Fall 2016/MA 314 Introduction to Metric Spaces

Geometry, both classical and modern, is about finding ways to mathematically investigate space. One of the primary tools for doing this is the theory of metric spaces. These notes are intended to help you learn the parts of metric space theory which will be useful for us in the rest of the semester. As always in mathematics, the best way to learn something is to try to solve a variety of problems, so there are a list of results for you to prove. You are encouraged to work with your classmates on these, but your write-ups must be your own. If you find yourself in the position of having to just listen to someone else explain something - or explaining something to someone else - you should find a different partner to work with. It is crucial you do your own thinking. After I am back from Germany, we'll talk about these problems and about how metric spaces fit into the goals of the semester. Finally, if you still feel uncomfortable with metric spaces by next week, that's fine! Learning mathematics is a process of constantly developing our understanding.

1. Metric Spaces

At its heart, a metric space is a set (whose elements are called **points**) together with a way of measuring distance (called a **metric**) between the elements.

Definition. A set X and a function $d\colon X \times X \to [0,\infty)$ is called a **metric space** if the following hold for all $x,y,z \in X$:

(M1) $d(x,x) = 0$ (the distance from a point to itself is zero)

(M2) If $d(x,y) = 0$ then $x = y$ (the only point y of distance 0 from x is x itself)

(M3) $d(x,y) = d(y,x)$ (the distance from x to y is the same as the distance from y to x)

(M4) $d(x,z) \leq d(x,y) + d(y,z)$ (the triangle inequality: taking a detour from x to z by going through y can only increase distance)

If X and d satisfy (M1), (M3), and (M4) (but not necessarily (M2)) then X and d are a **pseudo-metric space** and d is a **pseudo-metric** or **semi-metric**. We often call the pair (X,d) a metric space, to emphasize the dependence on both the set and the distance function.

Here is the simplest example of a metric. Prove that it is one.

Exercise 1.1. Let X be any set and define $d\colon X \times X \to [0,\infty)$ by

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

The function d is called the **discrete metric** on X.

The usual distance function (called the **euclidean metric**) on $\mathbb{R}^n$ is the most important of all metrics. It is easiest to define and work with if we use some linear algebra. If you don't remember much about the dot product, now would be a good time to go review it. Other than as a good review of the properties of the dot product, the proof isn't that important for us (though the theorem is!) There are shorter proofs - perhaps you can find one! Recall that the **norm** or **magnitude** of a vector $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$ is

$$||x|| = \sqrt{x \cdot x} = x_1^2 + \cdots + x_n^2.$$

If $n = 1$, we usually write $|x|$ instead of $||x||$.

Theorem 1.2. *The euclidean metric d on $\mathbb{R}^n$, defined by*

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

for all $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$, is a metric.

Perhaps you can come up with your own way of proving this, but here is one.

Proof. It is easily seen from the formula for dot product that (M1) and (M3) hold. To show (M2) holds, suppose that $d(x,y) = 0$. Then

$$\begin{aligned} d(x,y) &= 0 \quad \Rightarrow \\ \sqrt{(x-y)\cdot(x-y)} &= 0 \quad \Rightarrow \\ (x-y)\cdot(x-y) &= 0 \quad \Rightarrow \\ (x_1-y_1)^2+\cdots+(x_n-y_n)^2 &= 0 \end{aligned}$$

Each term in the sum on the left is non-negative, so the sum can only be zero if each term is zero. Consequently, $x = y$.

