Therefore \( d(x_1, x) < r \). Thus \( A_1 \cup A_2 \subseteq B_r(x_1) \), hence it is bounded.

Now suppose every collection of \( n \) bounded sets has a bounded union. Now let \( \{A_1, A_2, \ldots, A_{n+1}\} \) be a collection of bounded sets. By assumption, \( A = \bigcup_{j=1}^{n} A_j \) is bounded, and by what we showed, \( A \cup A_{n+1} \) is bounded. Therefore \( \bigcup_{j=1}^{n+1} A_j \) is bounded.

j. Let \( D \) be any nonempty discrete metric space. Then \( D \) is bounded. In particular, if \( x \in D \), then \( D \subseteq B_2(x) \). Since this ball by definition is contained within \( D \), it follows that \( D = B_2(x) \).

z. If we go back to the initial examples of metric spaces to the bounded metric (1), then \((X, d')\) is a bounded metric space.

### 2.4 Sequences, Limits, and Closed Sets

We are almost in a position to say what it means for a collection of points to approach a point \( x \). We just have to give a set a sense of direction, or some order. Fortunately we have a natural setting for order in \( \mathbb{N} \).
Definition 2.15. Let $Y$ be a set. A sequence is a subset $\{a_1, a_2, \ldots\}$ of a metric space $X$ along with an assignment $n \mapsto a_n$ for each $n \in \mathbb{N}$. Alternatively, we can think of a sequence as a function $f : \mathbb{N} \to Y$ by $f(n) = a_n$. We say that $a_n$ is the $n^{th}$ term of the sequence $\{a_n\}_{n \geq 1}$.

Sometimes we start a sequence with $n = 0$ instead of $n = 1$.

We also have the notion of a subsequence of a sequence; a sequence which is part of the original sequence but retains essential parts like the “approaching-ness” of the original sequence.

Definition 2.16. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence. Let $i(k)$ be a function $\mathbb{N} \to \mathbb{N}$ so that $j < k \Rightarrow i(j) < i(k)$. Then we say that that $\{a_{i(k)}\}_{k=1}^{\infty}$ is a subsequence of $\{a_n\}$. If there is some $k$ so that $i(k) > k$, i.e. that $i(k)$ skips some numbers, then we say that $\{a_{i(k)}\}_{k=1}^{\infty}$ is a proper subsequence of $\{a_n\}$. Sometimes we will use $i_k$ instead of $i(k)$, and we may write $\{a_{i_k}\}_{k \geq 1}$ for the subsequence.

Definition 2.17. Let $c_k$ be a sequence in a metric space $X$. Then the sequence is bounded if the set $\{c_k\}_{k \geq 1}$ is a bounded set and unbounded if $\{c_k\}_{k \geq 1}$ is an unbounded set.

Note that if a sequence is bounded, then any subsequence is also bounded.

Example 2.18.  

a. The sequence
\[
1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots
\]
is a sequence in $\mathbb{R}$ with $a_n = \frac{1}{n}$. The following are all subsequences:
\[
1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \frac{1}{9}, \ldots
\]
\[
1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \ldots
\]
\[
1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots
\]
with $i(n) = 2n - 1$, $i(n) = n^2$, and $i(n) = 2^n$ respectively.

b. The function $b_n = x^{n-1}$ is a sequence in $C[0, 1]$:
\[
1, x, x^2, x^3, x^4, \ldots
\]
If we consider $d_\infty(0, b_n) = 1$, then this sequence is a bounded sequence with this metric.

If we take the same assignment, but consider functions in $C[0, 2]$, then this sequence is an unbounded sequence using the $d_\infty$ metric. Namely $d_\infty(0, b_n) = 2^{n-1}$ in this metric space.

c. The sequence $1, 1, 2, 3, 5, 8, 13, 21, \ldots$ is known as the Fibonacci sequence and is defined recursively by $F_1 = 1$, $F_2 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 3$. This is an unbounded sequence in $\mathbb{R}$. 

19
The sequence $-1, 1, -1, 1, -1, \ldots$ so that $a_n = (-1)^n$ is a bounded real sequence.

We now have the language to say what convergence means for sequences. Let’s look at what we want. When a sequence $\{x_n\}$ approaches a point $x$, the elements of the sequence “cluster up” on $x$. If we go further along the sequence, the clustering gets tighter around $x$. In fact, if we go far enough, we could get as tight a clustering as we want.

Another way to say this is that given any tolerance, no matter how small, eventually everything in the sequence past a certain point will be within this tolerance. Then using this same property, we could choose a much tighter tolerance than we chose before, and once we get past some point in the sequence (it may be much, much further along this time), then everything beyond that point will be close to $x$ within that tolerance. We can repeat this procedure ad nauseam to get a system of ever tightening tolerances (ever shrinking balls $B_\varepsilon(x)$) around $x$. This is how we can say that a sequence approaches $x$. We now make this precise.

**Definition 2.19.** Let $\{x_n\}$ be a sequence in metric space $X$. Then for some point $x \in X$, we say

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

if for every $\varepsilon > 0$, there is some $N \in \mathbb{N}$ so that if $n \geq N$, then $d(x_n, x) < \varepsilon$.

If $\lim_{n \to \infty} x_n = x$, equivalent terminology includes:

- $x_n \to x$,
- $x_n$ converges to $x$,
- $x_n$ approaches $x$,
- $x_n \xrightarrow{n} x$, or
- the limit of $x_n$ is $x$.

**Example 2.20.** a. The real sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ converges to 0.

**Proof.** We note that $a_n = \frac{1}{n}$. Let $\varepsilon > 0$. Pick $N > \frac{1}{\varepsilon}$. Then if $n \geq N$, then $n > \frac{1}{\varepsilon} > 0$, and then $\varepsilon > \frac{1}{n} > 0$. Thus $|a_n - 0| < \varepsilon$. Therefore $\lim_{n \to \infty} a_n = 0$. □
[Issue: How do we know there is some \( N > \frac{1}{\varepsilon} \)? The quick answer is that \( \mathbb{R} \) has the archimedean property. We will prove this.]

b. Any constant sequence \( a, a, a, \ldots \) converges to \( a \).

c. A sequence converges to at most one limit.

Proof. Let \( \{s_n\} \) be a sequence in a metric space \( Z \), and suppose \( s_n \) converges to \( s \). Suppose it also converges to \( t \in Z \). We show \( s = t \) by showing \( d(s, t) = 0 \).

Let \( \varepsilon > 0 \). Let \( N_1 \in \mathbb{N} \) be so that \( n \geq N_1 \Rightarrow d(s_n, s) < \frac{\varepsilon}{2} \). Let \( N_2 \in \mathbb{N} \) be so that \( n \geq N_2 \Rightarrow d(s_n, t) < \frac{\varepsilon}{2} \). Let \( N = \max\{N_1, N_2\} \), so that \( n \geq N \) satisfies both conditions. Let \( n \geq N \). Then the triangle inequality implies

\[
d(s, t) \leq d(s, s_n) + d(s_n, t) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\]

Thus \( d(s, t) < \varepsilon \). Since \( \varepsilon \) was an arbitrary positive number, this implies that \( d(s, t) \) must be smaller than any positive number. This forces \( d(s, t) = 0 \), and thus \( s = t \).

This is a property of metric spaces. When studying general topological spaces, there are instances where a sequence converges to multiple limits. We may get to some examples of this in this course depending on how much time we have.

d. Any convergent sequence is bounded.

Proof. Consider a sequence \( \{x_n\} \subseteq X \) which converges to \( x \). Let \( \varepsilon > 0 \), and let \( N \) be a positive integer as in the definition of convergence. Then \( \{x_n, x_{N+1}, x_{N+2}, \ldots\} \subseteq B_\varepsilon(x) \), and since \( \{x_1, \ldots, x_{N-1}\} \) is finite, it is bounded, hence contained in some ball \( B_\alpha(y) \). We have expressed the sequence as a union of two bounded sets; therefore it is bounded.

e. Let \( b_n \) be a sequence in a discrete metric space \( B \). Then \( b_n \) converges if and only if it is eventually constant, that is there is some \( b \in B \) and \( N \in \mathbb{N} \) so that \( n \geq N \Rightarrow b_n = b \). In this case \( \lim_{n \to \infty} b_n = b \).

Quick proof. Let \( \varepsilon = \frac{1}{2} \) and \( N \) so that \( n \geq N \) implies \( d(b_n, b) < \frac{1}{2} \). But \( B \) is a discrete metric space, so the only point within a distance of \( \frac{1}{2} \) of \( b \) is \( b \) itself. Therefore, \( b_n = b \).

And if \( b_n = b \) for all \( n \) greater than some particular \( N \), then certainly \( b_n \to b \).

z. Let \( f_n(x) = 1 - e^{-nx} \), for \( n \geq 1 \), be a sequence of functions in \( C[0, 7] \). Then \( f_n(x) \to f(x) \equiv 1 \) under the \( d_1 \) metric.
Proof. We have to show that \( d(1, f_n) \to 0 \) as \( n \to \infty \). So, let \( \varepsilon > 0 \). For any \( n \in \mathbb{N} \),

\[
d_1(1, f_n) = \int_0^7 |1 - (1 - e^{-nx})| \, dx
\]

\[
= \int_0^7 e^{-nx} \, dx
\]

\[
= -\frac{e^{-nx}}{n} \bigg|_0^7
\]

\[
= \frac{1}{n} - \frac{e^{-7n}}{n}
\]

\[
< \frac{1}{n}.
\]

So let’s pick \( N > \frac{1}{\varepsilon} \). Then for \( n \geq N \), \( d_1(1, f_n) < \frac{1}{n} < \varepsilon \). Therefore \( f_n \to 1 \) under \( d_1 \).

Note that \( \{f_n\} \) converges to \( f(x) = 1 \) even though not all the points converge to 1. In particular, \( f_n(0) = 0 \) for every \( n \). This illustrates that convergence in one sense does not necessarily mean convergence in another sense.

An important fact is that a subsequence inherits a lot of properties of a sequence. In particular, if a sequence converges to a limit then the subsequence converges to the same limit.

**Theorem 2.21.** Let \( x_n \to x \) as \( n \to \infty \) in a metric space \( X \), and let \( \{x_{n_k}\}_{k \geq 1} \) be a subsequence. Then \( x_{n_k} \) converges to \( x \) as well.

**Proof.** Let \( x_n \to x \) and let \( \varepsilon > 0 \). By the convergence of the series, there exists \( N \in \mathbb{N} \) so that \( n \geq N \) implies that \( d(x_n, x) < \varepsilon \). Consider the subsequence \( \{x_{n_k}\} \). If \( k \geq N \), then \( n_k \geq n_N \geq N \), so \( d(x_{n_k}, x) < \varepsilon \). Thus we have shown that the same \( N \) works for the subsequence \( \{x_{n_k}\} \) as is does for the sequence \( \{x_n\} \). Hence \( x_{n_k} \xrightarrow{k} x \).

We have showed that any \( N \) we found for the original sequence also works for the subsequence. In principle, there may be a smaller \( M \) so that \( k \geq M \) implies \( d(x_{n_k}, x) < \varepsilon \).

We have seen that an open set doesn’t include its boundary, but in a sense the boundary is “attached” to the open set in that if \( x \in \partial A \), then every open set containing \( x \) also contains a point of \( A \) as well. We can build a sequence in \( A \) which converges to \( x \) by picking \( a_n \in A \cap B_{\frac{1}{n}}(x) \). Then it’s easy to show that \( a_n \to x \).

We would like to further study the different way points are “attached” to a set \( A \).
Definition 2.22. Let $A \subseteq X$. Then $x \in X$ is called a **cluster point** or **point of accumulation** of $A$ if every neighborhood of $x$ contains a point of $A$ other than $x$.

An **isolated point** of $A$ is a point $a \in A$ so that there is some neighborhood $U$ of $a$ so that $U \cap A = \{a\}$.

Sometimes cluster points are called **limit points**.

Loosely speaking, in a metric space, a cluster point of a set $A$ is a point which is close to infinitely many points of $A$. A cluster point doesn’t necessarily belong to $A$; it could belong to $X \setminus A$. An isolated point belongs to $A$, but it is set apart from the rest of $A$ by some amount; the gap could be extremely small, but there is some positive distance. Note that when speaking of cluster points of a set $A$, we are considering the metric space $X$ in which $A$ resides.

Example 2.23. 

a. In the metric space $\mathbb{R}$, let $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$ and let $B = A \cup \{0\}$. Then 0 is a cluster point of $A$ since any neighborhood of 0 contains some ball $B_r(0)$ which contains $1_n \in A$, where $n > \frac{1}{r}$. By the same reasoning, 0 is a cluster point of $B$.

Each of the points $\frac{1}{n}$ for $n = 1, 2, 3, \ldots$ is an isolated point of both $A$. We can see this since we claim the interval $(\frac{1}{n} - \frac{1}{n(n+1)}, \frac{1}{n} + \frac{1}{n(n+1)})$ is an open set around $\frac{1}{n}$ which contains no other points of $A$. To show this, we just observe that $\frac{1}{n} - \frac{1}{n(n+1)} = \frac{1}{n+1}$, and for $n \geq 2$, $\frac{1}{n} + \frac{1}{n(n+1)} < \frac{1}{n-1}$. By the same reasoning, these points are isolated points of $B$ as well.

Note that a point can be an isolated point of a set even if we use it to show that another point is a cluster point. That’s because the limit properties of a sequence $a_n$ happen as $n \to \infty$. If we change or discard some finite number of elements of the sequence, the properties as $n \to \infty$ still remain the same (that is we can still take $N$ to be beyond our changes).

b. The set of cluster points of the open ball $B_1(0)$ in $\mathbb{R}^2$ is the closed ball \( \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\} \).

c. Let $A$ be a subset of a discrete metric space $X$. Then $A$ has no cluster points, and in particular every point of of $A$ is an isolated point. This is because for any $x \in X$, the ball $B_{1/2}(x)$ contains only the point $x$.

d. If $A \subseteq B$, then every cluster point of $A$ is also a cluster point of $B$. Let $x$ be a cluster point of $A$ and $U$ a neighborhood of $x$. Then there is some point $a \in A$ such that $a \neq x$ and $a \in U$. But since $A \subseteq B$, we have $a \in B$ and $a \neq x$. Thus $x$ is a cluster point of $B$.

e. Let $A$ be a subset of metric space $X$, and let $x$ be a cluster point of $A$. Then every neighborhood of $x$ contains an infinite number of points of $A$. The proof is left for exercise 8.

Definition 2.24. If a set $F$ contains all its cluster points, then we say $F$ is a **closed set**.
Additionally, we have the further characterization:

**Theorem 2.25.** A set is closed if and only if its complement is open.

*Proof.* Suppose $F \subseteq X$ is closed. Let $x \in X \setminus F$. Since $F$ contains all its cluster points, there is some ball $B_r(x)$ such that $B_r(x) \setminus \{x\}$ contains no points of $F$. Since $x \notin F$, this shows that $B_r(x)$ is disjoint from $F$. Since $x$ was an arbitrary point of $X \setminus F$, this implies that $X \setminus F$ is open.

Conversely, suppose that $X \setminus F$ is open. Pick $x \notin F$. There exists some open ball $B_r(x)$ contained within $X \setminus F$. Therefore, in this neighborhood of $x$, there are no elements of $F$, so it cannot be a cluster point. Since $x$ was an arbitrary point of $X \setminus F$, we conclude that all cluster points of $F$ must be contained in $F$, hence $F$ is closed.

When we discuss general topological spaces and concepts of nearness and convergence when a metric is not necessarily present, we will take the definition of a closed set to be the complement of an open set. In proofs, sometimes the easiest way to show a set is closed is to show its complement is open. Similarly, in other situations, the easiest way to show a set is open is to show its complement is closed.

The words open and closed seem to be evocative of a door or other object you can open and close, and the fact that closed sets and open sets are complements seem to support this, but one must be careful about taking linguistic analogies too far. A set could be open, closed, both, or neither. We provide some examples.

**Example 2.26.**

a. Let $a < b$ in $\mathbb{R}$. Then the sets $[a, b]$, $(a, \infty)$, and $(\infty, a]$ are closed.

b. Any closed ball $C = \{y \in X : d(x, y) \leq r\}$ around $x \in X$ is closed. We can show this by showing that any point in its complement cannot be an accumulation point. If $z \in X \setminus C$, then $d(x, z) > r$. Let $s = r - d(x, z)$, and consider $w \in B_s(z)$. Then $d(w, z) < s$, and the triangle inequality implies $d(x, z) \leq d(x, w) + d(w, z)$, so $d(x, z) - d(w, z) \leq d(x, w)$. We obtain $d(x, z) - s < d(x, w)$, so $r < d(x, w)$, which shows that $w \notin C$. Thus $B_r(z) \subseteq X \setminus C$. Since $z$ was an arbitrary point of the complement of $C$, this shows that $X \setminus C$ is open; therefore $C$ is closed.

