

Metric Spaces

BY JOHN M. LEE

This handout should be read between Chapters 1 and 2 of the text. It incorporates material from notes originally prepared by Steve Mitchell and Isabella Novik.

1. CONTINUOUS FUNCTIONS BETWEEN EUCLIDEAN SPACES

We start by recalling what it means for a function $f: \mathbb{R} \rightarrow \mathbb{R}$ to be continuous. We say that such an f is **continuous at a point** $a \in \mathbb{R}$ if

$$\lim_{x \rightarrow a} f(x) = f(a),$$

that is, if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all x with $|x - a| < \delta$, we have $|f(x) - f(a)| < \varepsilon$. We say simply that **f is continuous** if it is continuous at every point of \mathbb{R} .

Intuitively, continuity says that if you want to make the function value $f(x)$ close to $f(a)$, all you have to do is to make x close enough to a . “Closeness” is measured in terms of the **distance** from x to a , which is the absolute value of their difference: $|x - a|$.

More generally, we could consider functions $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$, where \mathbb{R}^n is the set of n -tuples (x_1, \dots, x_n) of real numbers, also known as **n -dimensional Euclidean space**. (Here n is an arbitrary positive integer.) In this case, to measure the distance between two points, we have to replace the absolute value by the **norm** of a vector, defined as follows. First, if $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ are any two points in \mathbb{R}^n , their **dot product** is the real number $\mathbf{x} \cdot \mathbf{y}$ defined by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i.$$

Then the **norm** of any $\mathbf{x} \in \mathbb{R}^n$ is defined by

$$(3.1) \quad \|\mathbf{x}\| = (\mathbf{x} \cdot \mathbf{x})^{1/2} = \left(\sum_{i=1}^n (x_i)^2 \right)^{1/2}.$$

Finally, the **distance** between two points $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ is the norm of their difference:

$$(3.2) \quad \|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2}.$$

When $n = 1$, this just gives the absolute value of the difference, because $(x^2)^{1/2} = |x|$ for any real number x . The general formula is motivated by the familiar 2- and 3-dimensional cases (think Pythagorean theorem!), but it works just as well in general.

Once we have this way of measuring distances, we can define continuity exactly as we did in the case $m = n = 1$: We say that $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ is **continuous** if for every $\mathbf{a} \in \mathbb{R}^m$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that for all \mathbf{x} with $\|\mathbf{x} - \mathbf{a}\| < \delta$, we have $\|f(\mathbf{x}) - f(\mathbf{a})\| < \varepsilon$.

Of course, not all interesting functions are defined on the *whole* space \mathbb{R}^m ; sometimes we have to restrict the domain, and perhaps also the codomain, to some subsets $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$. For example, the formula $f(x, y) = 1/(x + y)$ does not define a function from \mathbb{R}^2 to \mathbb{R} ; its domain has to be a subset such as $X = \{(x, y) \in \mathbb{R}^2 : x + y \neq 0\}$. In cases such as this, the above definition of continuity still makes sense, except that we have to restrict both \mathbf{a} and \mathbf{x} to lie in X .

2. DEFINITION OF A METRIC SPACE

The above discussion leads to the following natural question: If f is a function whose domain and codomain are subsets $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$, exactly what sort of mathematical objects are X and Y ? They are probably not vector spaces, because they might not be closed under vector addition and scalar multiplication. But they are not just naked sets either, because they come equipped with *some* of the structure of \mathbb{R}^n , namely its distance function. Note that it is *only* the distance between two points that gets used in the definition of continuity; we don't really need to know anything about the dot product, or norms, or vector addition or subtraction, or scalar multiplication, as long as we know how to compute the distance between any two points. This suggests that continuity might make sense in a much more general setting.

Thus we make the following definition. If X is a set, a *metric* on X is a function $d: X \times X \rightarrow \mathbb{R}$ satisfying

- (1) $d(x, y) \geq 0$ for all $x, y \in X$, with equality if and only if $x = y$.
- (2) $d(x, y) = d(y, x)$ for all $x, y \in X$.
- (3) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

The last condition is called the *triangle inequality*, because in the familiar special case $X = \mathbb{R}^2$, it says that the length of any side of a triangle is no more than the sum of the lengths of the other two sides.

