CATEGORY THEORY
TOPIC 33: METRIC SPACES
(DRAFT)

PAUL L. BAILEY

Abstract. We define *metric spaces*, which are sets together with the additional structure provided by a distance function. We give examples and develop their basic properties. We go far enough in this direction to thoroughly motivate the next level of abstraction, the *topological space*.

1. Motivation

At the dawn of man, pastoral shepherds made sure that their sheep never strayed too far, and they counted them at the end of the day. Thus the two primary branches of mathematics originated.

- Counting led to arithmetic, which evolved into algebra, which is the study of binary operators.
- Measuring distance led to geometry, which generalizes to topology, which is the study of open sets.

The use of open intervals in Calculus ensures that the notion of limit can be applied accurately. Recall that an open interval is an interval that does not contain its endpoints. To say that a point a is in an open interval, and that a function f is defined on that interval, indicates that f is also defined for all values *near* a.

We recall the definition of a continuous function from Calculus.

Let $f : D \to \mathbb{R}$, where $D \subset \mathbb{R}$. Let $a \in D$ such that there exists an open interval $I \subset D$ with $a \in I$. We say that f is *continuous* at a if

$$\forall \epsilon > 0 \, \exists \delta > 0 \ni x \in I \text{ and } |x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon.$$

In words, f is continuous at a if whenever x is near a, $f(x)$ is near $f(a)$. By "near", we mean "within a small enough open interval". We wish to generalize the concept of continuous, and to do so, we generalize the concept of "near".

Date: March 17, 2018.

2. Metric Spaces

Definition 1. Let X be a set. A *metric* on X is a function

$$d : X \times X \to \mathbb{R}$$

satisfying

(M1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$ (Positivity);
(M2) $d(x, y) = d(y, x)$ (Symmetry);
(M3) $d(x, y) + dy, z) \geq d(x, z)$ (Triangle Inequality).

The pair (X, d) is called a *metric space*.

Example 1. The set of real numbers is a metric space. The distance from x to y is defined by $d(x, y) = |x - y|$.

Example 2. Let $X = \mathbb{R}^2$ and use the Pythagorean theorem to define the metric d by

$$d(p, q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

where $p = (x_1, y_1)$ and $q = (x_2, y_2)$.

Example 3. Let $X = \mathbb{R}^3$. Two applications of the Pythagorean theorem and some slight simplification leads to the definition of the metric d by

$$d(p, q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2},$$

where $p = (x_1, y_1, z_1)$ and $q = (x_2, y_2, z_2)$.

Example 4. Let $X = \mathbb{R}^n$. We need to slightly modify our notation to conveniently write the distance formula. Thus for $p = (x_1, x_2, \ldots, x_n)$ and $q = (y_1, y_2, \ldots, y_n)$, define

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}.$$

Example 5. Let $\mathbb{R}^\infty$ denote the set of all sequences of real numbers that are eventually zero, that is, sequences (x_n) such that $x_n = 0$ for all but finitely many n. Let $X = \mathbb{R}^\infty$ and for $x, y \in X$, define

$$d(x, y) = \sqrt{\sum_{i=1}^{\infty} (x_i - y_i)^2},$$

where $x = (x_n)$ and $y = (y_n)$. This make sense, since there are only finitely many nonzero summands. Then (X, d) is a metric space.

Example 6. Let ℓ^2 denote the set of all sequences of real numbers (x_n) that satisfy the convergence criterion

$$\sum_{i=1}^{\infty} x_i^2 < \infty.$$

Let $X = \ell^2$ and for $x, y \in X$, define

$$d(x, y) = \sqrt{\sum_{i=1}^{\infty} (x_i - y_i)^2},$$

where $x = (x_n)$ and $y = (y_n)$. That this series converges follows from the inequality

$$(a \pm b)^2 \leq 2(a^2 + b^2),$$

which the reader is welcome to verify. Then (X, d) is a metric space.

Example 7. Let X be any set and define $d : X \times X \to \mathbb{R}$ by

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

Then d is a metric on X, called the *discrete metric*, and (X, d) is called a *discrete metric space*.

Example 8. Let $\mathcal{F}_{[a,b]}$ denote the set of all bounded functions $f : [a, b] \to \mathbb{R}$. Let $X = \mathcal{F}_{[a,b]}$ and for $f, g \in X$ define

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

Then (X, d) is a metric space.

Example 9. Let $\mathcal{C}_{[a,b]}$ denote the set of all continuous functions $f : [a, b] \to \mathbb{R}$. Let $X = \mathcal{C}_{[a,b]}$ and for $f, g \in X$ define

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

Then (X, d) is a metric space.

3. Subspaces of Metric Spaces

When creating a new category of mathematical objects, it is typical to discuss what sort of subobjects are to be consider. These is relatively easy in the case of a metric space.

Definition 2. Let (X, d) be a metric space and let $A \subset X$. Let $d_A : A \times A \to \mathbb{R}$ be the restriction of d to $A \times A \subset X \times X$. Then d_A is a metric on A, and (A, d_A) is called a *subspace* of (X, d).