Note that essentially the same argument shows that $z$ is not an accumulation point of $C$, so $C$ must contain all its accumulation points.

c. The empty set $\emptyset$ is closed since it has no accumulation points, so vacuously the empty set contains the set of accumulation points.

d. The entire metric space $X$ is closed in $X$ since any accumulation points must lie within it. So automatically $X$ and $\emptyset$ are both closed and open.

e. The interval $[0, 1)$ is neither open nor closed in $\mathbb{R}$ since
1. the point 1 is a cluster point of the interval but is not contained in it (hence it is not closed), and
2. every neighborhood of 0 contains some points outside of the interval (hence it is not open).

f. Any subset of a discrete space is closed, since no set has cluster points.

g. Any finite subset \( F \) of a metric space is closed, since a finite set contains no accumulation points in a metric space. Around any point \( x \in X \), we can take a ball small enough to contain either exactly one or zero points of \( F \).

h. An intersection of any number of closed sets is closed. The proof is left for exercise 7 at the end of the chapter.

We’ve seen that the interior of \( A \) is the largest open set contained in \( A \) in that \( A^\circ \) is an open subset of \( A \) so that if \( U \subseteq A \) is open then \( U \subseteq A^\circ \), too. We can define a similar property involving closed sets.

**Definition 2.27.** Let \( B \) be a subset of a metric space \( X \). Let the **closure** of \( B \) be the set \( \overline{B} \) which is the union of \( B \) with its cluster points.

In a similar way, \( \overline{B} \) is the smallest closed set containing \( B \), because if \( C \supseteq B \) is closed, then \( \overline{B} \subseteq C \), too.

**Theorem 2.28.** An immediate result is that for any set \( A \), \( \overline{A} \) is a closed set.

**Proof.** We show that all cluster points of \( \overline{A} \) are also cluster points of \( A \), which will imply that \( \overline{A} \) contains all its cluster points. So let \( x \) be a cluster point of \( \overline{A} \) and let \( \varepsilon > 0 \). We aim to show that there is a point \( a \in A \) so that \( a \in B_\varepsilon(x) \). Since \( x \) is a cluster point of \( \overline{A} \), there is some \( y \in \overline{A} \) with \( y \neq x \) such that \( d(x, y) < \frac{\varepsilon}{2} \). Now either \( y \in A \) or \( y \) is a cluster point of \( A \). In the first case, \( y \in A \) means that \( d(y, x) < \varepsilon \), and we’re done. In the second case, since \( y \) is a cluster point of \( A \), there is some \( a \in A \) with \( a \neq y \) such that \( d(y, a) < \frac{\varepsilon}{2} \).

We also may take \( a \neq x \) since every neighborhood of \( y \) contains infinitely many points of \( A \) by exercise 8. The triangle inequality gives

\[
    d(x, a) \leq d(x, y) + d(y, a) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\]

Since \( x \) is a cluster point of \( A \), this implies that \( x \in \overline{A} \). Therefore \( \overline{A} \) contains all its cluster points and is closed. \( \square \)

This result also establishes that \( \overline{\overline{A}} = \overline{A} \).

**Definition 2.29.** Let \( A \subseteq B \) be two subsets of a metric space \( X \). Then we say that \( A \) is **dense** in \( B \) if \( \overline{A} = \overline{B} \).
This means that in a sense, the set \( A \) “fills up” the set \( B \). The set \( B \) has no nooks and crannies which are not reachable by the set \( A \).

In practice, we most often take \( B \) to be closed in \( X \) so that you often see the statement: \( A \) is dense in \( B \) if \( \overline{A} = B \). In our case, this is a slightly more general definition, but it works in all the important places.

**Example 2.30.**

a. Let \( K \) be any subset of a discrete metric space. Then \( \overline{K} = K \).

b. Let \( A = [a, b] \cap \mathbb{Q} \) be a subset of \( \mathbb{R} \). Then \( \overline{A} = [a, b] \).

c. Let \( A \) be as above, but consider it as a subset of \( \mathbb{Q} \). Then \( \overline{A} = A \) because if we have any sequence within \( A \) which converges within \( \mathbb{Q} \) also converges within \( A \).

d. Let \( A \) be a subset of a metric space \( X \). Then \( \overline{A} = X \setminus (X \setminus A) \).

**Proof.** (\( \subseteq \)). Let \( b \in \overline{A} \). Then we must show that \( b \) is not in \( (X \setminus A)^c \). Suppose for contradiction that \( b \in (X \setminus A)^c \); then from the definition of interior, there is some open ball \( B_r(b) \subseteq X \setminus A \). Thus in the neighborhood \( B_r(b) \), there are no points of \( A \). But that means that in particular, \( b \notin A \), and there are no points of \( A \) in this ball, so \( b \) is not a cluster point of \( A \). Then from the definition of closure, \( b \notin \overline{A} \), which contradicts our first assumption. Therefore, we must have that \( b \in X \setminus (X \setminus A)^c \).

(\( \supseteq \)). Let \( b \in X \setminus (X \setminus A)^c \). Then \( b \) is not an interior point of \( X \setminus A \), so for every neighborhood \( B \) of \( b \), there is a point \( x \in A \cap B \). Now if it is the case that \( b \in A \), then this implies that \( b \in \overline{A} \). If \( b \notin A \), then this shows that every neighborhood of \( b \) contains a point of \( A \) which is not \( b \), so then \( b \) is a limit point of \( A \). In either case, \( b \in \overline{A} \).

In the first paragraph we established \( \overline{A} \subseteq X \setminus (X \setminus A)^c \), and in the second we proved \( \overline{A} \supseteq X \setminus (X \setminus A)^c \). Putting these together, we arrive at our conclusion. \( \square \)

**Picture!**

e. Similarly, \( A^c = X \setminus (X \setminus A) \). The proof is left to exercise 11.

f. Since \( \text{ext}(A) = (X \setminus A)^c \), it follows that \( \overline{A} = A^c \cup \partial A \).

g. In \( \mathbb{R} \) with the usual metric,

- \( \mathbb{Q} \) is dense in \( \mathbb{R} \),
- \( (a, b) \) is dense in \( [a, b] \)
- \( [0, 1] \cup (1, 2] \) is dense in \( [0, 2] \),
- \( \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\} \) is dense in \( \{0\} \cup \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\} \)
2.5 Cauchy sequences and Completeness

We see that a sequence converges if there’s a point to which the sequence gets arbitrarily close. Depending on the space, even if the the points accumulate in a particular space, there may not be a point to which they converge.

Example 2.31. Consider $X = \mathbb{R}$ with the usual metric and $A = (0,2) \subseteq \mathbb{R}$. Let $a_n = \frac{1}{n}$, $n \geq 1$ be a sequence. We’ve already shown that $a_n \to 0$ in $\mathbb{R}$, but if we consider the same sequence in the subspace $A$, it doesn’t converge. Therefore, whether a sequence converges depends as much on the space it’s in as the elements of the sequence itself.

There are many examples of such spaces: $\mathbb{Q}$, $\mathbb{R}^2 \setminus (0,0)$, $B_1(0)$ in $\mathbb{R}^n$, and $\mathbb{R}$ with the metric $d(x,y) = |\tan^{-1}(x) - \tan^{-1}(y)|$ are some.

This begs the question, how do we say that a sequence is heading somewhere using only the elements of the sequence? How do we say that there ought to be a limit without specifying what that limit is?

Let’s examine the elements of a convergent sequence to discern what property we need. Let $x_n \to x$ in $X$ and $\varepsilon > 0$. From the definition of convergence, there is some $N \in \mathbb{N}$ so that $n \geq N$ implies $d(x_n, x) < \varepsilon$. So everything past $N$ is contained in the ball $B_\varepsilon(x)$. Now let’s take two elements past $N$ by letting $m, n \geq N$ and see how they relate. The triangle inequality gives

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \varepsilon + \varepsilon = 2\varepsilon.$$ 

Thus every element past $N$ is contained within the ball $B_{2\varepsilon}(x_N)$. By replacing $\varepsilon$ with $\frac{\varepsilon}{2}$ in the above argument, we arrive at the definition Cauchy did in the 19th century.

Definition 2.32. Let $x_n$ be a sequence in a metric space $X$. Then we say $x_n$ is Cauchy or that it is a Cauchy sequence if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ so that $m,n \geq N$ implies $d(x_m, x_n) < \varepsilon$.

We have seen above that every convergent sequence is Cauchy, and we also have examples of spaces in which some Cauchy sequences don’t converge. We would like to say when every Cauchy sequence converges.

Definition 2.33. Let $X$ be a metric space. We say that $X$ is complete if every Cauchy sequence converges in $X$.

Example 2.34. a. Our archetypical example of a metric space, $\mathbb{R}$ is complete. It would be a little disturbing if it weren’t. We will prove this fact later. Also, $\mathbb{R}^n$ is complete for any $n \in \mathbb{N}$.

b. Any discrete metric space $Y$ is complete. This is because any Cauchy sequence is eventually constant, and hence convergent.

c. The metric space $\mathbb{Q}$ is not complete. The Cauchy sequence

$$3, 3.1, 3.14, 3.141, 3.1415, 3.14159, \ldots$$

doesn’t converge in $\mathbb{Q}$. (Exercise: verify that this sequence is Cauchy)
d. The open ball $B_1(0) \subseteq \mathbb{R}^2$ is not complete because it doesn’t contain its boundary, i.e. we can create a Cauchy sequence in $B_1(0)$ which converges to a point on its boundary, which is not contained in the set.

Since every closed subset $F$ of a metric space $X$ contains all accumulation points of $F$, it follows that any closed subset of a complete metric space is complete. Similarly, a complete subset of any metric space is closed.

Notice that although closed sets and complete sets are related by the observation above, one significant difference is that closed sets are determined by which metric space is their universe. Completeness, however, is intrinsic to the set, and a set may be determined to be complete or not regardless if we consider it as a subspace of something else or as a space on its own.

This brings us to another question. We are able to identify sequences which “ought” to converge, namely Cauchy sequences, so given a metric space $X$ which isn’t complete, how can we “throw in” all the limits of Cauchy sequences to make it complete? If $X$ is a subspace of a complete space $Y$, then it’s pretty easy - just take the closure of $X$ in $Y$. However, a priori, we are not guaranteed such a space $Y$ to even exist, so how do we do it when we only have $X$?

We do have one important set, the set of all Cauchy sequences. Since these are the sequences we would like to converge, let’s build a space on them.

Let $X$ be a metric space, and $\{x_n\}$ and $\{y_n\}$ be two Cauchy sequences in $X$. Then we say $\{x_n\}$ and $\{y_n\}$ are equivalent, or $\{x_n\} \sim \{y_n\}$ if

$$\lim_{n \to \infty} d(x_n, y_n) = 0.$$ 

What this is saying is that two Cauchy sequences are equivalent if they would be heading toward the same spot. This definition is advantageous since it doesn’t mention whether or not $\lim x_n$ or $\lim y_n$ exists, and we continue to work entirely within the space $X$.

Loosely speaking, we define $Y$ to be the set of all Cauchy sequences in $X$ with the relation that two Cauchy sequences represent the same point in $Y$ if they are equivalent in the sense above. Then we can consider $X$ to sit inside $Y$ by identifying the point $x \in X$ with the Cauchy sequence $\{x, x, x, x, \ldots\}$ in $Y$. We even can induce a metric $d'$ on $Y$ from the one on $X$ by defining

$$d'(\{x_n\}, \{y_n\}) = \lim_{n \to \infty} d(x_n, y_n).$$

Then the metric $d'$ on the copy of $X$ sitting inside $Y$ is the same as our original metric $d$ on our original set $X$.

**Definition 2.35.** Let $(X, d)$ be a metric space. The $Y$ described above is the \textit{completion} of $X$.

**Remark 2.36.** Some basic facts about completions:

- The metric space $X$ is dense inside its completion $Y$. This means that if $X \subseteq Z$ where $Z$ is a complete metric space, then $Y \subseteq Z$ as well. In this sense, a completion has some of the flavor of the closure of a set but they are still different concepts.
Given a metric space, it essentially has a unique completion. Precisely, given two completions \( Y \) and \( Z \) of \( X \), there is a one-to-one and onto mapping from \( Y \) to \( Z \) which preserves the distances between points (i.e. a bijective isometric mapping, but we haven’t defined this, yet).

I deliberately left the construction of \( Y \) a bit vague and gave some of the ideas involved. The reason is that we haven’t yet talked about sets in enough detail to give as precise a treatment as it really should have. My main purpose is to show you that every metric space \( X \) is effectively contained within a smallest complete metric space \( Y \) which can be constructed from information contained entirely within \( X \). For more information on complete metric spaces, see section 7 in [2].

### 2.6 Compactness

We now turn to the study of swordfish. Just kidding, we’re studying compactness. If I were to put the definition right here, it might make little sense and seem to come from nowhere. That’s because it took a while to get the definition precisely right.

The reason we study compactness is that we like continuous functions \( f : X \rightarrow Y \) to behave nicely. We relate the metric of the space \( X \) to the metric of the space \( Y \) by using continuous functions. Without knowing more specific properties of \( X \) or \( Y \), we can’t say too much about how functions behave. However, if we know that some of the sets involved have nice properties, then we can say a lot about any continuous function between the two spaces.

For example, we like compact sets because they aren’t “too big” in a sense, although bigness in this sense is definitely not measured strictly in set cardinality. Probably a better description is that compact sets have some qualities of finiteness. This doesn’t mean that compact sets are finite; in fact, the most interesting compact sets tend to be infinite in size. What this means is that when we ask questions involving a compact set and how it relates to other sets, often the answers are finite, or that we know the values involved are bounded.

So if we want a set \( K \) to possess properties of finiteness, then we have to control the infinite subsets of \( K \) somehow. If a subset \( A \) of \( K \) is infinite, then \( K \) has to tie it down somehow. If we could guarantee that a cluster point of \( A \) is always within \( K \), then that might work, and in fact, it does.

**Definition 2.37.** Let \( K \) be a subset of a metric space \( X \). Then we say \( K \) is countably compact if every infinite subset of \( K \) has a cluster point inside \( K \). (See [2] p. 95)

This is sometimes called limit point compactness. This certainly works because it ties down the behavior of infinite subsets within \( K \). This definition seems to be similar to that of closed sets, but it has a crucial difference. The difference is that in a closed set \( F \), if an infinite subset has a cluster point, then that point must also belong to \( F \). The definition of closed sets says that all cluster points of \( F \) belong to \( F \), but it doesn’t say anything about the existence
of cluster points. The definition of compactness also contains an assertion about the existence of cluster points.

So even though the definitions of countable compactness and closed are different, the relation of the definitions by cluster points raises the question, are all countably compact sets closed? We can quickly see that the converse isn’t true:

**Example 2.38.** Let $X = \mathbb{R}\setminus\{0\}$ be a metric subspace of $\mathbb{R}$. Then $C = [-1, 0) \cup (0, 1]$ is a closed set in $X$, but it is not countably compact since the set $\{ \frac{1}{n} : n \in \mathbb{N} \}$ is infinite, but it lacks any cluster points in $C$.

One possible reason one might provide is that each infinite subset of a countably compact set $K$ has a cluster point inside $K$, so it is possible to construct some infinite subset which has a lot of cluster points, and only some of them are inside $K$? The answer is no because of the strength of the statement. It says **every** infinite subset has a cluster point, so given any infinite set, there is a way to pare it down so the resulting set has only one cluster point which must lie within $K$. More precisely, we have the following theorem.

**Theorem 2.39.** Let $K$ be a countably compact subset of a metric space $X$. Then $K$ is closed.

*Proof.* Let $x$ be a limit point of $K$. We just need to show that $x \in K$. From the definition of limit point, there is a point of $K$ different from $x$ contained within $B_1(x)$. Label this point $x_1$. Similarly, there is a point of $K$ distinct from $x$ contained within $B_{1/2}(x)$, and label this point $x_2$. Continue by recursion to find a point $x_n \neq x$ within $K \cap B_{1/2^n}(x)$. The set $\{x_1, x_2, x_3, \ldots\}$ is infinite since if it were finite, then there would be some $r > 0$ so that $d(x, x_k) \geq r$ for every $k \in \mathbb{N}$. Then the sequence $\{x_1, x_2, x_3, \ldots\}$ converges to $x$ since given any $\varepsilon > 0$, there is some $N \in \mathbb{N}$ such that $\frac{1}{2^n} < \varepsilon$. Thus for $n \geq N$, $d(x_n, x) < \frac{1}{2^n} < \varepsilon$. \(\square\)

Similarly, we can show that any countably compact set is bounded.

**Theorem 2.40.** A countably compact subset of a metric space is bounded.

*Proof.* Suppose by way of contradiction that $K$ is a countably compact subset of metric space $X$, yet $K$ is unbounded. Then pick $x_1 \in K$. Since we are assuming $K$ is unbounded, there is some point $x_2 \in K$ so that $d(x_1, x_2) > 1$. Now for $n \geq 3$, recursively pick points $x_n$ so that $d(x_1, x_n) > d(x_1, x_{n-1}) + 1$. Using the triangle inequality, we can show that any two points of this set $\{x_1, x_2, x_3, \ldots\}$ are more than a distance of 1 apart, so it is not a Cauchy sequence, and it has no Cauchy subsequence (try taking $\varepsilon < \frac{1}{2}$, for example in the definition of Cauchy), so this set has no cluster point. But it is a subset of $K$, so it must have a cluster point, and we arrive at a contradiction. Therefore, our initial assumption that a countably compact set $K$ could be unbounded is wrong. Hence countably compact sets are bounded. \(\square\)

We have another characterization of countably compact sets:
Theorem 2.41. Let $K \subseteq X$. Then $K$ is countably compact if and only if every sequence $\{c_n\} \subseteq K$ has a subsequence $\{c_{n_i}\}$ which converges in $K$.

