MATH 521, WEEKS 4 & 5:
Metric Spaces, Euclidean Spaces

1 Further Set Theory

We will pause briefly to introduce some more notions from set theory which will factor significantly moving forward. In particular, we will be interested in appreciating the differences which can arise when consider intersections and unions of *finite* versus *infinite* sets.

First of all, we generalized our notion of a family of sets. We have already seen *finite* families of sets $\{S_n\}_{n\in\{1,\ldots,N\}}$ and *countably infinite* families of sets $\{S_n\}_{n\in\mathbb{N}}$. We can also, however, consider *uncountable* families of sets $\{S_\alpha\}_{\alpha\in A}$ by choosing A to an uncountable set (for instance, an interval). Our definitions remain analoguous to before. We have

$$x \in \bigcup_{\alpha\in A} S_\alpha \tag{1}$$

if $x \in S_\alpha$ for some $\alpha \in A$, and

$$x \in \bigcap_{\alpha\in A} S_\alpha \tag{2}$$

if $x \in S_\alpha$ for all $\alpha \in A$. We will see that there can be significant differences between finite and infinite (countable or uncountable) sets with regards to intersection and union.

Example 1: Consider the family of sets $\{S_n\}_{n\in A}$ where

$$S_n = \left\{ x \in \mathbb{R} \mid 0 < x < \frac{1}{n} \right\} = \left(0, \frac{1}{n}\right).$$

Determine the union (1) and the intersection (2) when (a) A is any finite subset of $\mathbb{N}$, and (b), $A = \mathbb{N}$.

Solution: For any finite subset $A \subset \mathbb{N}$, we have that $\min(A) = m$ and $\max(A) = M$ for some $m, M \in \mathbb{N}$. We also clearly have that $S_m \subseteq S_n$ for all $m > n$. It follows directly that

$$\bigcup_{n\in A} S_n = S_m = \left(0, \frac{1}{m}\right), \quad \text{and} \quad \bigcap_{n\in A} S_n = S_M = \left(0, \frac{1}{M}\right).$$

Now consider $A = \mathbb{N}$. Notice that we still have $S_m \subseteq S_n$ for all $m > n$ so that

$$\bigcup_{\alpha \in \mathbb{N}} S_n = S_1 = (0, 1).$$

The intersection is not quite so clear since there is no element N such that $\max(A) = N$. Consider the question directly: What elements $x \in \mathbb{R}$ are in *every* set S_n, $n \in \mathbb{N}$? We quickly realize that we cannot have any strictly positive number $x > 0$ because there will always be an n sufficiently large so that $0 < 1/n < x$. It follows that

$$\bigcap_{n \in \mathbb{N}} S_n = \emptyset.$$

The moral of the story is that the infinite intersection is empty even though *every* finite intersection from the set is nonempty. In fact, every finite intersection has an infinite number of points (since it is an interval).

Example 2: Determine the union (1) and the intersection (2) for the family of sets

$$S_\alpha = \{x \in \mathbb{R} \mid -2 + \alpha \leq x < 1 + \alpha\} = [-2 + \alpha, 1 + \alpha)$$

where (a) A is any finite set of points from the interval $(0, 1)$, and (b) $A = (0, 1)$.

Solution: This example is slightly different than the previous one in that the index set A is an interval rather than a countable set. Nevertheless, we can draw individual sets in exactly the same way as before. For example, we have

$$S_{0.5} = [-1.5, 1.5), \quad S_{\frac{1}{\sqrt{2}}} = \left[-2 + \frac{1}{\sqrt{2}}, 1 + \frac{1}{\sqrt{2}}\right), \text{ etc.}$$

Now consider the intersection and union for any finite set of points drawn from the interval $(0, 1)$. That is to say, consider a set $A = \{\alpha_1, \ldots, \alpha_N\}$ where $\alpha_i \in (0, 1)$ for all $i = 1, \ldots, N$. From this set, we define $\min(A) = m$ and $\max(A) = M$ where $m \leq M$ so that the extreme sets are

$$S_m = [-2 + m, 1 + m) \quad \text{and} \quad S_M = [-2 + M, 1 + M).$$

Notice that $0 < m \leq M < 1$ implies that $-2 < -2 + m \leq -2 + M < -1$ and $1 < 1 + m < 1 + M < 2$ so that

$$-2 < -2 + m \leq -2 + M < 1 + m \leq 1 + M < 2.$$

It follows that we have

$$\bigcup_{n=1}^{N} S_{\alpha_n} = [-2+m, 1+M) \quad \text{and} \quad \bigcap_{n=1}^{N} S_{\alpha_n} = [-2+M, 1+m).$$

