Welcome to the first episode of the exciting metric space saga! In this trilogy, we generalize many of the results we’ve seen to $\mathbb{R}^k$ or more general spaces called metric spaces. Even though this is more abstract, it really gives us a nice perspective into what concepts in this course are essential, and which ones are not.

1. Metric Spaces

Video: What is a Metric Space?

(Optional) Full Version: Metric Spaces

The single, most important identity with absolute values that we have learned so far is the triangle inequality, which states:

\[
|x + y| \leq |x| + |y|
\]

And one of its consequences was:

Corollary

\[
|a - c| \leq |a - b| + |b - a|
\]

Interpretation: The third leg of a triangle is shorter than the sum of the other two legs.
There are other properties of $|x|$ that we have used, even though we didn’t really think of them:

For all $x, y, z \in \mathbb{R}$,

1. $|x - y| \geq 0$
2. $|x - y| = 0 \iff x = y$
3. $|x - y| = |y - x|
4. $|x - z| \leq |x - y| + |y - z|$

**Main Idea:** What if we forget everything about $\mathbb{R}$ and absolute values, except the four properties above? Then we get an extremely useful object called a **metric space**:
Definition:
If \( S \) is any set, then \([S, d]\) is called a metric space if the following 4 properties hold. Here \( x, y, z \in S \)

1. \( d(x, y) \geq 0 \)
2. \( d(x, y) = 0 \iff x = y \)
3. \( d(x, y) = d(y, x) \)
4. \( d(x, z) \leq d(x, y) + d(y, z) \)

\( d \) is any function from \( S \times S \) to \( \mathbb{R} \)

Notice the similarity between \( d(x, y) \) and \( |x - y| \), so in fact \( d \) (called a metric) is just a generalization of the absolute value.

To show how useful and powerful this concept is, let me give you 10 examples of metric spaces.

Note: The first 5 examples are important for your homework, but you can skip the last 5 if you wish, although they are pretty cool.

Example 1:
\( (\mathbb{R}, d) \) with \( d(x, y) = |x - y| \)

Example 2:
\( (\mathbb{R}^2, d_2) \)

\[
d_2((x_1, x_2), (y_1, y_2)) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}
\]
Example:

$$d_2((1, 2), (3, 4)) = \sqrt{(3 - 1)^2 + (4 - 2)^2} = \sqrt{8}$$

**Note:** Because this is such a natural distance function on $\mathbb{R}^2$, from now on we’ll write $d$ instead of $d_2$.

**Note:** This can be generalized to $\mathbb{R}^k$:

$$d_2((x_1, \ldots, x_k), (y_1, \ldots, y_k)) = \sqrt{(y_1 - x_1)^2 + \cdots + (y_k - x_k)^2}$$

**Example 3:**

$$(\mathbb{R}^2, d_1)$$

$$d_1((x_1, x_2), (y_1, y_2)) = |y_1 - x_1| + |y_2 - x_2|$$
In other words, just add the sum of the lengths of the legs of the triangle.

**Note:** This is sometimes called the taxicab (or Manhattan) metric. Because taxicabs in New York can’t just go diagonally from \((x_1, x_2)\) to \((y_1, y_2)\) without crashing into buildings, they have to go right, and then up.

**Example 4:**

\[
(\mathbb{R}^2, d_\infty)
\]

\[
d_\infty((x_1, x_2), (y_1, y_2)) = \max \{|y_1 - x_1| + |y_2 - x_2|\}
\]

**Note:** In other words, just calculate the length of the biggest leg.
Example 5: Discrete Metric

\((\mathbb{R}, d)\) with

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

(This is Example 10 in the video)

In other words, with the metric \(d\), all the points in \(\mathbb{R}\) are distance 1 apart. Freaky, isn’t it? But it’s a great source of counterexamples!
Note: The discrete metric seems weird for $\mathbb{R}$, but is more natural in other examples:

**Example:** $S = \{1, 2, 3\}$ with the discrete metric. Then $S$ is just an equilateral triangle!

Optional Example 6:

$S =$ Set of bounded sequences in $\mathbb{R}$ with

\[d((s_n), (t_n)) = \sup \{|s_n - t_n| \mid n \in \mathbb{N}\}\]
In other words, look at the largest possible difference between $s_n$ and $t_n$.

Note: The following is NOT a metric on $S$:

$$d((s_n), (t_n)) = \sum_{n=1}^{\infty} |s_n - t_n|$$

Because $d((s_n), (t_n))$ might be $\infty$ (for instance with $(s_n) = (1, 1, 1, \ldots)$ and $(t_n) = (0, 0, 0, \ldots)$), but for a metric we must have $d(x, y) < \infty$ for all $x$ and $y$.