Proof. ($\Rightarrow$). Let $K$ be countably compact, and let $\{a_n\}$ be a sequence in $K$. If there are only a finite number of distinct entries in the sequence, then some entry, say $a$ appears infinitely often. Then there is a subsequence $\{a_{n_i}\}$ so that $a_{n_i} = a$ for every $i = 1, 2, 3, \ldots$. Then $\{a_{n_i}\}_{i \geq 1}$ is a constant sequence, and so converges to $a$.

If $\{a_n\}$ has an infinite number of distinct elements, then it is an infinite set, and by the countable compactness of $K$, there is a cluster point $b \in K$. Since $b$ is a cluster point of the sequence $\{a_n\}$, there is some $a_{n_1}$ with $d(a_{n_1}, b) < 1$. We also know that there is some $n_2 > n_1$ so that $d(a_{n_2}, b) < \frac{1}{2}$. Continue this process by recursively finding $n_k > n_{k-1}$ such that $d(a_{n_k}, b) < \frac{1}{k}$ for all $k \geq 3$. Then $\{a_{n_i}\}_{i \geq 1}$ is a subsequence of $\{a_n\}$, and by this construction $a_{n_i} \to b$.

($\Leftarrow$). Suppose $K$ is a set such that every sequence has a subsequence which converges in $K$. Let $A \subseteq K$ be an infinite set. Then we can take a sequence $\{a_n\} \subseteq A$ so that $a_j \neq a_k$ if $j \neq k$. By our assumption, there is a subsequence $\{a_{n_i}\}$ which converges to $b \in K$. Since all the elements of $\{a_{n_i}\}$ are different, there is at most one $a_{n_j} = b$, so let $U$ be an arbitrary neighborhood of $b$. Since $a_{n_i} \to b$, there is an $N \in \mathbb{N}$ so that $i \geq N$ implies $a_{n_i} \in U$. In particular, for any $i \neq j$ bigger than $N$, $a_{n_i}$ is a member of $U$ which is different than $b$. Since we have picked an arbitrary neighborhood of $b$, this applies to every neighborhood, so we have shown that every neighborhood of $b$ contains a point of $A$ different than $b$. Hence $b$ is a cluster point of $A$, and $b \in K$. Therefore, every infinite subset of $K$ has a cluster point within $K$. \hfill \square

In the reals, it was noted that closed and bounded sets have special properties, so the definition of countable compactness was developed. In fact, for a while, it was called compactness [need citation] until a better and more general definition was discovered (we will give it below) It is a fact that subsets of $\mathbb{R}$ which are closed and bounded are also countably compact (we will prove this in the Heine-Borel theorem). One might ask whether this is true for all metric spaces.

The answer is no. There are metric spaces with closed and bounded subsets, yet these subsets are not compact. A quick example is to consider an infinite discrete metric space $Y$. The whole space $Y$ is bounded. It is closed as well since all subsets of a discrete space are closed. But it is not countably compact since $Y$ has no limit points.

This brings us to the definition of compactness. We need a bit of terminology first.

Definition 2.42. Let $A$ be a subset of a metric space $X$. An open cover of $A$ is a collection of open sets $\mathcal{U} = \{U_i\}_{i \in I}$ whose union contains $A$ (hence the collection covers $A$). Note that $I$ is an index set for the open cover.

Any subset $V \subseteq \mathcal{U}$ is called an open subcover of $A$ if the union of the elements of $V$ contains $A$ (i.e. if $V$ covers $A$ as well). Precisely, let $J \subseteq I$. If the
set \{U_j\}_{j \in J} is also an open cover of \(A\), then we say it is an open subcover.

Depending on the situation, it may be simpler to think of the open cover as open sets within \(A\) with the subspace metric. It would help, I think, to have some examples of open covers. Note that an open cover is a set whose elements are sets.

**Example 2.43.**

a. Let \(A \subseteq X\). Then \(\{X\}\) is an open cover of \(A\).

b. The set \(\{(x - 2, x + 2) : x \in [1, 9]\}\) is an open cover of the interval \([0, 10]\) in \(\mathbb{R}\).

c. Let \(X\) be a discrete space. Then for any \(A \subseteq X\), \(\\{\{a\} : a \in A\}\) is an open cover of \(A\).

**Definition 2.44.** Let \(A \subseteq X\). We say \(A\) is compact if every open cover of \(A\) has a finite subcover.

Compact sets share a lot of the same properties as countably compact sets. Notably compactness shares the same kind of “reining in infiniteness” as countable compactness. This is a good comparison because all compact subsets of metric spaces are countably compact.

**Theorem 2.45.** Let \(A \subseteq X\). If \(A\) is compact, then it is countably compact.

**Proof.** Let \(A\) be compact and \(B\) be an infinite subset of \(A\). For contradiction, assume that \(A\) is not countably compact and that there is some infinite subset \(B\) which does not have an accumulation point inside \(A\). In particular, for \(a \in A\), there is some neighborhood \(U_a\) around \(a\) which contains only a finite number of elements of \(B\). Since \(a \subseteq U_a\), it follows that

\[
A \subseteq \bigcup_{a \in A} U_a,
\]

Therefore \(\{U_a : a \in A\}\) is an open cover of \(A\). By compactness, there exists a subcover

\[
\{U_{a_1}, U_{a_2}, \ldots, U_{a_n}\}.
\]

Each of the sets \(U_{a_k}\) contains a finite number of elements of \(B\), and since there is a finite number of these sets, their union contains only a finite number of elements of \(B\). But

\[
B \subseteq A \subseteq \bigcup_{k=1}^n U_{a_k},
\]

which implies that \(B\) is a finite set. By this contradiction we conclude that any compact set in a metric space is also countably compact.

The converse is also true, in that a countably compact metric space is compact, but in the interest of time we will not prove this here. If you are interested in the proof, see section 11 in [2].
So in metric spaces these two concepts are equivalent. In general topology there are examples of “large” spaces where sequences are insufficient to carry all the convergence information of the space. In such spaces we can have compact sets which are not countably compact. In particular such a space can’t be assigned any compatible metric because in metric spaces, sequences are sufficient to tell the tale of convergence.

**Example 2.46.**  

a. Any finite set of points \( A \) in a metric space \( X \) is compact.

b. Any closed, bounded interval \( [a, b] \) in \( \mathbb{R} \) is compact. [**Issue**: This is a reassuring fact, but we have not proved it yet (See Heine-Borel theorem).]

c. A compact set \( K \) can have noncompact subsets. For example \((0, 1]\) is not compact even though \([0, 1]\) is. This can happen by removing one or more cluster points of \( K \) but not removing sets which have those points as cluster points.

d. Any closed subset \( F \) of a compact set \( K \) is compact. This can be seen by taking an open cover of \( F \) and add in the relatively open set \( K \setminus F \). This union is an open cover of \( K \), so there is a finite subcover of \( K \). Then we can remove \( K \setminus F \) from this subcover to get a finite open cover of \( F \) which is a subcover of our original cover.

By the existence of cluster points in the definition of countable compactness, we can assert that a compact metric space is also complete.

**Theorem 2.47.** A compact subset of a metric space is complete.

**Proof.** Let \( K \) be a compact subset of metric space \( X \). Let \( \{c_n\} \) be a Cauchy sequence.

Case 1: If only a finite number of elements of \( K \) appear in the sequence, then there is some element \( c_k \) which appears infinitely often. So let \( \varepsilon > 0 \). Since the sequence is Cauchy, we pick an \( N \in \mathbb{N} \) in concordance with the definition. Then there is some \( m \geq N \) so that \( c_m = c_k \). Now let \( n \geq N \); we have that

\[
d(c_k, c_n) = d(c_m, c_n) < \varepsilon.
\]

This is just the definition of convergence to \( c_k \), and we already know \( c_k \in K \). So \( c_n \to c_k \in K \).

Case 2: If the sequence has an infinite number of different elements, then it is an infinite subset of \( K \). Since \( K \) is compact, this implies that it is countably compact, hence the set \( C = \{c_n\}_{n \geq 1} \) has a cluster point \( p \) contained in \( K \). We now show that \( \lim_{n \to \infty} c_n = p \).

Let \( \varepsilon > 0 \), and pick \( N \in \mathbb{N} \) associated with \( \varepsilon \) since \( C \) is a Cauchy sequence. Since \( p \) is a cluster point of \( C \), there are an infinite number of points of \( C \) within \( B_\varepsilon(p) \) (see the exercises at the end of the chapter), so we can pick \( m \geq N \) so that \( d(c_m, p) < \varepsilon \). Now pick \( n \geq N \). Then

\[
d(c_n, p) \leq d(c_n, c_m) + d(c_m, p) < \varepsilon + \varepsilon = 2\varepsilon.
\]

To sum up, for any \( \varepsilon > 0 \), there is a \( N \in \mathbb{N} \) so that \( n \geq N \) implies \( d(c_n, p) < 2\varepsilon \), hence \( \lim_{n \to \infty} c_n = p \). \( \Box \)
In this proof you might say that we only showed \( d(c_n, p) < 2\varepsilon \), not \( \varepsilon \), but that is ok. The strength of the argument comes from showing that this argument holds for any positive value of \( \varepsilon \). For example, after picking \( \varepsilon > 0 \), we could have let \( \eta = \frac{\varepsilon}{2} \) and then used \( \eta \) in place of \( \varepsilon \) throughout the proof to end up with \( d(c_n, p) < 2\eta = \varepsilon \). This would have been in strict concordance with the definition of convergence, but because we could pick any positive value for \( \varepsilon \), this argument is contained within the use of “for all.” One can see that the notions of “for all” and “there exists” are powerful.

Since countably compact subsets of a metric space are closed and bounded and compact subsets of a metric space are countably compact, it follows that compact subsets of a metric space are closed and bounded. We can also prove this directly from the definition of compactness.

**Theorem 2.48.** Let \( X \) be a metric space and \( K \) a compact subset. Then \( K \) is closed and bounded.

**Proof.** 1. \( K \) is bounded. If \( K \) is empty, then we’re done, otherwise pick \( p \in K \). Since for any point \( x \in K \), \( d(p, x) < \infty \), it is contained in \( B_n(p) \) for some \( n \in \mathbb{N} \). Therefore \( \{B_n(p)\}_{n \in \mathbb{N}} \) is an open cover of \( K \). Since \( K \) is compact, it is covered by a finite number of these,

\[
\{B_{n_1}(p), B_{n_2}(p), \ldots, B_{n_k}(p)\},
\]

and let \( N = \max\{n_1, n_2, \ldots, n_k\} \). Then \( B_{n_j}(p) \subseteq B_N(p) \) for \( j = 1, \ldots, k \), hence \( K \subseteq B_N(p) \).

2. \( K \) is closed. We will show that \( X \setminus K \) is open, i.e. we’ll find an open ball around \( x \in X \setminus K \) disjoint from \( K \). **Note:** This argument is essentially the same as when we show that compact subsets of Hausdorff topological spaces are closed, so watch for this again.

Let \( x \in X \setminus K \). Pick \( y \in K \), and let \( r_y = \frac{d(x, y)}{2} \). Then the sets \( B_{r_y}(x) \) and \( B_{r_y}(y) \) are open and disjoint (why are they disjoint?). Then the set

\[
\{B_{r_y}(y) : y \in K\}
\]

is an open cover of \( K \) (why?), so there is a finite subcover

\[
\{B_{r_1}(y_1), \ldots, B_{r_n}(y_n)\},
\]

Now let

\[
U = \bigcup_{k=1}^{n} B_{r_k}(y_k), \quad \text{and}
\]

\[
V = \bigcap_{k=1}^{n} B_{r_k}(x).
\]

Then \( U \) is an open set containing \( K \), and \( V \) is an open set containing \( x \). We can also assert that \( U \) and \( V \) are disjoint because if they weren’t, there
would be some point \( p \in U \cap V \) so that \( p \) would be in every \( B_{r_n}(x) \). But since \( B_{r_n}(y_k) \) and \( B_{r_n}(x) \) are disjoint, this can’t happen.

Now since \( U \) and \( V \) are disjoint, we have that \( V \) is a neighborhood of \( x \) which is disjoint from \( K \). Since \( x \) was an arbitrary point of \( X \setminus K \), we can conclude that \( X \setminus K \) is open, whence \( K \) is closed.

\[ \square \]

One situation we see from time to time is that we have a decreasing sequence of nonempty sets \( A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \). We’d like the situation where their intersection \( \bigcap_{n=1}^\infty \) contains a point. This would be a kind of set analogue to convergence. Unfortunately for general sets, this doesn’t always happen.

**Example 2.49.** Let \( A_n = (0, \frac{1}{n}] \subseteq \mathbb{R} \), and \( A = \bigcap_{n=0}^\infty A_n \). Obviously, if \( x \leq 0 \), then \( x \notin A \) since \( x \notin A_n \) for every \( n \geq 1 \). If \( x > 0 \), then there is some \( k \in \mathbb{N} \) so that \( \frac{1}{k} < x \) [Issue! This is a consequence of the archimedean principle which we will prove later.], so \( x \notin A_k \), and thus \( x \) can’t be in the intersection of all the \( A_n \). Thus \( A = \emptyset \).

Fortunately, if all the sets \( A_n \) are compact, then this intersection is always nonempty.

**Theorem 2.50.** Let \( A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \) be a sequence of nonempty compact sets in a metric space \( X \). Then \( A = \bigcap_{n=1}^\infty A_n \) is nonempty.

**Proof.** Let \( B_n = A_1 \setminus A_n \). Then by considering \( A_1 \) as a metric space, then \( A_n \) is closed in \( A_1 \) for every \( n \in \mathbb{N} \). Therefore \( B_n \) is open in \( A_1 \). Additionally, \( \emptyset = B_1 \subseteq B_2 \subseteq B_3 \subseteq \cdots \), so the \( B_n \) are nested sets, but just nested in the opposite order of the \( A_n \).

For the sake of contradiction, assume that \( \bigcap_{n=1}^\infty = \emptyset \). Correspondingly, this assumption is equivalent to saying that \( \bigcup_{n=1}^\infty B_n = A_1 \). This implies that \( \{B_n\}_{n \geq 1} \) is an open cover of \( A_1 \). By the compactness of \( A_1 \), there are \( n_1 < n_2 < \cdots < n_k \) in \( \mathbb{N} \) so that \( A_1 = \bigcup_{i=1}^k B_{n_i} \). By the set inclusion in the previous paragraph, this union is equal to \( B_{n_k} \). Thus \( B_{n_k} = A_1 \). But \( B_{n_k} = A_1 \setminus A_{n_k} \), whence \( A_{n_k} = \emptyset \). This contradicts that each \( A_n \) is nonempty. Thus we must have \( \bigcap_{n=1}^\infty A_n \) is nonempty. \( \square \)

This result is handy because it gives us a condition so that a sequence of nested sets doesn’t disappear entirely. Note that this result holds for compact subsets of a metric space. It doesn’t necessarily hold for just some descending sequence of closed subsets:

**Example 2.51.** Let \( A_n = [n, \infty) \subseteq \mathbb{R} \). Then each \( A_n \) is closed, and \( A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \), but \( \bigcap_{n=1}^\infty = \emptyset \).

This result rests on the property that compact subsets of a metric space are closed in that space as well as that open covers have finite subcovers. In general topological spaces, there are examples where some compact sets are not closed (and in particular these spaces cannot be assigned a metric which is compatible with the topology), and we could have a nested sequence of compact sets which have empty intersection. Stay tuned!
2.7 Connectedness

Now we study what it means for a set to be connected. Let’s examine the archtypical example of a metric space, \( \mathbb{R} \). Between any two points \( a \) and \( b \), there is an unbroken path on which we can travel, namely the interval \([a, b]\), so can we say \( \mathbb{R} \) is connected since between any two points there is a path between them? The problem with this notion is that it is a bit tautological. We’re saying that \( \mathbb{R} \) is connected because it contains a subset that “looks like” a subset of \( \mathbb{R} \), and by doing that we’re appealing to our notions of what \( \mathbb{R} \) looks like (and remember, we haven’t even axiomatized it, yet, so we’re going on intuition only here).

So let’s take a different approach; let’s consider a metric space \( X \). If \( X \) is connected, then there shouldn’t be a way to break it up into a couple chunks without somehow ripping things apart. If we split it into two nonempty disjoint pieces (subsets \( A \) and \( B \) with \( A \cap B = \emptyset \) and \( A \cup B = X \)), then the pieces should be attached to each other in some way. We have seen a notion of attachment before. If \( A \) contains a cluster point of \( B \) or if \( B \) contains a cluster point of \( A \), then we can’t pull them apart cleanly without affecting what it means to be \( X \).

So in this case we can say that \( A \) and \( B \) are attached in some way. Then we can go through \( X \) and consider all such disjoint pairs \( A \) and \( B \) whose union is the whole space. If we find some pair in which neither contains an accumulation point of the other, then we can say that that pair shows that \( X \) is disconnected, since there is no way to travel without some jump from one set to the other.

[Issue: You may have noticed that I said let’s not look at the idea of connectedness directly from paths, but instead look in terms of the metric on the space, so you may say that in using the metric we are still comparing points with \( \mathbb{R} \). I say, you’re right, and get off my lawn. But seriously, the second notion is the more general of the two since we are looking at a form of attachment of sets, and there may be many surprising ways that sets can be attached, so we don’t want to limit ourselves by saying that the space has to contain the image of an interval. The first notion is useful, though, and it is called path connectedness. We’ll define it precisely later.]