We now show (M4), the triangle inequality. Let $x,y,z \in \mathbb{R}^n$. Define $a = (x-y)$, $b = (y-z)$, and $c = (x-z) = a+b$. We need to show that

$$||a||+||b|| \geq ||a+b||$$

This inequality holds if and only if the inequality obtained by squaring both sides holds:

$$a\cdot a+b\cdot b+2\sqrt{a\cdot a}\sqrt{b\cdot b} \geq (a+b)\cdot(a+b).$$

The right side is equal to $a\cdot a+b\cdot b+2a\cdot b$. Thus, we desire to show

$$\begin{aligned} \sqrt{a\cdot a}\sqrt{b\cdot b}-a\cdot b &\geq 0 \quad \Leftrightarrow \\ ||a||||b||-a\cdot b &\geq 0. \end{aligned}$$

The geometric interpretation of the dot product is that $a\cdot b = \sqrt{a\cdot a}\sqrt{b\cdot b}\cos\theta$ where θ is the angle between a and b. Since $\cos\theta \in [-1,1]$ and since we wanted to show

$$\sqrt{a\cdot a}\sqrt{b\cdot b}(1-\cos\theta) \geq 0$$

we have our desired result. □

Henceforth, if we do not specify a metric on $\mathbb{R}^n$, we will assume we are using the euclidean metric.

What are some other metric spaces? One way of getting a new metric space is simply to restrict to a subset of a given metric space. This is surprisingly important.

Prove the following theorem (it should be easy!)

Theorem 1.3. *Suppose that X is a metric space with metric d and that $A \subset X$. Let d_A denote the restriction of d to $A\times A \subset X\times X$. Then A is a metric space with metric d_A (it is called a **subspace** of X and d_A is the **subspace metric**.*

Exercise 1.4. Let $S^1 = \{x \in \mathbb{R}^2 : ||x|| = 1\}$ be the unit circle. Give S^1 the subspace metric from $\mathbb{R}^2$. What is the distance from $(1,0)$ to $(0,1)$? Is this equal to the distance you would travel as you go around the circle from $(1,0)$ to $(0,1)$?

The set $\mathbb{R}^2$ can have many different metrics. We have seen the discrete metric and the euclidean metric. Here is another metric, called the **Manhattan metric** - prove it is a metric.

Definition. For $(x,y),(a,b) \in \mathbb{R}^2$, define $d((x,y),(a,b)) = |x-a|+|y-b|$. Then d is called the **Manhattan metric** on $\mathbb{R}^2$.

Exercise 1.5. Prove the Manhattan metric is a metric. Why do you think it is called the Manhattan metric? (Hint: look at a map of Manhattan and think about how far you have to travel to get between two different points in a taxicab.)

Although we could continue to give lots of examples of metrics on lots of different sets, our time is better spent by continuing on.

2. Isometries

Definition. Suppose that (X,d_X) and (Y,d_Y) are metric spaces. A function $f\colon X\to Y$ is an **isometry** if it is a bijection and if, for all $a,b\in X$ we have

$$d_Y(f(a),f(b)) = d_X(a,b).$$

Theorem 2.1. *Let $X=\mathbb{R}^n$ (for $n=2$ or $n=3$). Then for each of the following functions $T\colon X\to X$, T is an isometry:*

1. *translations: $T(x)=x+a$ for some fixed $a\in\mathbb{R}^n$ and all $x\in X$.*
2. *orthogonal transformations: $T(x)=Ax$ where A is an orthogonal matrix, i.e. an $n\times n$ matrix such that $A^TA=AA^T=I$ (where I is the identity matrix and the superscript T denotes the transpose).*

(Hint: Recall that for vectors $v,w\in\mathbb{R}^n$ we can convert the dot product into matrix multiplication by $v\cdot w = v^Tw$.)

When $n=2$, there are two special types of orthogonal transformations: coordinate plane reflections and rotations. For a coordinate plane reflection, the matrix $A=\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}$ or $A=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$. For a rotation, $A=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$.

When $n=3$, we have similar orthogonal transformations. The corresponding matrices are

$$\begin{pmatrix}-1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix},\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -1\end{pmatrix}$$

for the coordinate plane transformations and

$$\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix},\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{pmatrix},\begin{pmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\theta\end{pmatrix}$$

for the coordinate plane rotations.

Exercise 2.2. Prove that if $f\colon X\to Y$ is an isometry, then $f^{-1}\colon Y\to X$ is also an isometry.