Optional Example 7:

$$S = \text{Continuous functions on } [a, b] \text{ (see Chapter 3) with}$$

$$d(f, g) = \max \{|f(x) - g(x)| \mid x \in [a, b]\}$$

This is a continuous analog of the previous example
Optional Example 8:

Same, but this time

\[ d(f, g) = \int_a^b |f(x) - g(x)| \, dx \]
Optional Example 9:

Same, but this time

\[ d(f, g) = \sqrt{\int_a^b |f(x) - g(x)|^2 \, dx} \]

This is a very natural metric on \( S \) if you remember that an integral is just a sum (so here we take the square root of the sum of squares). It’s also nice because \( S \) becomes a Hilbert space (in case you know what that is).

Optional Example 10:

If \( A \) and \( B \) are two subsets of \( \mathbb{R} \) (or of any metric space), then

\[ d(A, B) = \inf \{|a - b| \mid a \in A, b \in B\} \]

(Analogy: Think like two lovebirds on two different continents \( A \) and \( B \) who try to communicate with each other).
So you can even measure how far whole sets are, how cool is that?

**Take-away:** *Everything* we’re going to show in this trilogy holds for **ALL** 10 examples at once, so we’re really killing 10 birds with one stone! *THIS* is the power of abstract mathematics!

## 2. Convergence

**Video:** Convergence in $\mathbb{R}^k$

The neat thing about metric spaces is that it’s really easy to generalize the notion of convergence to those spaces.

### Recall:

If $(s_n)$ is a sequence in $\mathbb{R}$, then $s_n \to s$ if for all $\epsilon > 0$ there is $N$ such that if $n > N$, then $|s_n - s| < \epsilon$.

It’s *exactly* the same for metric spaces, except you replace the absolute value with $d$!
**Definition:**

If \((S, d)\) is a metric space and \((s_n)\) is a sequence in \(S\), then \(s_n \to s\) if for all \(\epsilon > 0\) there is \(N\) such that if \(n > N\), then \(d(s_n, s) < \epsilon\).

That said, even though the definition is the same, the way of thinking is a bit different. Here the good region is a ball (Points that are at most \(\epsilon\) apart from \(s\)) instead of a strip and the sequence gets closer and closer to \(s\).

### 3. Convergence in \(\mathbb{R}^k\)

Since \(\mathbb{R}^k\) is a metric space, in order to define convergence in \(\mathbb{R}^k\), we just need to apply the definition above to the space \(\mathbb{R}^k\).
Notation:

(1) Points in $\mathbb{R}^k$ will be denoted by $(x_1, \ldots, x_k)$

(2) The distance between $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_k)$ is defined as

$$d((x_1, \ldots, x_k), (y_1, \ldots, y_k)) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_k - y_k)^2}$$

(3) Sequences in $\mathbb{R}^k$ will be written as

$$(x^{(n)}) = (x_1^{(n)}, \ldots, x_k^{(n)})$$

So the first term is $(x_1^{(1)}, \ldots, x_k^{(1)})$, the second term is $(x_1^{(2)}, \ldots, x_k^{(2)})$ and so forth.

With the notation as above, we can now define what it means for a sequence $(x^{(n)})$ in $\mathbb{R}^k$ to converge to $x$.

Definition:

If $(x^{(n)})$ is a sequence in $\mathbb{R}^k$, then we say $(x^n)$ converges to $x$ if for all $\epsilon > 0$ there is $N$ such that if $n > N$, then $d(x^n, x) < \epsilon$
Luckily we never have to use this to prove that \( x^{(n)} \) converges to \( x \), because we have the following really useful result.

**Motivation:** Consider the following sequence in \( \mathbb{R}^2 \):

\[
x^{(n)} = \left( \frac{1}{n}, e^{-n} \right) = \left( x^{(n)}_1, x^{(n)}_2 \right)
\]

Then \( x^{(n)} \to (0, 0) = x \).

How did we figure this out? Well, notice that \( x^{(n)}_1 = \frac{1}{n} \to 0 \) and \( x^{(n)}_2 = e^{-n} \to 0 \), and from this we concluded that \( x^{(n)} \to x = (0, 0) \).