So let’s run with this idea and find an equivalent formulation.

**Theorem 2.52.** Let \( X \) be a metric space with \( A, B \subseteq X \) such that \( A \cap B = \emptyset \) and \( A \cup B = X \). Then \( A \) contains no cluster points of \( B \) and \( B \) contains no cluster points of \( A \) if and only if both \( A \) and \( B \) are open in \( X \).

**Proof.** \((\Rightarrow)\). Suppose that \( A \) contains no cluster points of \( B \) and \( B \) contains no cluster points of \( A \). Since all cluster points of \( A \) must lie within \( X \) and \( A \cup B = X \), we must have that \( A \) contains all its cluster points, hence it is closed. Therefore, \( B \) must be open. Similarly, we are able to show that \( A \) is open.

\((\Leftarrow)\). Suppose that \( A \) and \( B \) are both open. Since \( B = X \setminus A \), this implies that \( B \) is closed and contains all its cluster points. Therefore \( A \) contains no cluster points of \( B \). By a symmetric argument, \( B \) contains no cluster points of \( A \).

\( \square \)
We are now near to defining connectedness. We have to first be careful of one case. If \( A = X \) and \( B = \emptyset \), then we have seen before that both \( X \) and \( \emptyset \) are simultaneously both open and closed in \( X \), but this division of \( X \) doesn’t give us any useful information; we’re not separating points of \( X \) in this case, so we must exclude it.

**Definition 2.53.** Let \( X \) be a metric space. If there are two subsets \( A \) and \( B \) such that

- \( A \) and \( B \) are nonempty,
- \( A \cap B = \emptyset \),
- \( A \cup B = X \), and
- both \( A \) and \( B \) are open in \( X \),

then we say \( \{A, B\} \) is a **disconnection** of \( X \).

If \( X \) has a disconnection, then we say it is disconnected. If \( X \) has no disconnection, then we say that \( X \) is connected.

**Example 2.54.**

a. Any one point metric space \( X = \{x\} \) is connected.

b. We will prove later that any interval in \( \mathbb{R} \) is connected, and in fact that a subset \( C \subseteq \mathbb{R} \) is connected if and only if it is an interval. So our initial notion of connectedness via paths does have merit.

c. Let \( X = [0, 1) \cup (1, 2] \) under the subspace metric inherited from \( \mathbb{R} \). Then \( A = [0, 1) \) and \( B = (1, 2] \) is a disconnection of \( X \), i.e. these are nonempty complementary open sets. Therefore \( X \) is disconnected.

d. Let \( C \) be a finite metric space with at least two points. Then \( C \) is disconnected since if \( p \in C \), then the sets \( \{p\} \) and \( C\setminus\{p\} \) form a disconnection of \( C \).

e. The set \( \mathbb{Q} \) is disconnected because we can take the disconnection \( A = (-\infty, \pi) \cap \mathbb{Q} \) and \( B = [\pi, \infty) \cap \mathbb{Q} \). In fact, we may say that \( \mathbb{Q} \) is completely disconnected because no matter which two points we choose, we can find some disconnection which houses them apart. We will precisely define this later.

We will use connectedness later in the course. It is essential to prove the intermediate value property.

### 2.8 Continuity in metric spaces

We have referred to continuous functions without actually defining what it means for a function to be continuous. When we first encountered continuity many of us were presented with the idea that a function is continuous if we can draw its graph with a single line without holes or jumps. If we examine this
point by point of the domain, if we let \( x \) be in the domain, this idea means that for \( y \) nearby to \( x \), then \( f(y) \) ought to be nearby to \( f(x) \). Then if this property holds for all points in the domain, then we can say the function is continuous.

If we want \( f(y) \) to be near \( f(x) \) whenever \( y \) is near \( x \), then we have to have a standard of what “to be near” means. Our final condition is that the function values are close, so we have to set a tolerance on function values. So we let \( \varepsilon > 0 \), and say that \( f(y) \) is near \( f(x) \) if \( d(f(y), f(x)) < \varepsilon \). Now that we have set the tolerance on the range, \( f \) is continuous at \( x \) if there is some neighborhood sufficiently small of \( x \) so that all points in this neighborhood get mapped to points in the range near \( f(x) \). In other words, to be continuous, we have to find some \( \delta > 0 \) so that \( d(x, y) < \delta \) means that \( d(f(x), f(y)) < \varepsilon \).

Let’s look at what happens if there doesn’t exist such a \( \delta \). Then no matter how close we got to \( x \), then there would be some \( y \) so that \( d(f(x), f(y)) \geq \varepsilon \), that is, there are points arbitrarily close to \( x \) which get sent to points far away from \( f(x) \). In that case we don’t have the guarantee that \( y \) near to \( x \) implies that \( f(y) \) is near to \( f(x) \).

Definition 2.55. Let \((X, d)\) and \((Y, \rho)\) be metric spaces, and let \( f : X \to Y \) be a function. Then for some \( x \in X \), we say that \( f \) is continuous at \( x \) if for every \( \varepsilon > 0 \), \( \exists \delta > 0 \) so that if \( d(x, y) < \delta \) then \( \rho(f(x), f(y)) < \varepsilon \).

If \( f \) is continuous at every point in \( X \), then we say that \( f \) is continuous or that \( f \) is continuous on \( X \).

We have an equivalent definition of continuous functions in terms of solely open sets.

Theorem 2.56. Let \( f : X \to Y \) be a function between metric spaces. Then \( f \) is continuous if and only if \( f^{-1}(U) \) is open in \( X \) whenever \( U \) is an open set in \( Y \).

Another way of saying this is that a function is continuous if and only if the preimages of open sets are open.

Proof. Exercise. \( \square \)

We spent time with sequences, and we can describe continuity in terms of convergent sequences.

Theorem 2.57. Let \( X \) and \( Y \) be metric spaces and \( f : X \to Y \). Let \( x_0 \in X \). Then \( f \) is continuous at \( x_0 \) if and only if for every sequence \( \{x_n\} \) such that \( x_n \to x_0 \), then \( f(x_n) \to f(x_0) \).
Proof. $\Rightarrow$. Let $f$ be continuous at $x_0$ and $x_n \to x_0$. Let $\epsilon > 0$. Then from continuity, there is a $\delta > 0$ so that if $d(z, x_0) < \delta$ then $d(f(z), f(x_0)) < \epsilon$. Since $x_n$ converges to $x_0$, there is $N \in \mathbb{N}$ so that if $n \geq N$, then $d(x_n, x_0) < \delta$. Then in this case, if $n \geq N$, then $d(f(x_n), f(x_0)) < \epsilon$. Thus we have shown that $\lim_{n \to \infty} f(x_n) = f(x_0)$. Since $\{x_n\}$ was an arbitrary sequence converging to $x_0$, we have shown our conclusion.

$\Leftarrow$. Suppose that $f(x_n) \to f(x_0)$ for every sequence such that $x_n \to x_0$. Suppose, for contradiction, that $f$ is not continuous. Then there exists $\epsilon > 0$ such that for any $\delta > 0$ there exists a point $x \in X$ with $d(x, x_0) < \delta$, but $d(f(x), f(x_0)) \geq \epsilon$.

Set $\delta = 1$, and let $x_1 \in X$ so that $d(x_1, x_0) < 1$ but $d(f(x_1), f(x_0)) \geq \epsilon$. Now, let $\delta = \frac{1}{2}$, and find $x_2 \in X$ so that $d(x_2, x_0) < \frac{1}{2}$ but that $d(f(x_2), f(x_0)) \geq \epsilon$. Similarly for each $k \in \mathbb{N}$ let $x_k \in X$ with $d(x_k, x_0) < \frac{1}{k}$ but $d(f(x_k), f(x_0)) \geq \epsilon$. Now the sequence $\{x_n\}_{n \geq 1}$ converges to $x_0$, but by construction, for any $n$, $d(f(x_n), f(x_0)) \geq \epsilon$, so $f(x_n)$ does not converge to $f(x_0)$, which contradicts our assumption. Therefore, $f$ must be continuous.

Since continuous functions are the main way we link the metric on one space to the metric on another, we would hope they behave nicely with respect to the metrics, and indeed they do.

**Theorem 2.58.** Let $X$ be compact and let $f : X \to Y$ be a continuous function. Then $f(X)$ is compact.

Proof. Let $\{V_i\}_{i \in I}$ be an open cover of $f(X)$. For each $i \in I$, let $U_i = f^{-1}(V_i)$. In light of the theorem above, each $U_i$ is open in $X$, so $\{U_i\}_{i \in I}$ is an open cover of $X$. By the compactness of $X$ there is an open subcover $\{U_1, U_2, \ldots, U_n\}$. Since $X \subseteq \bigcup_{k=1}^n U_k$, it follows that $f(X) \subseteq \bigcup_{k=1}^n V_k$. Then we see that $\{V_1, V_2, \ldots, V_n\}$ is an open cover of $f(X)$. Since our original open cover was arbitrary and we were able to find a subcover, this shows that $f(X)$ is compact.

We also get the nice result that connectedness is preserved by continuity.

**Theorem 2.59.** Let $X$ be a connected metric space and $f : X \to Y$ be a continuous function. Then $f(X)$ is connected.

Proof. Suppose by way of contradiction that there exists a disconnection $\{A, B\}$ of $f(X)$. Then $f^{-1}(A)$ and $f^{-1}(B)$ are open sets in $X$, and $f^{-1}(A) \cup f^{-1}(B) = X$ since the union of $A$ and $B$ contains the image of $X$. Similarly, both $f^{-1}(A)$ and $f^{-1}(B)$ are nonempty and disjoint. Therefore $\{f^{-1}(A), f^{-1}(B)\}$ form a disconnection of $X$, and we arrive at a contradiction. Thus we must conclude that $f(X)$ is connected.

This is a significant result, and we will use it to prove the intermediate value theorem later.

We have established that continuous functions are significant tools to link one metric space to another. Sometimes we can use functions to show that parts of two metric spaces behave exactly alike.
Definition 2.60. Let \((X, d)\) and \((Y, \rho)\) be two metric spaces, and let \(f : X \to Y\) be a function. If for every \(a, b \in X\),
\[
d(a, b) = \rho(f(a), f(b)),
\]
then we say \(f\) is an **isometry**.

We say two spaces \(X\) and \(Y\) are **isometric** if there exists an isometric surjection \(f : X \to Y\).

Essentially, an isometry is a correspondence between \(X\) and its image so that the metric is preserved. In terms of how metrics behave, we can consider \(X\) and \(f(X)\) to be two representations of the same metric space.

Since an isometry is a function between metric spaces which preserves metrics, it would be a bit disturbing if it weren’t continuous. Fortunately, isometries are continuous.

**Lemma 2.61.** An isometry is continuous.

**Proof.** Let \(f : X \to Y\) be an isometry. Let \(x \in X\) and \(\varepsilon > 0\). Set \(\delta = \varepsilon\), and let \(d(a, b) < \delta\). Then \(\rho(f(a), f(b)) = d(a, b) < \delta = \varepsilon\). Thus \(f\) is continuous at \(x\), and since \(x\) was an arbitrary element of \(X\), it follows that \(f\) is continuous on \(X\). \(\Box\)

**Example 2.62.**

a. Let \(\mathbb{R}\) have the usual metric and \(\mathbb{R}^2\) have the euclidean metric. The map \(x \mapsto (x, 0)\) is an isometry \(\mathbb{R} \to \mathbb{R}^2\). Note that this map is an still an isometry if \(\mathbb{R}^2\) has the \(d_1\) or \(d_{\infty}\) metrics.

b. Let \(X\) be a metric space and \(A \subseteq X\). Consider \(A\) to be a metric space as a subspace of \(X\). Then the inclusion map \(i : A \hookrightarrow X\) is an isometry.

Note: The inclusion map is a function on sets in general. If \(P \subseteq Q\), then every \(x \in P\) is also in \(Q\). The inclusion map \(i : P \hookrightarrow Q\) is defined by \(i(x) = x\) for every \(x \in P\).

Also, in case you haven’t seen it before, the notation “\(\hookrightarrow\)” indicates that the function is an injection.

h. Let \(C[0, 1]\) have the \(d_1\) metric. Consider a function \(\phi : \mathbb{R} \to C[0, 1]\) by
\[
a \mapsto f_a(x) \equiv a.
\]
Then \(\phi\) is an isometry. We show this by letting \(a, b \in \mathbb{R}\). Then \(d(a, b) = |a - b|\), and
\[
d_1(\phi(a), \phi(b)) = d_1(f_a(x), f_b(x))
= \int_0^1 |f_a(x) - f_b(x)| \, dx
= \int_0^1 |a - b| \, dx
= |a - b|
= d(a, b).
\]
Note that if we take the corresponding map \( a \mapsto f(x) \equiv a \) from \( \mathbb{R} \) into \( C[0, 2] \) with the \( d_1 \) metric, then this map is \textit{not} an isometry. The integration in this case is taken from 0 to 2, not from 0 to 1. To find an isometry \( \mathbb{R} \to C[0, 2] \), please see exercise 20.

y. Let \( X \) be a set and \( d \) and \( p \) two different metrics on \( X \). Then the identity function \( f : X \to X \) by \( f(x) = x \) is \textit{not} an isometry if the domain and range possess different metrics.

z. There is no isometry \( \mathbb{R}^2 \to \mathbb{R} \) under the standard metrics.
   For contradiction, suppose there is. Then some \( (x, y) \in \mathbb{R}^2 \) gets sent to \( 0 \in \mathbb{R} \). Note that there are an infinite number of points in \( \mathbb{R}^2 \) which are a distance of 1 away from \( (x, y) \), but only two points in \( \mathbb{R} \) with a distance of 1 away from 0. So an infinite number of points get mapped to the points \( \{-1, 1\} \). But by exercise 15 in this chapter, an isometry is one to one. Thus there is no isometry \( \mathbb{R}^2 \to \mathbb{R} \) with the usual metrics.

2.9 Exercises

1. Show that \( \mathbb{R}^n \) with the usual euclidean distance is a metric space. Items a-c will guide you through the proof.

   a. Define a function called an \textbf{inner product} on pairs of points \( x = (x_1, x_2, \ldots, x_n) \) and \( y = (y_1, y_2, \ldots, y_n) \) in \( \mathbb{R}^n \) by
      \[
      \langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.
      \]
      Show that for any \( x \in \mathbb{R}^n \), \( \langle x, x \rangle \geq 0 \) and is equal to 0 if and only if \( x = 0 \) (i.e. the inner product is positive definite).
      Also show that for any \( x, y \in \mathbb{R}^n \), \( \langle x, y \rangle = \langle y, x \rangle \) (i.e. the inner product is symmetric). Third, show that the inner product is linear in each component, that is for \( \alpha, \beta \in \mathbb{R} \) and \( x, y, z \in \mathbb{R}^n \) show
      \[
      \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle
      \]
      to establish linearity in the first coordinate, then use the symmetry of the inner product to give linearity in the second coordinate.

   b. For \( x \in \mathbb{R}^n \) define the \textbf{norm} of \( x \) by
      \[
      \|x\| = \langle x, x \rangle^{\frac{1}{2}}.
      \]
      Note that this gives the standard euclidean distance in \( \mathbb{R}^n \) of \( x \) from the origin.
      We now establish the \textbf{Cauchy-Schwarz-Bunyakovsky} inequality:
      \[
      |\langle x, y \rangle| \leq \|x\| \|y\|.
      \]
      This is a significant inequality, and is essential in the study of inner product spaces.

41
Let \( \lambda \in \mathbb{R} \) and \( x, y \in \mathbb{R}^n \). Consider the inequality from positive definiteness
\[
\langle \lambda x + y, \lambda x + y \rangle \geq 0.
\]
Use linearity and symmetry to rewrite this as
\[
\langle x, x \rangle \lambda^2 + 2 \langle x, y \rangle \lambda + \langle y, y \rangle \geq 0.
\]
Think of this as a quadratic in the variable \( \lambda \), and since the left side is always nonnegative, the discriminant in the quadratic formula is nonpositive, so that
\[
4 \langle x, y \rangle^2 - 4 \langle x, x \rangle \langle y, y \rangle \leq 0.
\]
From this deduce that
\[
|\langle x, y \rangle| \leq \|x\| \|y\|.
\]

c. Define the function \( d(x, y) = \langle x - y, x - y \rangle^{\frac{1}{2}} \). We show that \( d \) is a metric on \( \mathbb{R}^n \).
   i. Show that \( d \) is positive definite.
   ii. Show that \( d \) is symmetric.
   iii. Consider three points \( x, y, z \in \mathbb{R}^n \). Since \( \mathbb{R}^n \) is a linear space and that in the function \( d \) we are considering differences of points, we may assume \( y = 0 \) for simplicity, so \( x - y = x \) and \( y - z = -z \). With this simplification, we have \( d(x, y) = \|x\| \) and \( d(y, z) = \|z\| \).
   Show that
\[
d(x, z)^2 = \|x\|^2 - 2 \langle x, z \rangle + \|z\|^2.
\]
   Using the Cauchy-Schwarz inequality, show that
\[
d(x, z)^2 \leq \|x\|^2 + 2 \|x\| \|z\| + \|z\|^2.
\]
   Taking square roots, conclude that
\[
d(x, z) \leq \|x\| + \|z\| = d(x, y) + d(y, z).
\]
   Hence, \( \mathbb{R}^n \) is a metric space using the metric derived from the inner product.

d. Given any real vector space \( V \), if there is a function \( \langle \cdot, \cdot \rangle \) which satisfies
   i. \( \langle x, x \rangle \geq 0 \), and \( \langle x, x \rangle = 0 \) if and only if \( x = 0 \) (positive definiteness),
   ii. \( \langle x, y \rangle = \langle y, x \rangle \) (symmetry), and
   iii. \( \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \) for all \( \alpha, \beta \in \mathbb{R} \) and \( x, y, z \in V \),
then we say $V$ is a **real inner product space**. The Cauchy-Schwarz inequality follows from these axioms, so $V$ becomes a metric space under the metric $d(x, y) = \langle x - y, x - y \rangle ^{\frac{1}{2}}$.