Recall[1] that a group G is a set with an operation $\circ$ such that the following axioms hold:

(G1) for all $a,b\in G$, there is a unique element $a\circ b$ in G

(G2) there exists $\mathrm{id}\in G$ such that for all $a\in G$, $a\circ\mathrm{id}=\mathrm{id}\circ a=a$

(G3) for all $a\in G$, there exists $a^{-1}\in G$ such that $a\circ a^{-1}=\mathrm{id}$.

(G4) for all $a,b,c\in G$, $(a\circ b)\circ c = a\circ(b\circ c)$.

[1] In math classes, this means, "I hope somewhere you've seen this before, but if not, keep reading and try to figure it out."

Theorem 2.3. *Let X be a metric space with metric d. Then the set* $\textsc{Isom}(X)$ *of isometries from X to itself forms a group with function composition as the operation.*

(Hint: You may take the associativity of function composition for granted.)

The next two theorems classify the isometries of $\mathbb{R}$ and $\mathbb{R}^2$. We will use them, and the method of proof, repeatedly. Notice how we begin with special cases and work our way toward the general statement. The complete proof of the first theorem is given. You should fill in the details for the proof of the second theorem.

Theorem 2.4. *Let $T:\ \mathbb{R} \to \mathbb{R}$ be an isometry. Then T is either a translation or the composition of a reflection and a translation.*

Proof. **Case 1:** $T(0) = 0$ and $T(1) = 1$.

We claim that T is the identity (which is a translation by 0). Let $x \in \mathbb{R}$. Since

$$|x| = d(x,0) = d(T(x),T(0)) = d(T(x),0) = |T(x)|,$$

we have $x = \pm T(x)$. If $x = -T(x)$ and $x \neq 0$ for some x, then

$$|x-1| = d(x,1) = d(T(x),T(1)) = d(T(x),1) = |T(x)-1| = |-x-1| = |1+x|$$

Thus,

$$x^2 - 2x + 1 = x^2 + 2x + 1$$

and so $x = 0$, contrary to our assumption. Thus, $T(x) = x$ for all $x \in \mathbb{R}$.

Case 2: $T(0) = 0$.

We claim that either T is the identity or T is the reflection R defined by $R(x) = -x$ for all $x \in \mathbb{R}$.

Observe that

$$1 = d(1,0) = d(T(1),T(0)) = d(T(1),0) = |T(1)|$$

Thus, $T(1) = \pm 1$. If $T(1) = 1$, then by the previous case, we are done. If $T(1) = -1$, then $R \circ T(0) = 0$ and $R \circ T(1) = 1$. Thus, by Case 1, $R \circ T = \text{id}$. The reflection R is its own inverse and so applying it both sides of the equation we get $T = R$.

General Case: Assume only that $T:\ \mathbb{R} \to \mathbb{R}$ is an isometry.

Let $a = T(0)$ and let $\tau:\ \mathbb{R} \to \mathbb{R}$ be the translation $\tau(x) = x - a$. Observe that $\tau^{-1}(y) = y + a$ for all $y \in R$ and, therefore, τ^{-1} is also a translation. Also, $\tau \circ T(0) = 0$. Thus, by Case 2, $\tau \circ T$ is either the identity, in which case $T = \tau^{-1}$, or is a reflection in which case T is the composition of the translation τ^{-1} with the reflection. $\square$

Theorem 2.5. *Let $T:\ \mathbb{R}^2 \to \mathbb{R}^2$ be an isometry. Then T is the composition of translations, reflections, and rotations.*

Proof. Let $e_1 = (1,0)$ and $e_2 = (0,1)$ be the standard basis vectors for $\mathbb{R}^2$. Let 0 denote the vector $(0,0)$.

Case 1: $T(0) = 0$, $T(e_1) = e_1$ and $T(e_2) = e_2$.

$\langle$ *Prove that T is the identity* $\rangle$

Case 2: $T(0) = 0$ and $T(e_1) = e_1$, but $T(e_2)$ is not necessarily e_2.