Now consider the union and intersection for the interval $A = (0,1)$ (which is infinite!). We notice that we do not have a maximal or minimal element within the set A. Instead, we identify the extreme sets

$$S_0 = [-2,1) \quad \text{and} \quad S_1 = [-1,2).$$

We notice that these sets are *not* contained in the family of sets $\{S_\alpha\}$ but anything arbitrarily close to them is, so long as we choosen the interval to approach from the correct side.

Consider the union. We notice that sets approaching the extreme sets S_0 and S_1 cover everything in the interval $(-2,2)$. (This can be seen by noticing that, for any small $\epsilon > 0$, there is an $\alpha \in (0,1)$ so that $-2 < -2+\alpha < -2+\epsilon$, and similarly for the other end point.) We might wonder about whether we can include the endpoint $x = -2$. After all, it is in the extreme set. In this case, we cannot, because $\alpha > 0$ implies $-2 < -2+\alpha$ so that $-2 \notin S_\alpha$ for any $\alpha \in (0,1)$. We have that the desired set is

$$\bigcup_{\alpha \in A} S_\alpha = (-2,2).$$

Now consider the intersection. We start by noticing that every $x \in S_0 \cap S_1 = [-1,1)$ is in the intersection. Consider, however, the endpoint $x = 1$. The have only excluded this point because it is not in S_0; however, S_0 is not in the family of subsets we are considering. We can see that for any $\alpha > 0$, we have $1 < 1+\alpha$ so that $1 \in S_\alpha$ for all $\alpha \in (0,1)$. We therefore have

$$\bigcap_{\alpha \in A} S_\alpha = [-1,1].$$

We should pause for a moment to consider how strange this truly is. We have just seen that, for *every* finite family drawn from the set, the union and intersection corresponded to an interval which has the same form as the original interval—one endpoint is included, one is not. For the interval corresponding to the infinite union, however, neither endpoint was included, while for the infinite intersection interval, both endpoints were included. So strange things can happen (and properties can change!) when we consider infinite families! □

We will see later on that we *will* be able to prove general properties about infinite intersections. For now, though, we will content ourselves with the following classical result.

Theorem 1.1 (DeMorgan's Laws). *Suppose $S_\alpha \subseteq X$, $\alpha \in A$, is a family of subsets of a universal set X. Then*

$$\left(\bigcup_{\alpha \in A} S_\alpha\right)^c = \bigcap_{\alpha \in A} S_\alpha^c$$

and

$$\left(\bigcap_{\alpha \in A} S_\alpha\right)^c = \bigcup_{\alpha \in A} S_\alpha^c$$

Proof. Take $x \in (\cup_{\alpha \in A} S_\alpha)^c$. It follows that

$$\begin{aligned} &\implies x \notin \bigcup_{\alpha \in A} S_\alpha \\ &\implies x \notin S_\alpha \text{ for any } \alpha \in A \\ &\implies x \in S_\alpha^c \text{ for all } \alpha \in A \\ &\implies x \in \bigcap_{\alpha \in A} S_\alpha^c. \end{aligned}$$

It follows that $(\cup_{\alpha \in A} S_\alpha)^c \subseteq \cap_{\alpha \in A} S_\alpha^c$. The other implication follows by the same steps above in reverse, so that $(\cup_{\alpha \in A} S_\alpha)^c = \cap_{\alpha \in A} S_\alpha^c$. The other law follows similarly. □

2 Metric Spaces

So far, we have spent the majority of our time investigating *ordered systems* and, in particular, *ordered fields*. We understand the relationships $>$, $<$, $+$, $\cdot$, etc., and know what it means for a set to have a maximum, minimum, supremum, and infimum. We also understand the real numbers $\mathbb{R}$ as the *completion* of the rational numbers $\mathbb{Q}$ with respect to least upper bound property (the property of containing the least upper bound of each subset, a property which $\mathbb{Q}$ does not possess).