If we want to consider **complex** vector spaces we have to change things a slight bit. First, we have to make sure the function respects complex linearity, so we must consider $\alpha, \beta \in \mathbb{C}$ instead of the reals. Second, the symmetric property must account for the complex variables as well. So we replace ii. with

ii. $\langle x, y \rangle = \langle y, x \rangle$.

so $V$ equipped with such a function is a **complex inner product space**.

e. There are a lot of inner product spaces out there. For example $C[0, 1]$ is a real (complex) inner product space with inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x) \, dx = \int_0^1 f(x)\overline{g(x)} \, dx.$$

f. There are a lot of spaces which are **not** inner product spaces. For example, the metrics $d_1$ and $d_\infty$ on $\mathbb{R}^2$ do not come from inner products.

Similarly, the metrics $d_1$ and $d_\infty$ on $C[0, 1]$ don’t come from inner products.

2. Given a metric space $(X, d)$ define a function $d' : X \times X \to \mathbb{R}$ by

$$d'(x, y) = \frac{d(x, y)}{1 + d(x, y)}.$$

a. Show that $(X, d')$ is a metric space. **Hint:** When showing the triangle inequality, at some point you will have something like $d(x, y) + d(y, z)$ in the denominator, and you replace this with $d(x, z)$ which will make the fraction larger (but that’s ok since you’re showing an inequality).

b. Let $\{x_n\}$ be a sequence in $X$. Show that $x_n \to x$ in the metric $d$ if and only if $x_n \to x$ in the metric $d'$. This property is important because it shows that transforming $d$ into $d'$ doesn’t change which points are close to others.

3. Show that an arbitrary union of open sets is open.

4. Let $X$ be a metric space and $S \subseteq X$. Prove that $\text{ext}(S) = (X \setminus S)^\circ$.

5. Show that $\mathbb{R} \setminus \mathbb{Z}$ is open in $\mathbb{R}$.

6. Prove that in $\mathbb{R}$, $\partial \mathbb{Q} = \mathbb{R}$. **Hint:** between any two real numbers, there is a rational.

7. Show that an arbitrary intersection of closed sets is closed.
8. Let $A$ be a subset of a metric space $X$, and let $p$ be a cluster point of $A$. Prove that any neighborhood of $p$ contains infinitely many points of $A$.

9. Let $A$ be a subset of a metric space. Show that

$$A^o = \bigcup_{U \text{ is open}} U.$$ 

This gives the interior of a set in terms of open sets without mention of the metric involved. This is useful when we look at general topology.

10. Let $A \subseteq X$. Prove that $\partial A = \overline{A} \cap (\overline{X\setminus A})$. In general topological spaces, we will take this to be the definition of the boundary of $A$.

11. Show that for $A \subseteq X$, $A^o = X\setminus(\overline{X\setminus A})$.

12. Show that the sequence

$$3, 3.1, 3.14, 3.141, 3.1415, 3.14159, \ldots$$

is a Cauchy sequence in $\mathbb{Q}$.

13. Let $X$ be a metric space. Prove that $X$ is connected if and only if the only sets which are both open and closed are $\emptyset$ and $X$.

14. Let $f_n(x) = x^n$ be a sequence in $C[0, 1]$. Show that this sequence converges to the $f(x) \equiv 0$ function under the $d_1$ metric but does not converge to $f(x)$ under the $d_{\infty}$ metric. This is an important example of how changing the metric changes how the space behaves.

15. Let $X$ and $Y$ be metric spaces, and let $f : X \to Y$ be an isometry. Prove that $f$ is one-to-one.

16. Let $X$ and $Y$ be metric spaces and $f : X \to Y$. Prove that $f$ is continuous if and only if it has the property that whenever $U$ is an open set of $Y$, then $f^{-1}(U)$ is open in $X$.

This property is important because it gives us a way to define continuity in terms of open sets on $X$ and $Y$ without mentioning the metric. We will later use this as the definition of continuity in general topological spaces.

17. Given any $B \subseteq X$, prove that $B^o$, $\partial B$, and $\text{ext}(B)$ are pairwise disjoint sets, and $X = B^o \cup \partial B \cup \text{ext}(B)$.

18. Let $X$ be a discrete metric space and $Y$ any metric space. Show that every function $f : X \to Y$ is continuous.

19. Consider the interval $X = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ with the metric $\rho(x, y) = |\tan^{-1}(x) - \tan^{-1}(y)|$, and consider $\mathbb{R}$ with the usual metric. Show that $f : X \to \mathbb{R}$ given by $x \mapsto \tan x$ is an isometry.
20. Consider $C[0, 1]$ with the $d_1$ metric and $C[0, 2]$ with its $d'_1$ metric defined by

$$d'_1(f, g) = \int_0^2 |f(x) - g(x)| \, dx,$$

for $f, g \in C[0, 2]$. Define a mapping $\phi : C[0, 1] \to C[0, 2]$ by

$$\phi(f)(x) = \frac{f(x^2)}{2}.$$

Show that $\phi$ is an isometry.

21. Let $X, Y,$ and $Z$ be metric spaces and $f : X \to Y$ and $g : Y \to Z$ be continuous functions. Prove that the composition $g \circ f : X \to Z$ is continuous.

### 2.10 Metric Spaces Epilogue

In this chapter we have seen a lot of concepts and issues which will form the basis of our study of the reals and functions on the reals.

- We will use that $\mathbb{R}$ is a metric space, and in particular it is the archetypical metric space.
- We use open sets, and in particular neighborhoods to determine closeness to a point.
- In convergence of a sequence or in continuity, we use that given any tolerance ($\varepsilon > 0$) we can test whether the sequence or function behaves within that tolerance.
- The closure of a set in some sense determines the reach of a set’s influence. In the same sense, the boundary of a set is the common region of a set and its complement’s influence. The interior of a set is the region within a set for which every point is some distance away from the set’s complement.
- Compactness and connectedness play large roles in studying the behavior of functions between spaces.
- Functions, and in particular continuous functions are the main way we relate the properties of one space to another.
- Continuity is important.
- Continuity is really important.
- Further in the course, we will see that differentiability at a point implies continuity at the point.
We have seen some definitions of concepts in metric spaces and equivalent formulations in terms of open sets. Keep these in mind because we will use these as the definitions when we study topological spaces where distances may no longer apply, but we do have notions of nearness in terms of open sets.

3 Sets

3.1 Definitions and basic operations

Math wouldn’t be math as we know it without sets. We need a way to define collections of things and membership in these collections. In a sense, sets are the building blocks of mathematics, and logic is the glue holding these blocks together.

Intuitively, we can think of a set as a well defined collection of objects. The formal notions of objects, sets, and membership are essentially symbolic and ultimately we reach a level where we accept the existence of these things as axioms. So the question, “What really is a set?” leads us to more philosophical questions.

We describe the notions of $\in$, $\notin$, and what it means to be a set.

**Definition 3.1.** We think of $a \in A$ to mean that $a$ is contained in the set $A$. We also think of $a \notin A$ to mean that $a$ is not contained in $A$. Logically, the statement $a \notin A$ is equivalent to the statement $\sim (a \in A)$.

Loosely speaking, a set is an object such that for any $a$, exactly one of $a \in A$ or $a \notin A$ is true.

We say two sets $A$ and $B$ are equal if for any $x$, $x \in A \iff x \in B$.

We can create other sets from existing ones.

**Definition 3.2.** Let $A$ and $B$ be sets. We define the operations **intersection** $\cap$ and **union** $\cup$:

\[
x \in A \cap B \iff x \in A \text{ and } x \in B
\]
\[
x \in A \cup B \iff x \in A \text{ or } x \in B
\]

From these operations, we would like the result to always be a set. If $A$ and $B$ are two sets which have no elements in common, then the intersection $A \cap B$ contains no elements. This indicates that we must have an empty set. Formally, we can say there exists a set $\emptyset$ so that for every $x$, $x \notin \emptyset$.

We also define intersection and union for any collection of sets. We use that the operator “for every” can be an elaborate version of “and” and that “there exists” can be an elaborate version of “or.”

**Definition 3.3.** Let $\{S_i\}_{i \in I}$ be a collection of sets. (Note the set $I$ is an index set, so we can address any particular $S_i$ individually.) Then we define the
intersection of the sets by
\[ x \in \bigcap_{i \in I} S_i \iff x \in S_i, \forall i \in I \]
and the union of the sets by
\[ x \in \bigcup_{i \in I} S_i \iff \exists j \in I \text{ such that } x \in S_j. \]

**Definition 3.4.** We say set \( A \) is contained in set \( B \), or \( A \subseteq B \) if for every \( a \), \( a \in A \implies a \in B \).

We also say \( C \supseteq D \) when \( D \subseteq C \).

Note that \( A \subseteq A \cup B \) and \( A \cap B \subseteq A \) for any sets \( A \) and \( B \).

One of the most common ways we show that two sets \( X \) and \( Y \) are equal is to show \( X \subseteq Y \) and \( Y \subseteq X \).

We can also create sets from other sets. The set
\[ B = \{ a \in A | P(a) \} \]
is all the elements \( a \) of \( A \) which satisfy property \( P \). In particular, if \( x \in B \), then \( x \in A \) as well as satisfying some property, so \( B \subseteq A \).

**Definition 3.5.** Let \( X \) be a set and \( A \subseteq X \). Then the complement of \( A \) in \( X \) is the set
\[ X \setminus A = \{ x \in X | x \notin A \}. \]

Complements, unions, and intersections distribute in the following ways:

**Theorem 3.6.** Let \( X \) be a set, \( \{ B_\alpha \}_\alpha \in A \) a collection of subsets of \( X \), and \( C \subseteq X \).

\[
\begin{align*}
C \cup \left( \bigcap_{\alpha \in A} B_\alpha \right) &= \bigcap_{\alpha \in A} (C \cup B_\alpha) \\
C \cap \left( \bigcap_{\alpha \in A} B_\alpha \right) &= \bigcap_{\alpha \in A} (C \cap B_\alpha) \\
C \cap \left( \bigcup_{\alpha \in A} B_\alpha \right) &= \bigcup_{\alpha \in A} (C \cap B_\alpha) \\
C \cup \left( \bigcup_{\alpha \in A} B_\alpha \right) &= \bigcup_{\alpha \in A} (C \cup B_\alpha) \\
X \setminus \left( \bigcap_{\alpha \in A} B_\alpha \right) &= \bigcup_{\alpha \in A} (X \setminus B_\alpha) \\
X \setminus \left( \bigcup_{\alpha \in A} B_\alpha \right) &= \bigcap_{\alpha \in A} (X \setminus B_\alpha)
\end{align*}
\]
The last two lines are known as DeMorgan’s laws.

Proof. We’ll prove the last equality. The other five are left to the reader.

(⊆). Let \( x \in X \setminus \bigcup_{\alpha \in A} B_\alpha \). Then \( x \notin B_\alpha \) for every \( \alpha \in A \), otherwise there would be some \( B_\beta \ni x \), and then \( x \in \bigcup_{\alpha \in A} B_\alpha \). Therefore, for every \( \alpha \in A \), \( x \in X \setminus B_\alpha \). So by the definition of intersection, \( x \in \bigcap_{\alpha \in A} (X \setminus B_\alpha) \).

(⊇). Let \( x \in \bigcap_{\alpha \in A} (X \setminus B_\alpha) \). Then \( x \) is in every \( X \setminus B_\alpha \). This implies that \( x \) is not in \( \bigcup_{\alpha \in A} B_\alpha \) because if it were, there would be some \( B_\beta \) so that \( x \in B_\beta \), and this would violate that \( x \in X \setminus B_\alpha \) for every \( \alpha \). Therefore \( x \in X \setminus \bigcup_{\alpha \in A} B_\alpha \).

3.2 Cartesian products, relations, orders, and functions

Definition 3.7. Let \( A \) and \( B \) be sets. We define the Cartesian product of \( A \) and \( B \) to be the set

\[
A \times B = \{(a, b) | a \in A \text{ and } b \in B\}.
\]

A quick parenthetical note: How do we define an ordered pair? A set on its own is not intrinsically equipped with an order, so how do we create one? Norbert Wiener came up with the first known definition of an ordered pair in terms of sets [5], but we may use the slightly improved one by Kuratowski [4]:

\[
(a, b) := \{\{a\}, \{a, b\}\}.
\]

Other than this paragraph, we won’t worry about the specific definition of ordered pairs; we will just use them and know they exist. I just wanted to show you that there is a way to define an ordered pair from unordered things.

Definition 3.8. We then define a relation \( R \) between set \( A \) and set \( B \) by \( R \subseteq A \times B \). We say that \( A \) is the domain of the relation and \( B \) is the codomain. Note that except for functions, we typically won’t use the terms domain and codomain.

If we consider two sets \( A \) and \( B \), an arbitrary relation \( R \) between them we may not have much to say about the two sets. However, if we examine relations which have particular properties, then we can build ways to usefully link two sets or a set with itself.

Definition 3.9. Let \( R \subseteq A \times A \) be a relation. Then we say \( R \) is an equivalence relation if for every \( a, b, c \in A \),

a. \((a, a) \in R\) (reflexivity),

b. if \((a, b) \in R\), then \((b, a) \in R\) (symmetry), and

c. if \((a, b) \in R\) and \((b, c) \in R\), then \((a, c) \in R\) (transitivity).

Depending on circumstance, we often notate an equivalence relation by \( a = b \), \( a \equiv b \), \( a \sim b \), or \( a \cong b \) for \((a, b) \in R\).
Another concept which is useful in set theory and very closely related to equivalence classes is the concept of partitions. If we take a set $X$ and slice it up into disjoint nonempty sets, then we have partitioned $X$; we have grouped elements together in $X$.

**Definition 3.10.** Let $X$ be a nonempty set and $\{A_i\}_{i \in I}$ be a collection of subsets of $X$. If $A_i$ is nonempty for every $i \in I$, $A_i \cap A_j = \emptyset$ if $i \neq j$, and $\bigcup_{i \in I} A_i = X$, then we say that $\{A_i\}$ is a partition of $X$.

Partitions and equivalence relations are linked because for any nonempty $X$, there is a 1-1 correspondence between equivalence relations on $X$ and partitions of $X$. We can see that if we have an equivalence relation, we can create a subset of $X$ corresponding to $x \in X$ by making the set $H_x = \{y \in X | y \sim x\}$. If we look at the collection of all sets created in this way, then that forms a partition of $X$. If, instead, we start with a partition, we can define an equivalence relation by $x \sim y$ if there is some $j \in I$ so that $x, y \in A_j$. We prove this now.

**Theorem 3.11.** Let $X$ be nonempty. Then there is a 1-1 correspondence between partitions of $X$ and equivalence relations on $X$.

**Proof.** Let $\sim$ be an equivalence relation on $X$. For $x \in X$, consider the subset $H_x = \{y \in X | y \sim x\}$. Every $H_x$ is nonempty since at least $x \in H_x$. We also observe that the union of all such $H_x$ is $X$ since for any $z \in X$, $z \in H_z \subseteq \bigcup_{x \in X} H_x$, and this union is contained in $X$ since each $H_x \subseteq X$.

To show the property that $A_i \cap A_j = \emptyset$ if $i \neq j$, this is equivalent to showing that if $H_x \cap H_y \neq \emptyset$, then $H_x = H_y$. This means that we can have some duplication in labeling, and using the axiom of choice we can choose a distinct label for each of our subsets of $X$, but the important idea is that any two sets we pick must either be disjoint or be the same set. So suppose that $z \in H_x \cap H_y$. Then $x \sim z$ and $z \sim y$, so by transitivity, $x \sim y$. Thus if $w \sim y$, then transitivity shows $w \sim x$, so we must have $H_y \subseteq H_x$. Similarly, if $w \sim x$, then transitivity shows $w \sim y$, whence $H_x \subseteq H_y$. Therefore $H_x = H_y$.

For the other direction, suppose that $\{B_i | i \in I\}$ is a partition of $X$. Define a relation on $X$ by $x \sim y$ if and only if there is some $j \in I$ so that $x, y \in B_j$. Then

- $z \sim z$ since $z \in B_j$ for some $j \in I$, and this implies $z \in B_j$.
- If $x \sim y$, then that means that $x \in B_j$ for some $j$, and $y \in B_j$. Then $y \in B_j$ and $x \in B_j$. Thus $y \sim x$.
- Suppose $x \sim y$ and $y \sim z$. Then there are $j, k \in I$ so that $x, y \in B_j$ and $y, z \in B_k$. From the definition of a partition, since the sets $\{B_i\}$ are pairwise disjoint, if $B_j \cap B_k \neq \emptyset$, then $B_j = B_k$. Since $y$ is in both of these sets, then it must follow that $B_j = B_k$, and thus $x$ and $z$ are contained in this set. Therefore $x \sim z$.