$\langle$ *Prove that T is either the reflection $R(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} (x,y)$ or the identity.* $\rangle$

Case 3: $T(0) = 0$ but T does not necessarily fix either e_1 or e_2.

⟨ *Prove that T is either a rotation or the composition of a rotation and the reflection R* ⟩

General Case: T is any isometry of $\mathbb{R}^2$.

⟨ *Prove that T is the composition of rotations, reflections, and translations (not necessarily using all of them).* ⟩

$\square$

3. Paths

In a previous example, we saw how using the subspace metric on the circle S^1 the distance from $(1,0)$ to $(-1,0)$ is 2, but if we were forced to stay on the circle we'd have to travel a distance of π. In this section, we make this precise by talking about path metrics. Later on we will generalize some of this discussion to other metric spaces. We will make use of basic multi-variable calculus, we take a moment to review some of it. For more, see Colley's Vector Calculus text. Mostly we'll be working in 2- and 3-dimensions but some of this works in all dimensions, so whenever there's no extra cost we'll phrase things for $\mathbb{R}^n$ for any $n \geq 1$.

Recall that if $[a,b] \subset \mathbb{R}$ is an interval and if $\gamma\colon [a,b] \to \mathbb{R}^n$ is a function, then for every $t \in [a,b]$, $\gamma(t) = (x_1(t), x_2(t), \ldots, x_n(t))$ is an element of $\mathbb{R}^n$. The function γ is continuous if each of the coordinate functions $x_i\colon [a,b] \to \mathbb{R}^n$ is continuous. We say that γ is of class C^1 if each coordinate function x_i is differentiable and has continuous derivative $\dot{x}_i$. If γ is C^1, then the derivative of γ is $\dot{\gamma}(t) = (\dot{x}_1(t), \dot{x}_2(t), \ldots, \dot{x}_n(t))$ for all t.

Suppose that $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable function. We may write $F(x) = (f_1(x), f_2(x), \cdots, f_m(x)) \in \mathbb{R}^m$. Its **derivative** DF at the point $a \in \mathbb{R}^n$ is the *linear* function given by

$$x \mapsto DF|_a x.$$

Where $DF|_a$ is the $m \times n$ matrix

$$DF|_a = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_2}(a) & \frac{\partial f_1}{\partial x_2}(a)\ldots & \frac{\partial f_2}{\partial x_n}(a) & \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_2}(a) & \frac{\partial f_m}{\partial x_2}(a)\ldots & \frac{\partial f_m}{\partial x_n}(a) & \end{pmatrix}.$$

The chain rule says that if $F\colon \mathbb{R}^n \to \mathbb{R}^m$ and $G\colon \mathbb{R}^m \to \mathbb{R}^p$ are differentiable functions, then the function $G \circ F$ is differentiable and has matrix

$$D(G \circ F)|_a = DG|_{F(a)} DF|_a$$

We now apply these facts to paths.

Definition. A **smooth path** in $U \subset \mathbb{R}^n$ is a C^1 function $\gamma\colon [a,b] \to U$ where $a < b$ are real numbers. We will usually (but not always) take $a = 0$ and $b = 1$. The **derivative** of γ is the function $\dot{\gamma}$

A **piecewise smooth path** in $U \subset \mathbb{R}^n$ is a continuous function $\gamma\colon [a,b] \to U$ such that there exist $t_0, \ldots, t_n \in [a,b]$ such that

$$a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$$

so that the restriction of γ to each interval $[t_i, t_{i+1}]$ is smooth. We say that γ is a path from $\gamma(a)$ to $\gamma(b)$.

Exercise 3.1. Draw a picture of a smooth path in $\mathbb{R}^2$ and a path in $\mathbb{R}^2$ which is piecewise smooth, but not smooth.

Exercise 3.2. Write down a formula for an example of a smooth path in $\mathbb{R}^2$ and write down a formula (perhaps a piecewise formula) for an example of a piecewise smooth, but not smooth, path in $\mathbb{R}^2$. Find the derivatives of the smooth and piecewise smooth paths. If you can, find their lengths; otherwise use Mathematica or WolframAlpha to find an approximation to their lengths.