We should not grow too comfortable, however, with the idea that *ordered sets* are the best structured sets we can imagine. For instance, consider the question of relating points on a *map* (say, different cities or landmarks). We might be interested in such questions as:

- How far is Madison from Milwaukee?
- Is Madison closer to Milwaukee or Chicago?
- If you travelled from Madison to Milwaukee, then to Chicago, how far would you have travelled? Is this shorter than traveling via a direct route?

These questions seem like questions we should be able to handle by the order operator. After all, we very clearly need some notion of "close" and "far" which necessitates and order where "far" is *greater* than "close".

So what's missing? Well, the topology (think topography, but in general mathematical spaces) of map is *not the same* as the topology of the real number line. In order to define where Madison is on a *map*, we need *two* coordinates rather than just one. More specifically, we need an ordered pair (a, b) where a corresponds to the latitude and b corresponds to the longitude (both of which could be considered to be real numbers, if we liked).

If we want to model anything which has any real-world meaning, we had best be able to accommodate this mathematically! We have the following definition.

Definition 2.1. *For any $n \in \mathbb{N}$, we define the n-dimensional **Cartesian product** over the reals to be*

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R}, i = 1, \ldots, n\}.$$

*The elements x_i, $i = 1, \ldots, n$, are called **coordinates**.*

We will commonly refer to the elements of $\mathbb{R}^n$ as **vectors** (or simply **points**, depending on the context) and denote them with boldface or vector notation, i.e. $\mathbf{x} \in \mathbb{R}^n$ or $\vec{x} \in \mathbb{R}^n$. We have the following definition.

Definition 2.2. *We will say that a Cartesian product $\mathbb{R}^n$ is a **Euclidean space** if, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and any $\alpha \in \mathbb{R}$, we have the following operations:*

(1) ***Vector addition:*** $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)$

(2) ***Scalar multiplication:*** $\alpha\mathbf{x} = (\alpha x_1, \ldots, \alpha x_n)$

We additionally endow Euclidean spaces with the following:

(3) ***Vector norm:*** $$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

(4) ***Dot product*** *(or* ***inner product****):* $$\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

Properties (1) and (2) turn $\mathbb{R}^n$ into a **vector space**. The norm (3) is a measure of the *magnitude* of elements in $\mathbb{R}^n$ (turning $\mathbb{R}^n$ into a **normed space**) while the inner product is related to the relative *orientation* of elements (turning $\mathbb{R}^n$ into an **inner product space**). We will not deal with these concepts significantly in this course, except to notice that $\mathbf{x}\cdot(\mathbf{y}+\mathbf{z}) = \mathbf{x}\cdot\mathbf{y}+\mathbf{x}\cdot\mathbf{z}$ (distribution) and $\mathbf{x}\cdot\mathbf{y} = \mathbf{y}\cdot\mathbf{x}$ (commutativity). (Note that we have already used the operator "$\cdot$" in our field axioms, but that $\mathbb{R}^n$ is *not* a field because the multiplication operator does not take two elements of $\mathbb{R}^n$ into $\mathbb{R}^n$. The context should make it clear which operation is meant.)

Returning to our previous example, it is clear that the space we are interested in for describing the location of our cities on a map is the two-dimensional Euclidean space

$$\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}.$$

So how do we know which city is closest to which city? We still have a non-trivial problem to overcome. We know that $\mathbb{R}$ is an ordered set, so we can tell which city is farther to the west or the east, or *independently* which is farther to the north or the south. But we want some *combination* of these two directions.

We might think that we could simply extend the order operator as we have defined it, but we would get quickly frustrated. In fact, there is no order operator which turns even $\mathbb{R}^2$ into an ordered set as we have defined it. (This should not be that surprising, since an order requires that the elements of the set can be arranged in some way from least to greatest. Imagine running your finger along a page of paper, picking up points from lowest to greatest, and being asked to not stop until all of the points have been captured. Good luck!)

Of course, we do not panic for too long. We have known since grade school that the distance between two points $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ is given by the Pythagorean theorem as

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

This is known as the **Euclidean distance** between two points. We can now very quickly determine how far Madison is from Milwaukee, or Chicago, etc.

It should nevertheless be slightly unsettling that mathematical notions defined so far as insufficient to handle this case. In order to overcome this, we introduce the following.