Thus $\sim$ satisfies the three properties of an equivalence relation. \qed
Equivalence relations are helpful because they can take a large set of things and reduce by a collection of equivalences, then the resulting equivalences can have useful properties.

Example 3.12. a. Suppose that one has a set $X$ with a function $d$ which has all the properties of a metric except that $d(x, x) = 0$ but there may be points $x, y \in X$ with $x \neq y$ so that $d(x, y) = 0$. Since this isn’t a metric space, we can’t use many of the results of metric spaces on it, so we might not be able to guarantee that it or some subsets possess certain properties. However, if we say that elements which are a distance of 0 apart are equivalent, then we can insure that the resulting set of equivalence classes forms a metric space. So although $X$ is not a metric space, it can produce one.

If we define an equivalence relation on $X$ so that $x \sim y$ iff $d(x, y) = 0$

Definition 3.13. Let $R$ be a relation on a set $A$ and notate that $a \preceq b$ if $(a, b) \in R$. Then we say $R$ is a partial order if

- a. for every $a \in A$, $a \preceq a$ (reflexivity),
- b. for every $a, b \in A$, if $a \preceq b$ and $b \preceq a$, then $a = b$ (anti-symmetry), and
- c. for any $a, b, c \in A$, if $a \preceq b$ and $b \preceq c$, then $a \preceq c$ (transitivity).

We also have another notation. If $a \preceq b$ and $a \neq b$, then we often say $a \prec b$. Note that for any two elements $a, b \in A$, we do not require either $a \preceq b$ or $b \preceq a$. It may be that neither of these occur; we say that this $a$ and $b$ are not comparable in the partial order.

If we have a partial order on a set $A$ so that for every two elements $a, b \in A$, either $a \preceq b$ or $b \preceq a$, then we say that our relation is a linear order, a total linear order, or in the notation of [1] and [3], an order.

Every linear order is a partial order, but not conversely so.

Example 3.14. a. The usual order $<$ on $\mathbb{R}$ and $\mathbb{Q}$ are orders by this definition. Since they are orders, they are partial orders as well.

b. Let $A$ be any set. The set of subsets of $A$ can be partially ordered by the relation: for every $B, C \subseteq A$, then $B \preceq C$ iff $B \subseteq C$. If $A$ contains two distinct elements $a, b$, then this order is a partial order and not a total linear order because neither of the sets $\{a\}$ and $\{b\}$ are subsets of each other and thus are not comparable in the order.

c. If we wanted, we could have defined the partial order by $B \preceq C$ if $B \supseteq C$. Defining it the other way around still satisfies the axioms of an order relation.

d. More generally, if we have any (partial) order $\preceq$ on a set $X$, then the reverse (partial) order $\preceq'$ on $X$ defined by $x \preceq' y$ if and only if $y \preceq x$ is
also a (partial) order on $A$ since $a \preceq b \preceq c$ is the same as $a \succeq b \succeq c$, and the other axioms are satisfied by flipping the sign around.

If we start with an order and flip everything around, we still have a way to tell whether we can compare any two elements, and “betweenness” is still preserved through transitivity.

e. Consider the set $\mathbb{Z} \times \mathbb{Z}$ so that $(a, b) \preceq (c, d)$ iff $a \leq c$ and $b \leq d$. Then $\preceq$ is a partial order on $\mathbb{Z} \times \mathbb{Z}$.

Geometrically considering $\mathbb{Z} \times \mathbb{Z} \subseteq \mathbb{R}^2$, if we fix some point $(a, b)$, then the set of points $\{(c, d) \in \mathbb{Z} \times \mathbb{Z} \mid (c, d) \preceq (a, b)\}$ is the collection of points which are above and to the right of $(a, b)$, including points directly above or to the right of $(a, b)$. Note that the points $(3, 4)$ and $(4, 3)$ are not comparable in this partial order.

f. Consider $\mathbb{R}^2$ with the relation $(a, b) \preceq (c, d)$ iff either $a^2 + b^2 > c^2 + d^2$ or $(a, b) = (c, d)$. This is a partial order.

Geometrically, $(x, y) \succeq (w, z)$ if $(x, y)$ is closer to the origin than $(w, z)$ or in the case where they are the same distance to the origin, if $(x, y) = (w, z)$.

Note that two points are not comparable only when they are the same distance to the origin but are distinct points.

g. Let $(A, \preceq)$ be a partially ordered set, and $B \subseteq A$ be a nonempty set. Then $B$ is a partially ordered set with a partial order inherited from $A$ if we say for any $x, y \in B$ that $x \preceq_B y$ if $x \preceq y$ in $A$. This concept is similar in flavor to that of a metric subspace.

If we consider $\mathbb{R}$ under the usual ordering and $\mathbb{Q} \subseteq \mathbb{R}$, then $\mathbb{Q}$ with its regular order satisfies this condition.

In partially ordered sets, between any two elements we have some order or partial order relation, or the two elements are not comparable. We’d like to use the order properties to relate various subsets.

**Definition 3.15.** Let $A$ be a partially ordered set and $X \subseteq A$. Then we say that $a \in A$ is an **upper bound** of $X$ if for every $x \in X$, $x \preceq a$. We then say that $X$ is **bounded above**.

Similarly, if there is a $b \in A$ so that $x \in X$ implies $b \preceq x$, then we say that $b$ is a **lower bound** of $X$, and that $X$ is **bounded below**.

Sometimes for a set which we have an element which can represent all upper bounds or an element which represents all lower bounds.

**Definition 3.16.** Let $(A, \preceq)$ be a partially ordered set and $B \subseteq A$.

- Suppose $B$ is bounded above. If there is an $x \in A$ so that $x$ is an upper bound of $B$ and for any other upper bound $y$ of $B$, $x \preceq y$, then we say $x$ is a **least upper bound** or **supremum** of $B$. We say that $x = \text{sup } B$.
  (Sometimes you’ll see the notation that $x = \text{lub } B$.)
Let $B$ be bounded below. If there is an $w \in A$ so that $w$ is a lower bound of $B$ and for any other lower bound $z$ of $B$, $z \preceq w$, then we say $w$ is a **greatest lower bound** or **infimum** of $B$. We write $w = \inf B$, and sometimes you’ll see $x = \text{glb} B$.

Note that the supremum of a set doesn’t have to belong to the set itself.

**Example 3.17.**

a. In $\mathbb{R}$ or $\mathbb{Q}$ any finite set is bounded above and also has a least upper bound. In general, any finite subset of a linearly ordered set is bounded above and has a least upper bound.

b. Consider the set $B = \{x \in \mathbb{R} | x^2 < 2\} \subseteq \mathbb{R}$. This set is bounded above and has least upper bound $\sqrt{2}$. Note that the supremum of $B$ doesn’t belong to $B$.

d. In a partial order, a finite subset does not necessarily have an upper bound.

Let $A = \{0, 1\} \times \mathbb{R}$. Say that $(x, y) \preceq (w, z)$ if $x = w$ and $y \leq z$. Geometrically, this can be represented by two vertical lines in $\mathbb{R}^2$ where two points are comparable if and only if one point is directly above the other (or they are the same point). If the two points are from different lines, then the points are not comparable. Then the set $B = \{(0, 5), (1, 5)\}$ has no upper bound.

e. The set $\mathbb{R}$ with the regular order has the supremum property. We will take this to be an axiom, and it has important consequences.

The set $\mathbb{Q}$ with the usual order does not have the supremum property. For example, the subset $\{x \in \mathbb{Q} | x^2 < 2\}$ is bounded above (by 10, for example), but there is no least upper bound.

f. Consider the order on $\mathbb{Z} \times \mathbb{Z}$ in example 3.14. The set $B = \{(x, y) | x+y = 0\}$ has no upper bound. For any $(a, b) \in \mathbb{Z} \times \mathbb{Z}$, $(a + 1, -a - 1) \in B$, and $(a + 1, -a - 1) \not\preceq (a, b)$, so $(a, b)$ is not an upper bound for $B$.

functions (also ordered tuples and sequences as functions)

**Definition 3.18.** Let $A$ and $B$ be sets and $f$ a relation on $A$ and $B$. Suppose $f$ has the property that for every $a \in A$, there exists $(a, b) \in f$, and if $(a, b), (a, c) \in f$, then $b = c$. Any relation which has this property is called a **function**.

We often write $f : A \to B$ for the function $f$, and also $f(a) = b$ if $(a, b) \in f$.

Notation: Let $f : X \to Y$ be a function and $A \subseteq X$. Then $f(A) = \{f(a) | a \in A\}$.

injections, surjections, bijections

**Definition 3.19.** Let $f : A \to B$ be a function.

We say that $f$ is an **injection** or is **one to one** if it has the property that $f(a) = f(b)$ implies $a = b$.

We say that $f$ is a **surjection** or is **onto** if for every $b \in B$ there is an $a \in A$ so that $f(a) = b$.

We say that $f$ is a **bijection** if it is both a surjection and an injection.
3.3 Well-ordering and cardinality

**Definition 3.20.** The natural numbers has the property that for any nonempty subset \( B \subseteq \mathbb{N} \), the set \( B \) contains a smallest element. This property is called the well-ordering property.

This is somewhat similar to the greatest lower bound property, but there are two significant differences.

1. We’re not assuming that the subset \( B \) is bounded below. That it is bounded below is a consequence of this property.
2. The smallest element of \( B \) automatically belongs to \( B \).

The well-ordering property implies the greatest lower bound property since \( \mathbb{N} \) has a smallest element, namely 1, therefore 1 is a lower bound for any subset of \( \mathbb{N} \). So any nonempty set is bounded below, and there is a greatest lower bound. We get the bonus that the greatest lower bound is automatically contained in the set under consideration.

The well ordering property implies that \( \mathbb{N} \) is a totally linearly ordered set, the proof of which is left to the reader.

Each set has a notion of size. Roughly speaking, we can say that the size of \( A \) is the number of elements in \( A \). Specifically, each set \( A \) has a cardinality \( |A| \). We say two sets \( A \) and \( B \) have the same cardinality if there is a one to one matching of the elements of \( A \) with all the elements of \( B \).

**Definition 3.21.** We say that two sets \( A \) and \( B \) have the same cardinality or \( |A| = |B| \) if there is a bijection \( f : A \to B \).

We say that the cardinality of \( A \) is less than or equal to the cardinality of \( B \), or \( |A| \leq |B| \) if there is an injection \( g : A \to B \).

What follows is seemingly obvious, but it is a profound result.

**Theorem 3.22 (Cantor-Bernstein).** Let \( A \) and \( B \) be two sets so that \( |A| \leq |B| \) and \( |B| \leq |A| \). Then \( |A| = |B| \).

**Proof.** A very nice proof with explanation and an example can be found on the web at http://www.mathpath.org/proof/Sch-Bern/proofofS-B.htm. \( \square \)

If the sets \( A \) and \( B \) are finite, then this can be done with a simple induction. Of course, we’d like a definition of a finite set.

**Definition 3.23.** Let \( A \) be a set.

- We say \( A \) is **finite** if there is some \( n \in \mathbb{N} \) so that there is a bijection \( f : \{1, 2, \ldots, n\} \to A \).
- We say \( A \) is **infinite** if there is no such \( n \in \mathbb{N} \).
- We say that \( A \) is **countably infinite** if there is a bijection \( f : \mathbb{N} \to A \).
• We say $A$ is **countable** if it is either finite or countably infinite.

It’s reassuring to know that $\{1, 2, \ldots, n\}$ and $\mathbb{N}$ are different sizes, especially since seeing that one of the sets is finite and the other is infinite.

**Theorem 3.24.** Let $n \in \mathbb{N}$. Then there is no bijection between $\{1, 2, \ldots, n\}$ and $\mathbb{N}$.

**Proof.** Suppose for contradiction that $f : \mathbb{N} \to \{1, 2, \ldots, n\}$ is a bijection. Let $A_k = \{x \in \mathbb{N} | f(x) = k\}$. Then for each $k = 1, 2, \ldots, n$, the set $A_k$ is nonempty since $f$ is a surjection, so by the well-ordering principle $A_k$ has a smallest element $y_k$.

Let $B = \{y_1, y_2, \ldots, y_n\}$. By induction, this set has a greatest element, say $y_j$. Then $y_j + 1 \neq y_k$ for any $k = 1, 2, \ldots, n$. Now consider $f(y_j + 1) \in \{1, 2, \ldots, n\}$. Then by induction again, there is some $m$ with $1 \leq m \leq n$ so that $f(y_j + 1) = f(y_m)$, but $y_j + 1 \neq y_m$. Thus the function $f$ is not an injection. This contradicts that it is a bijection, hence no such bijection exists. $\square$

There is also an equivalent definition of an infinite set, which we state as a theorem.

**Theorem 3.25.** Let $B$ be a set. Then $B$ is infinite if and only if it has a proper subset $A$, that is $A \subset B$, so that there is a bijection $f : B \to A$.

**Proof.** The proof depends on the well-ordering principle of the natural numbers. The well-ordering principle, which is equivalent to the principle of induction, is a form of the axiom of choice, so this proof is related to the proof on the web for the Cantor-Bernstein theorem.

$(\Leftarrow)$. Consider the inclusion function $i : A \hookrightarrow B$. Let $x_1 \in B \setminus A$. For $n \geq 2$, define $x_n = i(f(x_{n-1})) \in B$. Note that if $n > 1$, then $x_n \neq x_1$ since $x_n$ is in the image of $f$ in $A$, hence $x_n \in A$, but $x_1 \notin A$. Now suppose that $x_n = x_m$ for some $m, n \geq 2$. We can assume without loss of generality that $m \leq n$. By one of the exercises, since $i$ and $f$ are injective, $i \circ f$ is injective, so it follows that $x_{m-1} = x_{n-1}$. By induction (here’s the well-ordering principle at work) we can show that $x_1 = x_{n-m+1}$, thus $1 = n - m + 1$ because if it weren’t, then $x_1 \neq x_{n-m+1}$, hence $m = n$. Now we create an injection $g : \mathbb{N} \to B$ by $g(k) = x_k$.

If $B$ were finite, then there would be $n \in \mathbb{N}$ and bijection $f : \{1, 2, \ldots, n\} \to B$. Then the function $g^{-1} \circ f : \{1, 2, \ldots, n\} \to \mathbb{N}$ would be a surjection, which would contradict the proof of theorem 3.24, hence $B$ must be infinite.

$(\Rightarrow)$. Let $B$ be infinite. Pick $x_1 \in B$. For $n \geq 2$ pick $x_n \in B \setminus \{x_1, \ldots, x_{n-1}\}$ recursively (again applying the well-ordering principle). Consider the set $A = \{x_1, x_2, x_3, \ldots\}$. Define a function $f : B \setminus \{x_1\}$ by

$$f(x) = \begin{cases} x, & x \in B \setminus A \\ x_{k+1}, & x \in A. \end{cases}$$

Then $f$ is an injection since if $f(x) = f(y)$, then it is easy to show that $x = y$, and $f$ is surjective since if $x \in B \setminus A$, then by setting $y = x$, then $f(y) = x$, and

54
if $x = x_n$ where $n \geq 2$, then $f(x_{n-1}) = x_n$. Therefore $f$ is a bijection from $B$ to $B \setminus \{x_1\}$, which is a proper subset of $B$. 

### 3.4 Exercises

1. Consider $\mathbb{R}^2$ with the identification $(x, y) \sim (w, z)$ if and only if $x^2 + y^2 = w^2 + z^2$. Prove that $\sim$ is an equivalence relation. What are the equivalence classes of $\sim$?

2. Let $A$ be a partial ordered set and $B$ a subset of $A$.
   (a) If $B$ has an upper bound $x$, show that if $y \geq x$, then $y$ is also an upper bound of $B$.
   (b) If $B$ has a lower bound $x$, show that $y \leq x$ implies that $y$ is a lower bound of $B$.

3. Let $(A, \preceq)$ be a partial order. Prove that $A$ has the least upper bound property if and only if it possesses the greatest lower bound property.

4. Using the definition of least upper bound here and in [1], show that in a partially ordered set $A$, if $B \subseteq A$ has a least upper bound, then it is unique.

5. Consider the definition of least upper bound found in [3] (1.8), but apply it to partially ordered sets instead of only to totally ordered sets. Find an example of a partially ordered set $A$ and a subset $B \subseteq A$ so that $B$ has more than one least upper bound. Are any of these upper bounds comparable?

   In the case of totally ordered sets, these two definitions are equivalent, but when we extend the ideas to more general order structures, we see that there are two different and non-equivalent generalizations.

6. Let $X$ be a partially ordered set with the property such that if $x, y \in X$, then there exists $z \in X$ so that $x \preceq z$ and $y \preceq z$. Prove that if $M$ is a finite subset of $A$, then $M$ has an upper bound.