Definition. Suppose that $\gamma\colon [a,b] \to \mathbb{R}^2$ is a piecewise smooth path. Then the (euclidean) **length** of γ is

$$L(\gamma) = \int_a^b ||\dot{\gamma}(t)||\,dt = \int_a^b \sqrt{\dot{x}^2(t) + \dot{y}^2(t)}\,dt.$$

We will often let $ds = ||\dot{\gamma}||\,dt$ and so $L(\gamma) = \int_a^b ds$.

Since $||\dot{\gamma}||$ is the speed of γ, we obtain length by integrating speed.

Exercise 3.3. The line segment in $\mathbb{R}^2$ between points (x_0, y_0) and (x_1, y_1) can be parameterized as

$$\gamma(t) = (1-t)\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + t\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

for $t \in [0,1]$. Calculate $L(\gamma)$ and show it is equal to the euclidean distance from (x_0, y_0) to (x_1, y_1).

In fact, as you probably already know, a straight line segment is the shortest path between two points. You will be asked to prove in a bit. First we show that isometries of $\mathbb{R}^2$ preserve the length of curves. We will be constructing variations of these arguments for different geometries down the road.

Theorem 3.4. *Suppose that $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry and that $\gamma\colon [a,b] \to \mathbb{R}^2$ is a piece-wise smooth path. Then*

$$L(\gamma) = L(T \circ \gamma).$$

(Hint: Use the classification of isometries of $\mathbb{R}^2$ and the matrix form of the chain rule from multi-variable calculus. If you haven't seen this try looking it up in Colley's vector calculus textbook.)

Theorem 3.5. *Suppose that $a = (x_0, y_0)$ and $b = (x_1, y_1)$ are points in $\mathbb{R}^2$ and that γ is a smooth path from a to b. Prove that $L(\gamma) \geq d(a,b)$ with equality only if the range of γ is a line segment.*

Proof. **Case 1:** $x_0 = x_1$ and γ is smooth.

Let $\gamma(t) = (x(t), y(t))$ for all $t \in [a,b]$. We have

$$L(\gamma) = \int_a^b \sqrt{\dot{x}^2(t) + \dot{y}^2(t)}\,dt.$$

$\langle$ *Explain why*

$$L(\gamma) \overset{(*)}{\geq} \int_a^b |\dot{y}|\,dt \geq y(b) - y(a) = y_1 - y_0 = d(a,b).$$

$\rangle$

If the range of γ is not a line segment, then $x\colon [a,b] \to \mathbb{R}^2$ is non-constant and $\dot{x}$ is not the zero function.

$\langle$ *Show in this case that the inequality (*) is strict* $\rangle$

The general case: γ is smooth.

$\langle$ *Prove the theorem in this case by using an isometry to convert it to Case 1.* $\rangle$

$\square$

Challenge! Prove the previous theorem for the case when γ is only piecewise-smooth.

The next task will be to show how, whenever, we can measure the length of a path we can create a corresponding pseudo-metric. Sometimes the pseudo-metric will actually be a metric! But that will wait until I am back from Germany.

We conclude this initial set of notes with some comments on where we will be going over the course of the semester.

We will be studying (mainly) three different (2–dimensional) geometries: Euclidean geometry (which you already know about), Spherical geometry, and Hyperbolic geometry. Each of these geometries incorporates the following:

- A metric space. In the case of Euclidean geometry, this is the euclidean metric from above.
- A way of measuring the lengths of paths which has the property that for any two points there is a path between those points whose length is the distance between the points.
- An isometry group. The isometry group preserves both the distance between points and the length of (piece-wise smooth) paths.
- Geodesics. These are paths which are "locally length-minimizing". In Euclidean geometry, they are line segments.

Our main goal will be to understand how every 2-dimensional space (subject to minor restrictions) admits one of these geometries. This will involve equivalence relations, group actions, and all sorts of other fun stuff.