In other words, to figure out if \( x^{(n)} \to x \), it is enough to check if each component \( x^{(n)}_1 \) and \( x^{(n)}_2 \) converges to \( x_1 \) and \( x_2 \) respectively, where \( x = (x_1, x_2) \). And in fact, this is always true:

**Theorem:**

If \( (x^{(n)}) = (x^{(n)}_1, \ldots, x^{(n)}_k) \) is a sequence in \( \mathbb{R}^k \), then

\[
(x^{(n)}) \to x \iff x^{(n)}_j \to x_j \quad \text{for each } j = 1, \ldots, k
\]

Where \( x = (x_1, \ldots, x_k) \)

**Note:** In terms of triangles, this makes sense: This just says that if the legs of the triangle are small, then the hypotenuse is small. And conversely, if the hypotenuse is small, then each leg is small as well. The picture below illustrates the case \( k = 2 \):
In order to prove this, we need a small lemma, which is kind of like a Squeeze Theorem, but for distances:

**Lemma: [Squeeze Theorem for Distances]**

If \((x_1, \ldots, x_k)\) and \((y_1, \ldots, y_k)\) are points in \(\mathbb{R}^k\), then for all \(j = 1, \ldots, k\), we have:

\[
|x_j - y_j| \leq d(x, y) \leq \sqrt{k} \max \{|x_1 - y_1|, \ldots, |x_k - y_k|\}
\]

Where \(x = (x_1, \ldots, x_k)\) and \(y = (y_1, \ldots, y_k)\)
This Lemma says two things: First of all, it says that the hypotenuse \( d \) of the triangle is bigger than each of its legs \( |x_j - y_j| \). On the other hand, the hypotenuse cannot be that big either. It is always smaller than a constant (Here \( \sqrt{k} \), think for instance \( \sqrt{2} \) in the case of \( \mathbb{R}^2 \)) times the biggest leg of the triangle. In the picture above, the red diagonal is smaller than the green vertical line.

**Proof:** On the one hand, we have:

\[
d((x_1, \ldots, x_k), (y_1, \ldots, y_k)) = \sqrt{\sum_{j=1}^{k} (x_j - y_j)^2} \geq \sqrt{(x_j - y_j)^2} = |x_j - y_j|
\]
On the other hand, let $M = \max \{|x_1 - y_1|, \ldots, |x_k - y_k|\}$, then each $|x_j - y_j| \leq M$

$$d((x_1, \ldots, x_k), (y_1, \ldots, y_k)) = \sqrt{\left(\frac{(x_1 - y_1)^2}{\leq M^2} + \cdots + \frac{(x_k - y_k)^2}{\leq M^2}\right)} \leq \sqrt{M^2 + M^2 + \cdots + M^2} \leq \sqrt{kM^2} = \sqrt{k} \max \{|x_1 - y_1|, \ldots, |x_k - y_k|\}$$

**Proof of Theorem:** We need to show that

$$(x^{(n)}) \to x = (x_1, \ldots, x_k) \iff x_j^{(n)} \to x_j \text{ (for each } j = 1, \ldots, k)$$

$(\Rightarrow)$ Let $\epsilon > 0$ be given, then since $(x^{(n)}) \to x$, there is $N$ such that if $n > N$, then $d(x^{(n)}, x) < \epsilon$. But then, for the same $N$, if $n > N$, then by the Lemma above, for each $j$, we have

$$\left| x_j^{(n)} - x_j \right| \leq d((x_1^{(n)}, \ldots, x_k^{(n)}), (x_1, \ldots, x_k)) = d((x^{(n)}), x) < \epsilon$$

Hence $|x_j^{(n)} - x_j| < \epsilon$ and therefore $x_j^n$ converges to $x_j$.

$(\Leftarrow)$ Let $\epsilon > 0$ be given.

Then, since $x_1^{(n)}$ converges to $x_1$, there is $N_1$ such that if $n > N_1$, then $|x_1^{(n)} - x_1| < \frac{\epsilon}{\sqrt{k}}$. 

Since \( x_2^{(n)} \) converges to \( x_2 \), there is \( N_2 \) such that if \( n > N_2 \), then
\[
\left| x_2^{(n)} - x_2 \right| < \frac{\epsilon}{\sqrt{k}}
\]
And in general, since \( x_j^{(n)} \) converges to \( x_j \), there is \( N_2 \) such that if \( n > N_j \), then
\[
\left| x_j^{(n)} - x_j \right| < \frac{\epsilon}{\sqrt{k}}
\]
Now if \( N = \max \{N_1, \ldots, N_k\} \), then if \( n > N \), by the Lemma, we have
\[
d(((x^{(n)}), x) = d((x_1^{(n)}, \ldots, x_k^{(n)}), (x_1, \ldots, x_k))
\leq \sqrt{k} \max \left\{ \left| x_1^{(n)} - x_1 \right|, \ldots, \left| x_k^{(n)} - x_k \right| \right\}
< \sqrt{k} \left( \frac{\epsilon}{\sqrt{k}} \right)
= \epsilon
\]
Hence \( (x^{(n)}) \) converges to \( x \) \( \Box \)

4. \( \mathbb{R}^k \) IS COMPLETE

Video: \( \mathbb{R}^k \) is complete

As a neat consequence of the above, we get that \( \mathbb{R}^k \) is complete. To do this, let’s quickly adapt the definition of Cauchy and completeness to \( \mathbb{R}^k \) (the same definition is valid for metric spaces)
**Definition:**

If \((x^{(n)})\) is a sequence in \(\mathbb{R}^k\), then we say \((x^{(n)})\) is **Cauchy** if, for all \(\epsilon > 0\), there is \(N\) such that if \(m, n > N\), then \(d((x^{(n)}), (x^{(m)})) < \epsilon\).