7. Let $f : X \to Y$ and $g : Y \to Z$ be functions.
   (a) Prove that if both $f$ and $g$ are injective, then $g \circ f$ is injective.
   (b) Prove that if both $f$ and $g$ are surjective, then $g \circ f$ is surjective.
   (c) Prove that if $g \circ f$ is injective, then $f$ is injective.
   (d) Prove that if $g \circ f$ is surjective, then $g$ is surjective.

8. Let $A$ and $B$ be finite sets so that $|A| \leq |B|$ and $|B| \leq |A|$. Prove that $|A| = |B|$.

9. Let $A$ be a set. Prove that $A$ is countable if and only if there is an injection $f : A \to \mathbb{N}$.
4 The real numbers and topology

4.1 Algebraic properties of the reals

We begin by building the algebraic structure of the reals.

**Definition 4.1.** Let $F$ be a set equipped with two functions $+: F \times F \to F$ and $\cdot : F \times F \to F$ (Note: we’ll typically use the notation $a + b$ for $+(a, b)$ and $ab$ or $a \cdot b$ for $\cdot(a, b)$). Then $F$ is a field if

1. for every $a, b, c \in F$, $(a + b) + c = a + (b + c)$;
2. there exists an element $0 \in F$ so that for every $x \in F$, $0 + x = x + 0 = x$;
3. for every $x \in F$ there is an element $-x \in F$ so that $x + (-x) = 0$;
4. for every $a, b \in F$, $a + b = b + a$;
5. for every $x, y, z \in F$, $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$;
6. for every $a, b, c \in F$, $(ab)c = a(bc)$;
7. there exists an element $1 \in F$ so that for every $x \in F$, $1x = x1 = x$;
8. for every $a, b \in F$, $ab = ba$;
9. for every $x \in F$ such that $x \neq 0$, there is an element $x^{-1} \in F$ so that $xx^{-1} = 1$;

For those who have taken abstract algebra, properties (1), (2), and (3) make $F$ a group under addition, and adding (4) makes $F$ a commutative, or abelian group. Property (5) shows how addition and multiplication interact, and by adding (6), we get a ring. Most rings that we care about contain a unity, so many that often the definition of a ring includes a unity, property (7). Adding in property (8) we get a commutative ring with unity, and by insuring multiplicative inverses with (9) we get a field.

uniqueness of additive and multiplicative inverses

**Theorem 4.2.**

a. $0$ is the unique additive identity.

b. $1$ is the unique multiplicative identity. Sometimes $1$ is called the unity.

c. Let $a \in \mathbb{R}$. Then $a$ has a unique additive inverse.

d. Let $a \in \mathbb{R}$ with $a \neq 0$. Then $a$ has a unique multiplicative inverse.

h. $0 \cdot a = 0$.

j. For any $a \in \mathbb{R}$, $-a = (-1)(a)$.
Proof. a. Suppose \(0'\) is any additive identity. Then

\[
0' = 0' + 0 = 0.
\]

The left equality arises since 0 is an identity. The right one comes from that \(0'\) is an identity.

b. Suppose \(1'\) is any unity. Then

\[
1' = (1')(1) = 1.
\]

The same reasoning applies here.

c. Suppose \(b\) is any additive inverse. Then

\[
b = b + 0 = b + (a + (-a)) = (b + a) + (-a) = 0 + (-a) = -a.
\]

d. Suppose \(d\) is any multiplicative inverse. Then

\[
d = d \cdot 1 = d \cdot (a \cdot a^{-1}) = (d \cdot a) \cdot a^{-1} = 1 \cdot a^{-1} = a^{-1}.
\]

h. First observe

\[
0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a.
\]

Subtracting \(0 \cdot a\) from both ends, we get \(0 = 0 + 0 \cdot a\), and thus \(0 = 0 \cdot a\).

j. We are given an additive inverse \(-a\) of \(a\), so compute

\[
a + (-1)a = 1 \cdot a + (-1)a = (1 + (-1))a = 0 \cdot a = 0.
\]

Thus \((-1)a\) is also an additive inverse of \(a\), but by the uniqueness of additive inverses proved above, we must have \((-1)a = -a\).

k. \(\square\)

4.2 Order properties of the reals

Definition 4.3. Let \(F\) be a field. Then we say that \(F\) is an ordered field if there is a subset \(\mathcal{P} \subseteq F\) so that

1. if \(a, b \in \mathcal{P}\), then \(a + b \in \mathcal{P}\) (closed under sums);
2. if \(a, b \in \mathcal{P}\), then \(ab \in \mathcal{P}\) (closed under products);
3. if \(a \in F\), then either \(a \in \mathcal{P}\), \(a = 0\), or \(-a \in \mathcal{P}\), and exactly one of these is true (the law of trichotomy).

We often call \(\mathcal{P}\) the set of positive elements of \(F\).
We have already encountered a couple of ordered fields: \( \mathbb{R} \) and \( \mathbb{Q} \) with the set \( \mathcal{P} \) being the usual set of positive numbers in each case. There are other ordered fields. [There exists a way to order the rational functions \( \mathbb{R}(x) \) so that they are an ordered field. If you are interested, look up the works of Artin.]

**Definition 4.4.** For any ordered field \( F \), we define for any \( a, b \in F \), \( a < b \) if and only if \( b - a \in \mathcal{P} \). We say \( a \leq b \) if and only if either \( a < b \) or \( a = b \).

In particular, \( a \in \mathcal{P} \) if and only if \( a > 0 \).

**Definition 4.5.** Let \( I \) be a subset of \( F \). We say that \( I \) is an **interval** if it has the property that if \( x, y \in I \) with \( x \leq y \), and \( z \in \mathbb{R} \) with \( x \leq y \leq z \), then \( z \in I \).

Essentially, if two points \( x, y \) are contained in an interval, then the interval also contains all points between \( x \) and \( y \).

In \( \mathbb{R} \), there are ten possible types of intervals:

\[ \mathbb{R}, (-\infty, b), (-\infty, b], (a, \infty), [a, \infty), (a, b), [a, b], (a, b], [a, b], \emptyset. \]

I leave it for an exercise to show that these are the only types.

**4.3 Combining algebra and order**

**Theorem 4.6.** Let \( a, b, c, d \in \mathbb{R} \).

a. Let \( a, b, c \in \mathbb{R} \) with \( a < b \). Then \( a + c < b + c \).

b. If \( a < 0 \) and \( b > 0 \), then \( ab < 0 \).

c. If \( a < b \) and \( c > 0 \), then \( ac < bc \).

d. If \( a < b \) and \( c < 0 \), then \( ac > bc \).

e. If \( a < b \) and \( c < d \), then \( ac < bd \).

h. \( 1 > 0 \).

m. If \( 0 < a < b \), then \( 0 < \frac{1}{a} < \frac{1}{b} \).

**Proof.**

a. \( (b + c) - (a + c) = b - a \in \mathcal{P} \).

b. Since \( -a \in \mathcal{P} \) and \( b \in \mathcal{P} \), it follows that \( -(ab) \in \mathcal{P} \). Thus \( -(ab) \in \mathcal{P} \), which is equivalent to saying \( 0 - ab \in \mathcal{P} \), whence \( ab < 0 \).

c. \( b - a > 0 \) and \( c > 0 \), so \( (b - a)c > 0 \).
archimedean property
sequences
convergence
completeness of \( \mathbb{R} \)
monotonic sequences (bounded monotonic implies convergent)
limsup and liminf
uncountability of \( \mathbb{R} \)
Bolzano-Weierstrass theorem
Heine-Borel theorem
Compactness of \([a, b]\)
Connectedness of \( \mathbb{R} \)
Cantor set

4.4 Point-set topology

Definition and examples

**Theorem 4.7.** The continuous image of a connected set is connected.

**Proof.** Let \( X \) and \( Y \) be topological spaces with \( X \) connected, and let \( f : X \to Y \) be a continuous function.

Assume for contradiction that \( f(X) \) is disconnected. Then \( f(X) \) has a disconnection, that is, there are two disjoint nonempty relatively open sets \( A, B \) in \( f(X) \) so that \( A \cup B = f(X) \). Since \( A \) and \( B \) are relatively open in \( f(X) \) and that \( f \) is continuous, it follows that \( f^{-1}(A) \) and \( f^{-1}(B) \) are open in \( X \). Since each of \( A \) and \( B \) are nonempty subsets of \( f(X) \), it follows that \( f^{-1}(A) \) and \( f^{-1}(B) \) are also nonempty. Because \( A \) and \( B \) are disjoint and that \( f \) is a well-defined function, we must have that \( f^{-1}(A) \) and \( f^{-1}(B) \) are disjoint. Since \( A \cup B = f(X) \), it follows that for any \( x \in X \), \( f(x) \) must be a member of either \( A \) or \( B \). So this implies that \( f^{-1}(A) \cup f^{-1}(B) = X \).

Thus we have that \( f^{-1}(A) \) and \( f^{-1}(B) \) satisfy all the requirements of a disconnection of \( X \). But this is a contradiction since we assumed that \( X \) was connected. Hence we can conclude that \( f(X) \) is connected. \( \square \)

From this theorem it is a quick consequence that if \( X \) is connected and \( f : X \to Y \) is a continuous surjection, then \( Y \) must also be connected.

We also have another important consequence in \( \mathbb{R} \):

**Theorem 4.8 (Intermediate Value Theorem).** Let \( f : [a, b] \to \mathbb{R} \) be continuous, and let \( y \) be a real number between \( f(a) \) and \( f(b) \). Then there is some \( c \in [a, b] \) so that \( f(c) = y \).

**Proof.** If \( f(a) = f(b) \), then we can take either \( c = a \) or \( c = b \), and we’re done. So let’s assume that \( f(a) < f(b) \). First, we note that \([a, b]\) is continuous. Now suppose that there is no \( c \in [a, b] \) so that \( f(c) = y \). Then the set \( f([a, b]) \) is contained within the set \( \mathbb{R}\backslash \{y\} \). Since \( f(a) < y < f(b) \), this implies that the image of \( f \) is disconnected. But this is a contradiction since \([a, b]\) is connected.
Now if $f(b) < f(a)$, define $g : [a, b] \to \mathbb{R}$ by $g(x) = -f(x)$ and apply the paragraph above to find $c \in [a, b]$ so that $g(a) < g(c) < g(b)$. From this we can conclude that $f(b) < f(c) < f(a)$. \hfill \square

4.5 Exercises

1. Let $X$ be a metric space with a countably infinite number of elements. Prove that $X$ is disconnected.

2. Prove that every metric space is Hausdorff.

5 Series

In a way, series are just sequences which are defined by using addition. By adding the extra algebraic structure, though, many new issues and interesting properties crop up.

- partial sums
- alternating series
- absolute convergence
- rearrangements

6 Functions

6.1 Continuity

Continuity

- sum/product of two continuous functions is continuous
- Uniform continuity continuous on compact set implies uniform continuity.

6.2 Some nice things about continuity

continuous image of compact set is compact.

- continuous image of connected set is connected.
- Intermediate value theorem.
- fixed point definition

**Definition 6.1.** Let $X$ be a set, and let $f : X \to X$. If there is some $x \in X$ so that $f(x) = x$, then we say $x$ is a **fixed point** of $f$.

Given an arbitrary function $X \to X$, there is no guarantee that $f$ has any fixed points, but if we are given some specific properties of $X$ and $f$, we may deduce that in some circumstances $f$ must have a fixed point.

**Theorem 6.2.** Every continuous function $f : [a, b] \to [a, b]$ has a fixed point.
Proof. If either we have \( f(a) = a \) or \( f(b) = b \), then we’re done, so let’s assume that \( f(a) \neq a \) and \( f(b) \neq b \). Since both \( f(a) \) and \( f(b) \) lie in the interval \([a, b]\), we must have that \( f(a) > a \) and \( f(b) < b \). So \( f(a) - a > 0 \) and \( f(b) - b < 0 \). Consider the function \( g(x) = f(x) - x \). Then \( g \) is continuous, \( g(a) > 0 \), and \( g(b) < 0 \). Therefore, by the intermediate value theorem, there is some \( y \in [a, b] \) so that \( g(y) = 0 \). By the definition of \( g \), this implies \( f(y) - y = 0 \), or \( f(y) = y \).

One issue to remember is that for this theorem to hold true, the domain and range must be a closed and bounded interval. If the domain is not closed and bounded or the function not continuous, then the function may not have a fixed point.

Example 6.3. Each of the following functions does not have a fixed point. Which condition(s) of the previous theorem do(es) not apply?

a. Let \( f : (0, 1) \to (0, 1) \) by \( f(x) = \frac{x}{2} \).

b. Let \( g : [-3, 3] \to [-3, 3] \) be defined by

\[
g(x) = \begin{cases} 
3, & x \leq 0 \\
-3, & x > 0.
\end{cases}
\]

c. Let \( h : [0, \infty) \to [0, \infty) \) by \( h(x) = x + 1 \).

d. Let \( f : [-2, -1] \cup [1, 2] \to [-2, -1] \cup [1, 2] \) be defined by \( f(y) = -y \).

e. Let \( g : [1, 2] \to [0, 1] \) be defined by \( g(z) = z - 1 \).

There is a generalization of this fixed point theorem to certain subsets of \( \mathbb{R}^n \), and it is known as the Brouwer fixed point theorem. The proof of this theorem involves algebraic topology, so we will not include it here because it involves many concepts beyond the scope of this course, but you should look it up.

6.3 Sequences of functions

sequences of functions, different notions of convergence: pointwise, \( d_1, d_\infty \). Uniform convergence, Cauchy sequence under \( d_\infty \) has continuous limit \( \to \) \( C[a, b] \) under \( d_\infty \) is complete.

Can have Cauchy sequence in \( C[a, b] \) under \( d_1 \) which doesn’t converge to a continuous function:

Example 6.4. Consider the space \( C[0, 2] \) under the \( d_1 \) metric. Let \( f_n \) be defined by

\[
f_n(x) = \begin{cases} 
x^n, & 0 \leq x < 1, \\
1, & 1 \leq x \leq 2.
\end{cases}
\]
Then $f_n(x)$ is continuous for all $n$ since $\lim_{x \to 1^-} f_n(x) = \lim_{x \to 1^+} f_n(x) = f_n(1) = 1$. From the definition of $d_1$, it is easy to see that if we define

$$f(x) = \begin{cases} 0, & 0 \leq x < 1, \\ 1, & 1 \leq x \leq 2, \end{cases}$$

then $d_1(f_n, f) = \frac{1}{n+1}$. Thus it follows that $\{f_n\}$ is a Cauchy sequence, but it does not converge within the space $C[0, 2]$.

This example shows that $C[0, 2]$ under the $d_1$ metric is not complete.

### 6.4 Series of functions

**Series of functions**

- Absolute and uniform convergence

### 6.5 More function awesomeness

**Weierstrass approximation theorem**

- Upper and lower semicontinuity
  - Sup of continuous functions is lsc; inf of continuous functions is usc.

### 6.6 Metric spaces revisited

Some properties of complete metric spaces are quite powerful. In this section we will look at two results which have far reaching consequences. First, the contraction mapping theorem states that given any complete metric space $X$ and a function $f : X \to X$ which brings points closer together (we’ll be more precise in a minute), then the function $f$ has a unique fixed point. This is immensely useful in dynamics and integral equations since we may set up a situation where we can iterate $f$ to show that a particular system has a stable behavior or that an integral equation has a unique solution. We won’t necessarily get to these applications in this course, but it is handy to have the background.

The second result is Baire’s theorem which states roughly that a complete metric space can’t be made up of a countable number of “thin” parts. This is a powerful and subtle theorem that indicates that by defining metric spaces in terms of $\mathbb{R}$, completeness carries in some way some uncountable information about $\mathbb{R}$ along for the ride. This is true, even if the complete metric space has a finite number of points - the consequence in this case is that nonempty sets aren’t “thin.” Baire’s theorem is crucial in the proof of several theorems fundamental to functional analysis. This is one reason it is so near and dear to my heart.

First, though, we give a proof that we glossed over before. We gave a hand-waving explanation of why a metric space has a completion in which it is dense, but now that we have more tools under our belt we can give a proper proof.
Theorem 6.5. Let $X$ be a metric space. Then it has a unique completion $\tilde{X}$ up to isometry, that is, if $Y$ is any other completion, then there is a bijective isometry $f : \tilde{X} \to Y$.

The idea of this proof is that the information contained in all the Cauchy sequences of $X$ tells us what points need to be added to $X$ to make the result complete. How do we do this in a reasonable manner without knowing if $X$ is contained in another metric space? We consider the set $Z$ of all Cauchy sequences of $X$ with the identification that two sequences $\{x_n\}$ and $\{y_n\}$ are equivalent if they would be heading to the same spot, i.e. that $d(x_n, y_n) \to 0$ as $n \to \infty$. By defining a metric on this set of equivalences in a reasonable way, then this becomes the completion we’re looking for.

Proof. Proof is a beautiful thing. \[\Box\]

Let us now turn to contraction mappings.

Definition 6.6. Let $X$ be a metric space and $f : X \to X$ be a function. If there is a $c$ with $0 \leq c < 1$ such that for any $x, y \in X$,

$$d(f(x), f(y)) \leq cd(x, y),$$

then we say that $f$ is a contraction mapping.

The idea for the contraction mapping theorem is that in a contraction mapping on a complete metric space, the function $f$ “shrinks” the space, so that if we apply it repeatedly, all points head toward one particular point which is the fixed point of $f$.