Definition 2.3. *Consider the pairing* (X, d) *where* X *is a set and* $d(x, y)$ *is an operator which takes elements from* X *into* $\mathbb{R}$*. We will say that* (X, d) *is a* ***metric space*** *with* ***metric*** *(or* ***distance****)* $d(x, y)$ *if*

1. $d(x, y) \geq 0$ *for all* $x, y \in X$*, and* $d(x, y) = 0$ *if and only if* $x = y$*;*

2. $d(x, y) = d(y, x)$ *for all* $x, y \in X$*; and*

3. $d(x, z) \leq d(x, y) + d(y, z)$ *for all* $x, y, z \in X$*.*

Notice here that $\mathbb{R}$ is an ordered set while the original space X may not be. So what we are doing here is recovering a sense of order by *mapping* elements in the original set into a set we know is ordered—namely, the real numbers.

We can interpret the axioms for "distance" in this way:

1. The distance between two points is not negative and is zero if and only if the two points coincide.

2. The distance from one point to another is the same as the distance back.

3. The distance from one point to any other is greater than the combined distance of going through an intermediary point (i.e. the direct route is the shortest route).

In other words, this operation obeys the conventions we normally associate (without thought!) to our colloquial notion of "distance". The third property is commonly called the *triangle inequality* due to its geometric interpretation (see Figure 1).

Note (for those taking further analysis courses): We will primarily consider Euclidean spaces in this course. It should be noted, however, that, while every vector space with a norm and/or an inner product is a metric space, not every metric space has a norm or an inner product. The general sequence, from weakest to strongest is:

$$\text{Metric spaces} \supseteq \text{Normed vector spaces} \supseteq \text{Inner product spaces.}$$

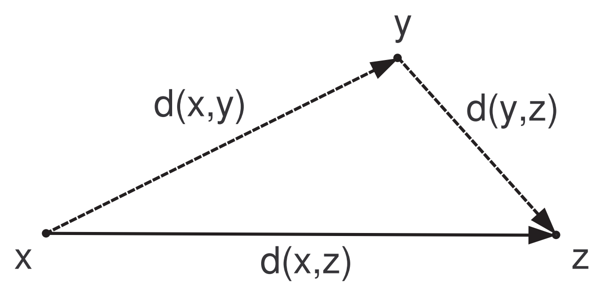

Figure 1: Graphic interpretation of the triangle inequality. We require that $d(x, z) \leq d(x, y) + d(y, z)$.

That is to say, every inner product space has a norm, and every normed vector space has a metric, but the converses do not necessarily hold.

Theorem 2.1. *The Euclidean space* $(\mathbb{R}^n, d_2)$ *where*

$$d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

is a metric space.

What we need to do is show that show that the conventionally defined distance between two points is a distance in the sense of Definition 2.3. We expect, based on our intuition, that this is true but may not have ever stopped to consider how we might explicitly *prove* it. We will see that it is not quite as trivial as we might expect, especially proving the triangle inequality. Before we proceed with the proof of this result, therefore, we introduce the following classical result.

Theorem 2.2 (Cauchy-Schwarz Inequality)**.** *Suppose* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*. Then*

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\| \, .$$

Proof. Note that the compact statement above is not fully illustrative of what is really happening. Expanding out the terms for $\mathbf{x} \cdot \mathbf{y}$, $\|\mathbf{x}\|$, and $\|\mathbf{y}\|$, we have

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{j=1}^{n} y_j^2}$$

We prove this result, we cannot simply expand the left-hand side and rearrange to obtain the right-hand side. Instead, we start with the observation that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i y_j - x_j y_i)^2 \geq 0$$

where the positivity follows from the basic field properties of squaring real numbers. We now expand this term. We have

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i y_j - x_j y_i)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i y_j - x_j y_i)(x_i y_j - x_j y_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_i^2 y_j^2 + x_j^2 y_i^2 - 2x_i y_i x_j y_j\right) \\
&= 2\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right) - 2\left(\sum_{i=1}^{n} x_i y_i\right)^2 \\
&= 2\left\|\mathbf{x}\right\|^2\left\|\mathbf{y}\right\|^2 - 2\left(\mathbf{x}\cdot\mathbf{y}\right)^2 \geq 0.
\end{aligned}$$

It follows immediately that $(\mathbf{x}\cdot\mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2$. The desired inequality follows from taking the square root, and we are done. □

Now consider the original claim.