In other words, the terms of the sequence \(x^{(n)}\) eventually get closer and closer together. Notice that this definition makes no mention of limits.

**Note:** Using almost the exact same proof as above, one can show that \((x^{(n)})\) is Cauchy (in \(\mathbb{R}^k\)) if and only if each component \((x^n_j)\) is Cauchy (in \(\mathbb{R}\)).

**Theorem:**

\(\mathbb{R}^k\) is **complete**, meaning every Cauchy sequence in \(\mathbb{R}^k\) converges.

**Proof:** Easy! We’ve done the hard part above.

Let \((x^n)\) be a Cauchy sequence in \(\mathbb{R}^k\).
Then, by the above, each component \((x_j^{(n)})\) is Cauchy in \(\mathbb{R}\) (for \(j = 1, \ldots, k\))

But since \(\mathbb{R}\) is complete, each \(x_j^{(n)}\) converges to some \(x_j\).

But then by the Theorem above, \((x^{(n)})\) converges to \(x\), where \(x = (x_1, \ldots, x_k)\). □

5. **Bolzano-Weierstrass for \(\mathbb{R}^k\)**

**Video:** [Bolzano-Weierstraß in \(\mathbb{R}^k\)]

Lastly, we can generalize the Bolzano-Weierstraß Theorem to \(\mathbb{R}^k\), which says that every bounded sequence has a convergent subsequence.

Bounded just means that every component is bounded:

**Definition:**

If \((x^{(n)})\) is a sequence in \(\mathbb{R}^k\), then \((x^{(n)})\) is bounded if there is \(M > 0\) such that \(|x_j^{(n)}| < M\) for all \(n\) (and all \(j = 1, \ldots, k\))

Basically, all it means is that the length of each leg of the triangle is \(\leq M\).
It also implies that the sequence \((x^{(n)})\) is inside a box (of side \(M - (-M) = 2M\)), but we won’t really need that fact.

**Bolzano-Weierstraß for \(\mathbb{R}^k\):**

Every bounded sequence in \(\mathbb{R}^k\) has a convergent subsequence

(Strictly speaking, the figure above should be in \(\mathbb{R}^k\))
Proof: Basically apply Bolzano-Weierstraß to each component.

**STEP 1:** Let \((x^{(n)})\) be a bounded sequence in \(\mathbb{R}^k\)

Then there is \(M > 0\) such that for all \(j = 1, \ldots, k\), \(|x_j^{(n)}| \leq M\).

But with \(j = 1\) we get \(|x_1^{(n)}| \leq M\), so \((x_1^{(n)})\) is bounded, and hence
by Bolzano-Weierstraß for \(\mathbb{R}\) we get that \((x_1^{(n)})\) has a convergent subsequence \((x_1^{n_k})\) that converges to some \(x_1 \in \mathbb{R}\).

**STEP 2:**

**Note:** We can’t just apply Bolzano-Weierstraß to the whole sequence \((x_2^{(n)})\) because that might a priori give us a subsequence \((x_2^{(n_k)})\) for a different \(n_k\), which we don’t want (it’s kind of like getting an express train for a different track \(n_k\))

To get around this, consider the subsequence \((x_2^{(n_k)})\) of \(x_2^{(n)}\), where \(n_k\) is as in **STEP 1**. Then since \(|x_2^{(n)}| \leq M\) (by boundedness with \(j = 2\)),
in particular the same is true for \(x_2^{(n_k)}\) and therefore \((x_2^{(n_k)})\) is bounded
in \(\mathbb{R}\) and therefore has a subsequence \((x_2^{(n_{k_l})})\) that converges to some \(x_2 \in \mathbb{R}\) as \(l \to \infty\) (think of an express-express train).
But notice then that $(x_{kl}^{(n)})$ is a subsequence of $(x_{k}^{(n)})$ which therefore also converges to $x_1$ (the express train converges, and hence the express-express one converges as well) and hence $(x_{kl}^{(n_1)}, x_{kl}^{(n_2)}) \to (x_1, x_2)$

**STEP 3:** Continuing this way at most $k$ times (you can do an inductive construction if you want), we therefore obtain a subsequence of $(x^{(n)})$ with the property that each component converges to some $x_j \in \mathbb{R}$ (for $j = 1, \ldots, n$). So if you let $x = (x_1, \ldots, x_k)$, then that subsequence of $(x^{(n)})$ converges to $x$