Theorem 6.7 (Contraction mapping theorem). Let $X$ be a nonempty complete metric space, and $f : X \to X$ a contraction mapping. Then $f$ has a unique fixed point.

Proof. First, let’s prove uniqueness. Suppose $x$ and $y$ are fixed points of $f$. Then $d(f(x), f(y)) \leq cd(x, y)$. If $x \neq y$, then we have

$$d(x, y) = d(f(x), f(y)) \quad (\text{since } x \text{ and } y \text{ are fixed points})$$

$$\leq c \cdot d(x, y)$$

$$< d(x, y),$$

which is a contradiction. Therefore we must have $x = y$ to make $d(x, y) \leq cd(x, y)$ a true statement.

Now we prove existence. Pick any $x_0 \in X$. For $n \in \mathbb{N}$, let $x_n = f(x_{n-1})$. 

63
Claim: \( \{x_n\} \) is a Cauchy sequence. We will use that

\[
d(x_1, x_2) \leq c \cdot d(x_0, x_1),
\]

\[
d(x_2, x_3) \leq c^2 d(x_0, x_1),
\]

\[
d(x_3, x_4) \leq c^3 d(x_0, x_1),
\]

\[\vdots\]

\[
d(x_n, x_{n+1}) \leq c^n d(x_0, x_1),
\]

\[\vdots\]

Let \( \varepsilon > 0 \), and pick \( N \in \mathbb{N} \) so that \( \frac{c^{N-1}}{1-c} < \varepsilon \). Let \( m, n \geq N \), and without loss of generality, assume \( m < n \). (If \( m = n \), then \( d(x_m, x_n) = 0 \), and we’re ok.) Then

\[
d(x_m, x_n) \leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n)
\]

\[\leq c^m d(x_0, x_1) + c^{m+1} d(x_0, x_1) + \cdots + c^{n-1} d(x_0, x_1)\]

\[= d(x_0, x_1) \sum_{k=m}^{n-1} c^k\]

\[\leq d(x_0, x_1) \sum_{k=m}^{\infty} c^k\]

\[= d(x_0, x_1) c^m \sum_{j=0}^{\infty} c^j\]

\[= d(x_0, x_1) \frac{c^m}{1-c}\]

\[\leq d(x_0, x_1) \frac{c^N}{1-c}\]

\[< d(x_0, x_1) \cdot \varepsilon.
\]

So for any \( \varepsilon > 0 \) we can find \( N \in \mathbb{N} \) so that \( d(x_m, x_n) < \varepsilon \cdot d(x_0, x_1) \). Since \( d(x_0, x_1) \) is fixed, this implies that \( \{x_n\} \) is a Cauchy sequence.

Now since \( X \) is complete, there is some \( x \in X \) so that \( x_n \xrightarrow{n} x \). We show that \( x \) is a fixed point of \( f \) by showing \( d(x, f(x)) = 0 \).

Since \( \{x_n\} \) is a Cauchy sequence converging to \( x \), for \( \varepsilon > 0 \) there is some \( N_\varepsilon \in \mathbb{N} \) so that \( m, n \geq N_\varepsilon \) implies that

\[
d(x, x_n) < \varepsilon, \quad d(x, f(x_n)) < \varepsilon, \quad \text{and} \quad d(f(x_n), f(x)) < \varepsilon.
\]

The first of these inequalities follows from the convergence of \( \{x_n\} \), the second since we can go far enough in the sequence to make \( c^n \) as small as we like. The third inequality follows from \( d(f(x_n), f(x)) \leq c \cdot d(x_n, x) \). Thus

\[
d(x, f(x)) \leq d(x, x_n) + d(x_n, f(x_n)) + d(f(x_n), f(x)) < 3\varepsilon.
\]
Since \( \varepsilon > 0 \) was arbitrary, this shows that \( d(x, f(x)) = 0 \), and thus \( x \) is a fixed point of \( f \).

A consequence of the contraction mapping theorem is that suppose you are walking on campus and have a map of campus in your possession, then no matter where you are on campus, there is a unique point on the map which is exactly over the point to which it corresponds (in a theoretical sense - in the real world, there are, like, atoms and stuff).

**Example 6.8.**

a. In \( \mathbb{R} \), the mapping \( f(x) = -2 - \frac{x}{3} \) is a contraction mapping. It is an exercise to prove this and find its fixed point.

b. Consider \( C[0,1] \) with the \( d_\infty \) metric. We have shown that this space is complete. Consider the function \( \phi : C[0,1] \to C[0,1] \) by

\[
\phi(f)(x) = 1 + \frac{1}{2} \int_0^x f(t) \, dt.
\]

Then we claim that \( \phi \) is a contraction mapping, and the fixed point of this mapping is the function \( g(x) = e^{x^2} \).

To show that \( \phi \) is a contraction mapping, let’s take two functions \( f, g \in C[0,1] \). Then

\[
d_\infty(\phi(f), \phi(g)) = \sup_{0 \leq x \leq 1} \left| \left( 1 + \frac{1}{2} \int_0^x f(t) \, dt \right) - \left( 1 + \frac{1}{2} \int_0^x g(t) \, dt \right) \right|
\]

\[
= \sup_{0 \leq x \leq 1} \left| \frac{1}{2} \int_0^x f(t) - g(t) \, dt \right|
\]

\[
\leq \sup_{0 \leq x \leq 1} \frac{1}{2} \int_0^x |f(t) - g(t)| \, dt
\]

\[
= \frac{1}{2} \int_0^1 |f(t) - g(t)| \, dt
\]

\[
\leq \frac{1}{2} \int_0^1 d_\infty(f, g) \, dt
\]

\[
= \frac{1}{2} d_\infty(f, g).
\]

Therefore, \( \phi \) is a contraction mapping, and we can take \( c = \frac{1}{2} \).

To be a fixed point, we must have a \( g \in C[0,1] \) so that \( \phi(g) = g \), i.e.

\[
g(x) = 1 + \int_0^x g(t) \, dt.
\]  

(2)

Since \( g \) is a fixed point of \( \phi \), it follows that it is a differentiable function since we can take the derivative of the right hand side. So taking the derivative of both sides, we arrive at the differential equation

\[
g'(x) = \frac{1}{2} g(x),
\]
such that \( g(0) = 1 \) (by plugging in \( x = 0 \) in (2)), we solve it to get
\[ g(x) = e^x. \]
Since \( g \) is differentiable, it is continuous (we will prove this later), so \( g \in C[0,1] \). If we evaluate \( \phi(g)(x) \), we get back \( g(x) \), so by the contraction mapping theorem, \( g(x) = e^x \) is the unique fixed point of \( \phi \).

**Definition 6.9.** Let \( A \) be a subset of a metric space \( X \). Then we say \( A \) is **nowhere dense** if its closure has empty interior.

**Example 6.10.**

a. A finite point set \( A \subseteq \mathbb{R} \) is nowhere dense. Any finite point set in \( \mathbb{R} \) is closed, so \( \overline{A} = A \), and \( A^o = \emptyset \) since any open set in \( \mathbb{R} \) has an uncountable number of points.

b. The Cantor set is nowhere dense.

m. The set \( \mathbb{Q} \subseteq \mathbb{R} \) is *not* nowhere dense since \( \overline{\mathbb{Q}} = \mathbb{R} \), and certainly \( \mathbb{R} \) contains a nonempty open set.

y. Let \( X \) be a discrete metric space and \( D \) a nonempty subset. Then \( D \) is *not* nowhere dense since by virtue of \( X \) being discrete, then \( D \) is also open.

**Theorem 6.11 (Baire’s theorem).** Let \( X \) be a complete metric space. Then \( X \) cannot be written as a countable union of nowhere dense sets.

**6.7 Exercises**

1. For each item in Example 6.3, show why the function fails to have a fixed point.

2. From example 6.8, show that \( f : \mathbb{R} \to \mathbb{R} \) by \( x \mapsto -2 - \frac{x}{2} \) is a contraction mapping on \( \mathbb{R} \), and find its fixed point. Prove that this is the only fixed point of \( f \).

3. In the proof of the contraction mapping theorem, it is important to have a \( c < 1 \) so that \( d(f(x), f(y)) \leq c \cdot d(x, y) \). It is not sufficient to just have \( d(f(x), f(y)) < d(x, y) \) when \( x \neq y \).

Consider \( A = [1, \infty) \subseteq \mathbb{R} \). Verify that \( A \) is a complete metric space. Let \( f : A \to A \) by
\[ f(x) = x + \frac{1}{x}. \]
Show that if \( x \neq y \), \( d(f(x), f(y)) < d(x, y) \) but that \( f \) does not have a fixed point.

4. (*) Let \( f \in C[0,1] \) so that \( f(0) = f(1) = 0 \). Define a **chord** of \( f \) to be a number \( c \in [0,1] \) so that there exists some \( x \in [0,1] \) such that \( x + c \in [0,1] \) and \( f(x) = f(x + c) \). Let \( \mathcal{F} \) be the set of all the functions \( f \) in \( C[0,1] \) so
that \( f(0) = f(1) = 0 \). Define a universal chord of \( F \) to be a number \( c \in [0,1] \) so that \( c \) is a chord of every \( f \in F \). What are the universal chords of \( F \)?

This problem was given to me by Carl Cowen who got it from Paul Halmos. I will try to find the original reference.

7 Differentiation

**Definition 7.1.** Let \( A \subseteq \mathbb{R} \) be an open set and \( f : A \to \mathbb{R} \). Then for \( x_0 \in A \), we say that \( f \) is **differentiable** at \( x_0 \) or that \( f \) has a **derivative** at \( x_0 \) if the limit

\[
\lim_{y \to x_0} \frac{f(x_0) - f(y)}{x_0 - y}
\]

exists and is finite, and we say that this limit is the derivative of \( f \) at \( x \), and we call this limit \( f'(x_0) \) or \( \frac{df}{dx} \mid_{x=x_0} \).

If \( f \) is differentiable at every point \( x \in A \), then \( f' \) is also a function on \( A \), and we call this function \( f'(x) \) or \( \frac{df}{dx} \).

left-sided and right-sided derivatives; a function is differentiable at \( x \) if and only if it has both a left side and right side derivative at \( x \).

examples (and non-examples and surprises)

Differentiability implies continuity

7.1 The Mean Value Theorem

(surprising snail introduction)

Rolle’s Theorem

Mean-value theorem

Taylor’s theorem

7.2 More about differentiability

Example of a function which is continuous everywhere in \([0,1]\) but is differentiable nowhere

The spaces \( C^n[a,b] \) and \( C^\infty[a,b] \)

7.3 Exercises

1. a. Let \( f \) be a polynomial with real coefficients which possesses a real root. Then show that there is some \( p \in \mathbb{R} \) so that \( f(p) = f'(p) \).

b. (Generalization due to Daniel Brice and Jorge Garcia) Let \( f \) be a polynomial as in part a. Then prove that there is some \( p \in \mathbb{R} \) so that for any \( n \in \mathbb{N} \)

\[
\sum_{k=0}^{n} (-1)^k \binom{n}{k} f^{(k)}(p) = 0.
\]
Remember that $f^{(n)}(x)$ is the $n^{\text{th}}$ derivative of $f$ at $x$ and $f^{(0)}(x) = f(x)$.

8 The Riemann-Stieltjes integral

A generalization of the Riemann integral - the Riemann integral is a special case

$|\int f \, dx| \leq \int |f| \, dx$ (integral analogue of triangle inequality)

FTC - both forms (Riemann integral)

9 Function Spaces

9.1 linear spaces

**Definition 9.1.** Let $\mathbb{F}$ be a field of scalars and $X$ a set equipped with an addition and a scalar multiplication so that for any $x, y \in X$ and $\alpha \in \mathbb{F}$, $x + y \in X$ and $\alpha x \in X$. Then we say $X$ is a $\mathbb{F}$-linear space if

1. there is an element $0 \in X$ so that for every $x \in X$, $0 + x = x$;
2. for every $x \in X$, there is an element $-x \in X$ so that $x + (-x) = 0$;
3. for every $x, y, z \in X$, $(x + y) + z = x + (y + z)$;
4. for every $x, y \in X$, $x + y = y + x$;
5. for every $\alpha \in \mathbb{F}$ and $x \in X$, there is an element $\alpha x \in X$;
6. for every $x \in X$, $1x = x$;
7. if $\alpha, \beta \in \mathbb{F}$ and $x \in X$, then $(\alpha \beta)x = \alpha(\beta x)$;
8. if $\alpha, \beta \in \mathbb{F}$ and $x \in X$, then $(\alpha + \beta)x = \alpha x + \beta x$; and
9. if $\alpha \in \mathbb{F}$ and $x, y \in X$, then $\alpha(x + y) = \alpha x + \alpha y$.

Sometimes we call $X$ a linear space if the field is understood, or at times we call $X$ a vector space.

The first four properties imply that $X$ is an abelian group. Properties 5-7 tell us that there is a scalar multiplication, and that it is compatible with the multiplication on $\mathbb{F}$. The last two properties tell us that scalar multiplication distributes over both addition in $X$ and addition in $\mathbb{F}$; they tell us how addition and multiplication interact.

For our purposes, our field $\mathbb{F}$ will either be $\mathbb{C}$ or $\mathbb{R}$, although other fields like $\mathbb{Q}$ could be used. If $X$ is a $\mathbb{F}$ vector space, then we also say that $X$ is a vector space over $\mathbb{F}$.

**Fact:** If $X$ is a $\mathbb{F}_1$-vector space and $\mathbb{F}_2 \subseteq \mathbb{F}_1$ is a subfield, then $X$ is also a $\mathbb{F}_2$-vector space. A consequence of this is that every complex vector space is also a real vector space. Also, every real vector space is a $\mathbb{Q}$-vector space.
Every field is a vector space over itself; the scalar multiplication is just the regular field multiplication.

**Definition 9.2.** Let $X$ be a vector space over $\mathbb{R}$ or $\mathbb{C}$. Let $\| \cdot \| : X \to \mathbb{R}$ be a function so that

1. for any $x \in X$, $\| x \| \geq 0$, and $\| x \| = 0$ if and only if $x = 0$;
2. for any $\alpha \in \mathbb{F}$ and $x \in X$, $\| \alpha x \| = |\alpha| \| x \|$; and
3. for any $x, y \in X$, $\| x + y \| \leq \| x \| + \| y \|$.

Then we say the function $\| \cdot \|$ is a **norm** on $X$, and that $X$ is a **normed linear space**.

Normed linear spaces are nice because we can define a distance on such a space which is compatible with the linear structure. In particular, if we define a distance $d(x, y) = \| x - y \|$, then this distance is nice.

**Theorem 9.3.** The function $d$ defined above is a metric.

**Proof.** The function $d$ is positive definite since $d(x, y) = \| x - y \|$ is nonnegative and is equal to zero if and only if $x - y = 0$. It is symmetric since $\| x - y \| = \| y - x \|$. The triangle inequality is established because for any three points $x, y, z \in X$, we have $x - z = (x - y) + (y - z)$, and by the third property of a norm, $\| x - z \| \leq \| x - y \| + \| y - z \|$. \qed

We often call this $d$ to be the metric induced by the norm $\| \cdot \|$. One especially nice thing about norms is that the distances they induce are compatible with the linear structure. The vector addition and scalar multiplication are continuous operations on $X \times X$ and $\mathbb{F} \times X$ respectively. One way that we can express this is that open sets are translation invariant, that is for any point $x \in X$ and open set $U \subseteq X$, the set

$$x + U = \{ x + u | u \in U \}$$

is also an open set in $X$. Additionally, if $U$ and $V$ are open, then $U + V$ is also open in $X$. This does not necessarily extend to closed sets since in many spaces exist closed sets $F$ and $G$ so that $F + G$ is not closed. Here we are taking the sum of two sets $A + B = \{ a + b | a \in A \text{ and } b \in B \}$ in the usual way.

**Example 9.4.**

a. In $\mathbb{R}^n$, if we let $x = (x_1, x_2, \ldots, x_n)$ and

$$\| x \|_2 = (x_1^2 + x_2^2 + \cdots + x_n^2)^{\frac{1}{2}},$$

then this is a norm on $\mathbb{R}^n$. We can call this the $L^2$ norm, and it induces the $d_2$ metric on $\mathbb{R}^n$.

b. Similarly, in $\mathbb{R}^n$, if we define $\| x \|_1 = \sum_{k=1}^{n} |x_k|$, then this norm induces the $d_1$ metric on $\mathbb{R}^n$. 

69
c. Also \( \|x\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|) \) is a norm which induces the \( d_\infty \) metric.

d. The three metrics we have discussed on \( C[a, b] \) are all induced from norms. For a function \( f \in C[a, b] \), the three norms

\[
\|f\|_1 = \int_a^b |f(x)| \, dx
\]
\[
\|f\|_2 = \left( \int_a^b |f(x)|^2 \, dx \right)^{\frac{1}{2}}
\]
\[
\|f\|_\infty = \sup_{x \in [a, b]} |f(x)|
\]

induce the \( d_1 \), \( d_2 \), and \( d_\infty \) metrics respectively. We call these norms the \( L^1 \), the \( L^2 \) and \( L^\infty \) norms respectively.

functionals
linear functionals

9.2 Exercises

1. Let \( X \) be a normed linear space with \( x \in X \) and \( U \) an open set in the metric derived from the norm. Prove that \( x + U \) is open in \( X \).

References