Proof of Theorem 2.1. We prove the three properties of Definition 2.3 hold.

Property 1: We clearly have $d_2(\mathbf{x},\mathbf{y}) \geq 0$ since the square root is never negative. Since $\mathbf{x} = \mathbf{y}$ if and only if $x_1 = x_2$ and $y_1 = y_2$, it follows that $d_2(\mathbf{x},\mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$.

Property 2: It is clear that we have

$$d_2(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} = d_2(\mathbf{y},\mathbf{x}).$$

Property 3: It will be convenient to recast the metric $d_2(\mathbf{x},\mathbf{y})$ in terms of the Euclidean norm. We can quickly see that, for any $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, we have

$$d_2(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}. \tag{3}$$

We furthermore have that

$$\mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^{n} x_i \cdot x_i = \left(\sqrt{\sum_{i=1}^{n} x_i^2} \right)^2 = \|\mathbf{x}\|^2 .$$

That is to say, we can recast the Euclidean metric as a norm, and recast the norm as an inner product. (This is not true of all metric spaces but for Euclidean spaces it will often be a helpful notational shortcut!)

We now use basic properties of norms and inner products (and the Cauchy-Schwarz inequality) to prove that

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{4}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We have

$$\begin{aligned}
\|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) \\
&= \mathbf{x} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} \\
&= \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2 \\
&\leq \|\mathbf{x}\|^2 + 2|\mathbf{x} \cdot \mathbf{y}| + \|\mathbf{y}\|^2 \\
&\leq \|\mathbf{x}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 \qquad \text{(Cauchy-Schwarz)} \\
&= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.
\end{aligned}$$

It follows after taking the square root that $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Terrific, but what does all of this have to do with the triangle inequality? We have verified that $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, but what we want to verify is that

$$d_2(\mathbf{x}, \mathbf{z}) \leq d_2(\mathbf{x}, \mathbf{y}) + d_2(\mathbf{y}, \mathbf{z}).$$

We recall that this is equivalent to

$$\|\mathbf{x} - \mathbf{z}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{z}\| . \tag{5}$$

We immediately suspect that the similarity between (4) and (5) is not coincidal. If we switch out $\mathbf{x}$ for $\mathbf{x} - \mathbf{y}$ and $\mathbf{y}$ for $\mathbf{y} - \mathbf{z}$, we have

$$\|\mathbf{x} - \mathbf{z}\| = \|(\mathbf{x} - \mathbf{y}) + (\mathbf{y} - \mathbf{z})\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} + \mathbf{z}\|$$

by the previous result, so that the triangle inequality has been shown. □

There are a few notes worth making about the (lengthy!) process we just went through.

- It probably seemed like a lot of work to introduce the notions of an "inner product" and a "norm" to prove a simple inequality. It turns out that, even for $\mathbb{R}^2$, it is *extremely* cumbersome to prove the triangle inequality directly, i.e. without resorting to norm and inner product notation. (Try it!)

- It is only a slight generalization to extend the proof we just used to show that the triangle inequality to *all* spaces where the metric follows from the norm, and the norm follows from the inner product. In particular, it can be extended to the complex field $\mathbb{C}$. We will not consider these spaces, or the general inner product operator, in significant depth in this course, but we should feel comfortable using $\mathbf{x} \cdot \mathbf{y}$ and $\|\mathbf{x}\|$ as short-hand notation when dealing with the Euclidean space $\mathbb{R}^n$.

3 Metric Spaces on $\mathbb{R}^n$

We probably feel supremely confident at this point. We have introduced a new mathematical concept, that of a *metric*, and it turned out to be exactly the same as the conventionally defined distance between two points. So why bother with this abstraction?

The reason is obvious after a little thought: the notion of measuring the *distance* between two objects of interest depends on the context of the problem involved. For instance, consider the problem in image processing of how "close" a processed image (say, a digital copy) is to the original. How do we measure the "distance" between two images? This is not a trivial problem.

Another problem, which is frequently encountered in statistics, is that of determining a model which "most closely" fits the given data (i.e. for which the "distance" between the model and the observations is minimized). The Euclidean metric givens a reasonable "penalty" for component-wise differences—namely, it squares the difference. Depending on the context, however, it may be appropriate to either penalize component-wise differences more severely (i.e. pick a higher power) or not as much (i.e. pick a lower power). So another metric may be more appropriate depending on the context of the problem being investigated.