Example 10. Define

$$\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

We call $\mathbb{S}^1$ the *unit circle*. It inherits the metric $d_{\mathbb{S}^1}$ from $(\mathbb{R}^2, d)$.

Define

$$\mathbb{D}^2 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}.$$

We call $\mathbb{D}^2$ the *(closed) unit disk*, and $(\mathbb{D}^2, d_{\mathbb{D}^2})$ is a metric space.

Example 11. Let $\mathbb{S}^1$ be the unit circle, and let d be as in Example 10. We may define a metric

$$\alpha : \mathbb{S}^1 \times \mathbb{S}^1 \to \mathbb{R} \quad \text{by} \quad \alpha(p_1, p_2) = 2 \arcsin(d(p_1, p_2)).$$

where $p_1, p_2 \in \mathbb{S}^1$. Then $\alpha(p_1, p_2)$ is the angle, measured in radians, from p_1 to the origin and then to p_2; this is the arclength of the shortest path between these two points.

This produces a different metric on $\mathbb{S}^1$. In due course, we will investigate the relationship between these metrics and related consequences for the structure of the metric space.

4. Classification of Points

Definition 3. Let (X, d) be a metric space. Let $A \subset X$ and let $p \in X$.

We say that p is an *interior point* of A if there exists $\epsilon > 0$ such that

$$\forall x \in x : d(p, x) < \epsilon \Rightarrow x \in A.$$

We say that p is an *exterior point* of A if there exists $\epsilon > 0$ such that

$$\forall x \in x : d(p, x) < \epsilon \Rightarrow x \notin A.$$

We say that p is a *boundary point* of A if for every $\epsilon > 0$ there exists $a \in A$ and $x \in X \smallsetminus A$ such that $d(p, a) < \epsilon$ and $d(p, x) < \epsilon$.

It is clear that interior points are in A, exterior points are in the complement of A, and boundary points may or may not be in A.

Example 12. Let $X = \mathbb{R}$ and $A = [0, 2)$. Then 1 is an interior point of A, 3 is an exterior point of A, and 0 and 2 are both boundary points of A.

Example 13. Let $X = \mathbb{R}^2$ and $A = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x < 2 \text{ and } 0 \leq y < 2\}$. Then $(1, 1)$ is an interior point of A, $(0, 3)$ and $(3, 0)$ are exterior points of A, $(0, 0)$ is a boundary point of A with is an element of A, and $(1, 2)$ is a boundary point of A which is not an element of A.

What is written above is the type of definition used by mathematical analysts. We wish to discuss this in terms of open sets, so that we can use the language and thought processes of mathematical topologists. This first step is to rephrase our definitions in terms of open balls.

Definition 4. Let (X, d) be a metric space. Let $p \in X$ and let $r > 0$.

The *open ball* about p or radius r is

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

The *closed ball* about p or radius r is

$$B_r(p) = \{x \in X \mid d(p, x) \leq r\}.$$

A *neighborhood* of p is a set which contains an open ball about p.

A *deleted neighborhood* of p is a set of the form $A \setminus \{p\}$, where A is a neighborhood of p.

It is clear that a neighborhood of p contains p, but a deleted neighborhood of p does not.

Setwise complements are useful in the study of metric spaces, so we dedicate a special notation for them. Thus, if A is a subset of a metric space X, let A^c denote the complement of A with respect to the entire space X; that is, $A^c = X \setminus A$.

Let A and B be any sets. We say that A *intersects* B if $A \cap B$ is nonempty. Clearly, A intersects B if and only if B intersects A. Using this, the definitions above are equivalent to the following.

Definition 5. Let (X, d) be a metric space. Let $A \subset X$ and let $p \in X$.

We say that p is a *closure point* of A if every neighborhood of p intersects A.

We say that p is an *interior point* of A if there exists neighborhood of p which is contained in A.

We say that p is a *boundary point* of A if every neighborhood of p intersects A and A^c.

We say that p is an *accumulation point* of A if every deleted neighborhood of p intersects A.

We say that p is an *isolated point* of A if $p \in A$ and there exists a deleted neighborhood of p which is contained in A^c.

The *closure* of A is the set of closure points of A, and is denoted $\operatorname{Clo} A$.

The *interior* of A is the set of interior points of A, and is denoted $\operatorname{Int} A$.

The *boundary* of A is the set of boundary points of A, and is denoted $\operatorname{Bnd} A$.

The *derived set* of A is the set of accumulation points of A, and is denoted A'.

5. Open and Closed Sets

Definition 6. Let (X, d) be a metric space. Let $A \subset X$.

We say that A is *open* if for every $a \in A$ there exists $\epsilon > 0$ such that

$$\forall x \in X : d(a, x) < \epsilon \Rightarrow x \in A.$$

The next may seem obvious, but it is required to go forward.