We now define the following metrics on $\mathbb{R}^n$:

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max_{i \in \{1,\ldots,n\}} |x_i - y_i|.$$

Theorem 3.1. *$(\mathbb{R}^n, d_1)$ and $(\mathbb{R}^n, d_\infty)$ are metric spaces.*

Proof. We first show that $d_1(\mathbf{x}, \mathbf{y})$ is a metric.

Property 1: We clearly have $d_1(\mathbf{x}, \mathbf{y}) \geq 0$ and $d_1(\mathbf{x}, \mathbf{y}) = 0$ if and only $\mathbf{x} = \mathbf{y}$.
Property 2: It is clear that

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} |y_i - x_i| = d_1(\mathbf{y}, \mathbf{x}).$$

Property 3: Recall that for any ordered set S, we have

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

from Assignment 2. Since $\mathbb{R}$ is an ordered set, we can replace x with $x_i - y_i$ and y with $y_i - z_i$ so that, for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, we have

$$\begin{aligned} d_1(\mathbf{x}, \mathbf{z}) &= \sum_{i=1}^{n} |x_i - z_i| \\ &= \sum_{i=1}^{n} |x_i - y_i + y_i - z_i| \\ &\leq \sum_{i=1}^{n} |x_i - y_i| + \sum_{i=1}^{n} |y_i - z_i| \\ &= d_1(\mathbf{x}, \mathbf{y}) + d_1(\mathbf{y}, \mathbf{z}). \end{aligned}$$

It follows that $(\mathbb{R}^n, d_1)$ so defined is a metric space. The proof for $(\mathbb{R}^n, d_\infty)$ is left as homework. □

So even on $\mathbb{R}^n$, we may define multiple metrics which satisfying Definition 2.3. It is worth emphasizing that the "distance" between two points in these metrics need not necessarily coincide. For instance, consider the points $\mathbf{x} = (0,0)$ and $\mathbf{y} = (1,1)$. We have

$$\begin{aligned} d_1(\mathbf{x},\mathbf{y}) &= |x_1 - y_1| + |x_2 - y_2| = 2, \\ d_2(\mathbf{x},\mathbf{y}) &= \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2} = \sqrt{2}, \text{ and} \\ d_\infty(\mathbf{x},\mathbf{y}) &= \max\{|x_1-y_1|, |x_2-y_2|\} = \max\{1,1\} = 1. \end{aligned}$$

So, even though each of these metrics give a measure of distance, they do in fact measure distance in a different way. We have the following interpretations:

1. The metric d_2 measures distance as the crow flies, so to speak. In this sense, it is our most natural notion of distance.

2. The metric d_1 measures distance *independently* in each direction (i.e. each x_i). For this reason, it is sometimes called the *taxicab* metric, since it measures distance as though $\mathbb{R}^n$ were partitioned into city blocks where buildings obstruct direct passage from one point to another unless they happen to lie on the same street.

3. This metric d_∞ is also sometimes called the *chessboard* metric. The analogy is made by considering how many steps it would take the king to move from one square to another on a chessboard. (Note that in a single step the the king may move to any adjacent square, including the diagonal.) If the king desires to move to the square one positive left, and three positions up, it will take $\max\{1,3\} = 3$ moves to get there. If the desired square is four positions left, and three positions up, however, it will take $\max\{4,3\} = 4$ moves to get there.

4 Other Metric Spaces

We are now somewhat familiar with metrics on the familiar Euclidean space $\mathbb{R}^n$. What other spaces could be define metrics on? Consider the following.

Example: Consider the space (I,d) where $I = \{[a,b] \mid a,b \in \mathbb{R}\}$ and, for all $I_1, I_2 \in I$,

$$d(I_1, I_2) = \max\left\{\max_{x\in I_1}\min_{y\in I_2}|x-y|, \max_{y\in I_2}\min_{x\in I_1}|x-y|\right\}. \tag{6}$$

Theorem 4.1. *(I,d) is a metric space.*

Proof. Homework! ☐

We should pause to consider exactly what we are talking about, and how this differs from the previous examples. The set I is the set of all *intervals* of the form $[a,b] \subseteq \mathbb{R}$. So when we talk about the "distance" between two elements in the set, are are talking about the distance between two *intervals*. (To emphasize again, the objects or interest are not the *points* $x \in I_1$ or $y \in I_2$ where $I_1, I_2 \in I$, but the *intervals* I_1 and I_2 themselves!)

The "distance" (6) between two intervals is also a little strange. We want a maximum, where we choose over the following two objects. The first objects tells us to consider an $x \in I_1$ (the first inverval), and then choose the $y \in I_2$ (the second interval) which is *closest* to x. We then want to take the $x \in I_1$ for which this closest element is the greatest distance away. For the second element, we reverse the process. That is to say, we take an element $y \in I_2$, choose the closest element $x \in I_1$, and then choose the $y \in I_2$ where this value is the greatest. Despite the strange formulation, it is possible to show that this is a metric. It is a special case of the *Hausdorff metric*. (The full proof is challenging and a bit beyond the scope of the course, although we will do a little more work with this metric on the homework!)

Let's consider the following cases. Determine the "distance" between the following sets, as defined by the Hausdorff metric (6). (See Figure 2.)

1. $I_1 = [-1, 0]$ and $I_2 = [0.5, 1]$
2. $I_1 = [-1, 0]$ and $I_2 = [-0.5, 1]$
3. $I_1 = [-1, 1]$ and $I_2 = [-0.5, 1]$

Solution (1): We can clearly see that the $x \in I_1$ which is the farthest away from any element $y \in I_2$ is $x = -1$ (corresponding to $y = 0.5$). This gives the value

$$\max_{x \in I_1} \min_{y \in I_2} |x - y| = 1.5.$$

The $y \in I_2$ which is the farthest away from any element $x \in I_1$ is $y = 1$ (corresponding to $x = 0$) so that

$$\max_{y \in I_2} \min_{x \in I_1} |x - y| = 1.$$

It follows that

$$d(I_1, I_2) = \max\{1.5, 1\} = 1.5.$$

Solution (2): Notice that the closest element for any $x \in I_1$ or $y \in I_2$ in the overlap is zero. This is not, however, enough to make the distance between the sets zero! We have

$$\max_{x \in I_1} \min_{y \in I_2} |x - y| = 0.5$$

where $x = -1$ and $y = -0.5$ (no element in I_2 is closer to $x = -1$ than $y = -0.5$). We also have

$$\max_{y \in I_2} \min_{x \in I_1} |x - y| = 1$$

where $y = 1$ and $x = 0$ (no element in I_1 is closer to $y = 1$ than $x = 0$). It follows that

$$d(I_1, I_2) = \max\{0.5, 1\} = 1.$$

Solution (3): Notice that we have $I_2 \subset I_1$. It follows immediately that

$$\max_{y \in I_2} \min_{x \in I_1} |x - y| = 0$$

because, for every $y \in I_2$, there is an $x \in I_2$ such that $y = x$ (so that $|x - y| = 0$). We *still* may not conclude that the distance between the sets is zero! This is because we have

$$\max_{x \in I_1} \min_{y \in I_2} |x - y| = 0.5$$

where $x = -1$ and $y = -0.5$. It follows that

$$d(I_1, I_2) = \max\{0.5, 0\} = 0.5.$$

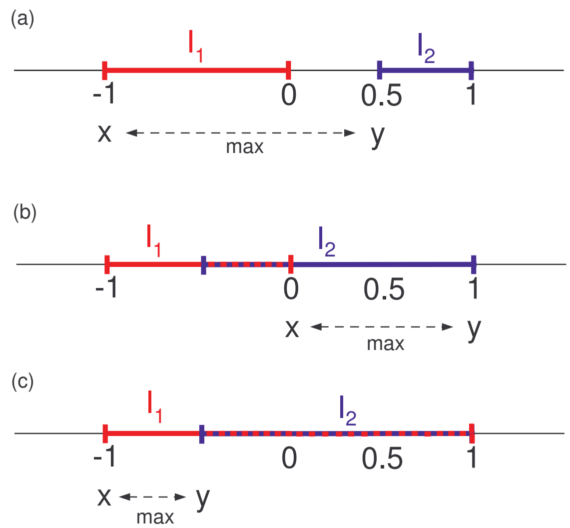

Figure 2: Illustration of Hausdorff metric on the intervals I_1 and I_2 given in the example.