Proposition 1. *Open balls are open sets.*

Proof. Let (X, d) be a metric space, $a \in X$, and $r > 0$. Now $B_r(a)$ is an arbitrary open ball in X. Let $b \in B_r(a)$ and let $s = d(a, b)$. Let $t = r - s$ and let $c \in B_t(b)$. Then $d(a, c) \leq d(a, b) + d(b, c) < s + t = r$. Thus $c \in B$, so $B_t(b) \subset B_r(a)$, so $B_r(a)$ is open. □

Clearly, a set is open if every point in it has a neighborhood which is contained in it. Equivalently, a set is open if it is a union of open balls.

Proposition 2. *Let X be a metric space. Then*

(a) *The empty set $\varnothing$ and the whole space X are open sets.*
(b) *The union of an arbitrary number of open sets is an open set.*
(c) *The intersection of finitely many open sets is an open set.*

Proof. A set is open if every point in it has an open ball around it which is entire in the empty set. This is vacuously true for the empty set, since it contains no points (so every point in it has this property). The whole space is open, since every ball in X is in X.

A set is open if it is the union of a collection of open balls. The union of a collection of open sets is the union of a collection of unions of open balls. So, it is itself a union of open balls.

The third property requires a little more care. Thus let n be a (finite) positive integer, and $U_1, U_2, \dots, U_n$ be open sets. Let $A = \cap_{i=1}^n U_i$ be their intersection. If A is empty, it is open by part **(a)**. Otherwise, it contains a point, say $a \in A$. Then $a \in U_i$ for each $i \in \{1, \dots, n\}$. Since U_i is open, there exists ϵ_i such that $a \in B_{\epsilon_i}(a) \subset U_i$, for each i. Let $\epsilon = \min\{\epsilon_1, \dots, \epsilon_n\}$. Then $B_\epsilon(a) \subset U_i$ for each i, so $B_\epsilon(a) \subset A$. Thus A is open. □

Note that the intersection of arbitrarily many open sets is not necessarily open. For example, in $\mathbb{R}$, we have

$$\bigcap_{n=1}^{\infty} \Big(-\frac{1}{n}, \frac{1}{n} \Big) = \{0\}.$$

6. Continuous Functions

Recall that a real-valued function f defined on an interval containing a is continuous at a if

$$\forall \epsilon > 0 \, \exists \delta > 0 \ni x \in I \text{ and } |x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon.$$

We extend the definition of continuous functions as follows.

We may rephrase the continuity condition using balls as

$$\forall \epsilon > 0 \, \exists \delta > 0 \ni x \in B_\delta(a) \Rightarrow f(x) \in B_\epsilon(f(a)).$$

This clearly has the identical meaning in the case of $\mathbb{R}$, but is immediately applicable to define continuity at a point in any metric space.

It will be convenient, in some cases, to cut loose of the metric altogether. Thus, we redefine a neighborhood of a point a to be any set containing an open set which contains a. All we have previously said remains as it was with this redefinition.

Indeed, we can reword the definition above using the language of neighborhoods, by stating that f is continuous at a if for every neighborhood V of $f(a)$ there exists a neighborhood U of a such that $f(U) \subset V$. This is clearly equivalent.

The traditional approach starts with this definition.

Definition 7. Let (X, d_X) and (Y, d_Y) be metric spaces. Let $f : X \to Y$ and let $a \in X$. We say that f is *continuous at* a if

$$\forall \epsilon > 0 \, \exists \delta > 0 \ni d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon.$$

We say that f is *continuous* if f is continuous at a for every $a \in X$.

With a bit of introspection, one sees that $f : X \to Y$ is continuous at a if for every neighborhood V of $f(a)$ there exists a neighborhood U of a such that $f(U) \subset V$. We would like to discuss the proof of the following.

Proposition 3. *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$. Then f is continuous if and only if the preimage of every open set in Y is open in X.*

Proof. ($\Rightarrow$) Suppose that f is continuous, and let $V \subset Y$ be open. Let $A = f^{-1}(V)$, and let $a \in A$. Let $b = f(a)$. Since V is open, there exists $\epsilon > 0$ such that $d_Y(y, b) < \epsilon$ implies that $y \in V$. Since f is continuous at a, there exists $\delta > 0$ such that $d_X(x, a) < \delta$ implies that $d_Y(f(x), b) < \epsilon$; by the previous sentence, this in turn implies that $f(x) \in V$, so that $x \in f^{-1}(V) = A$. Thus if $d_X(x, a) < \delta$, we know that $x \in A$, so A is open in X.

($\Leftarrow$) Suppose that the preimage of every open set in Y is open in X. Let $a \in X$; we wish to show that f is continuous at a. Let $\epsilon > 0$. Let $b = f(a)$ and set $V = B_\epsilon(a)$. Let $U = f^{-1}(V)$. Since V is open in Y, U is open in X. So there exists $\delta > 0$ such that $x \in U$ whenever $d_X(a, x) < \delta$. Then $f(x) \in f(U) = V = B_\epsilon(a)$; thus $d_X(x, a) < \delta$ implies that $d_Y(f(x), f(a)) < \epsilon$. This shows that f is continuous at a. □

Department of Mathematics and CSci, BASIS Mesa
E-mail address: paul.bailey@basised.com