Exercise 3.1. Show that the formula $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ defines a metric on \mathbb{R}^n , as follows. In these statements \mathbf{x} , \mathbf{y} , and \mathbf{z} represent arbitrary elements of \mathbb{R}^n .

- (a) Show that $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ and $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = (\mathbf{x} \cdot \mathbf{y}) + (\mathbf{x} \cdot \mathbf{z})$.
- (b) Show that $\mathbf{x} \cdot \mathbf{x} \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{0}$.
- (c) Show that $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\|$. [Hint: If $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$, let $a = 1/\|\mathbf{x}\|$ and $b = 1/\|\mathbf{y}\|$, and use the fact that $\|a\mathbf{x} \pm b\mathbf{y}\|^2 \geq 0$.]
- (d) Show that $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. [Hint: Expand $(\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y})$ and apply part (c).]
- (e) Verify that d is a metric.

This metric d is called the *Euclidean metric* on \mathbb{R}^n .

A *metric space* is a set X together with a specific choice of metric d on X . Sometimes we will say "Let (X, d) be a metric space," if we want to emphasize the specific metric we have in mind. But sometimes we will simply say "Let X be a metric space," if the metric is understood or irrelevant. (This is a very common and harmless shortcut in mathematical terminology: For example, when talking about

vector spaces, we typically say, “Let V be a vector space,” with the understanding that it’s not just the set V we’re interested in, but the set V together with its operations of vector addition and scalar multiplication, its zero vector, and its additive inverses.)

If X is a metric space with metric d , the elements of X are usually called its **points**, because we think of X as a “space” having a certain “shape,” rather than just as a set. The number $d(x, y)$ is called the **distance from x to y** .

Example 3.2 (Metric Spaces).

- (a) (\mathbb{R}^n, d) is a metric space, where d is the Euclidean metric.
- (b) Suppose X is any subset of \mathbb{R}^n , and define d again by (3.2). Then conditions (1)–(3) hold automatically because they hold in \mathbb{R}^n . Thus any subset of \mathbb{R}^n is itself a metric space with the Euclidean metric. Unless otherwise is specified, we always consider \mathbb{R}^n and any of its subsets as metric spaces with the Euclidean metric.
- (c) Now let X be any set whatsoever, and define a metric δ on X by the rule

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

This is called the **discrete metric**. The proof that it satisfies (1)–(3) is completely trivial.

3. CONVERGENT SEQUENCES

Once the notion of a metric space is introduced, we can generalize many concepts from real analysis to metric spaces. For instance, the concept of a convergent sequence can be extended to metric spaces in a straightforward way. Suppose (X, d) is a metric space, $(x_n)_{n \in \mathbb{Z}^+}$ is a sequence of points in X , and $a \in X$. We say that **x_n converges to a** if for every $\varepsilon > 0$ there exists a positive integer N such that $n \geq N$ implies $d(x_n, a) < \varepsilon$. Symbolically, the statement “ x_n converges to a ” is denoted by either

$$\lim_{n \rightarrow \infty} x_n = a \quad \text{or} \quad x_n \rightarrow a.$$

Can a sequence in a metric space converge to two distinct points? The answer, not surprisingly, is *no*:

Proposition 3.3. *Let (X, d) be a metric space and let $(x_n)_{n \in \mathbb{Z}^+}$ be a sequence of points in X . If x_n converges to a and also to b , then $a = b$.*

Proof: Assume for contradiction that x_n converges to both a and b with $a \neq b$, and let $r = d(a, b)$. Then, by condition (1) in the definition of a metric, $r > 0$. Take $\varepsilon = r/2 > 0$. Since $x_n \rightarrow a$, there exists a positive integer N_1 such that for all $n \geq N_1$, $d(x_n, a) < \varepsilon = r/2$. Similarly, there exists a positive integer N_2 such that for all $n \geq N_2$, $d(x_n, b) < \varepsilon = r/2$. Hence for all $n \geq \max\{N_1, N_2\}$, we have

$$r = d(a, b) \leq d(a, x_n) + d(x_n, b) = d(x_n, a) + d(x_n, b) < r/2 + r/2 = r,$$

which is clearly impossible. (Note that in the above formula we have used both conditions (2) and (3) in the definition of a metric.) \square

4. CONTINUOUS MAPS BETWEEN METRIC SPACES

If (X, d) and (Y, ρ) are metric spaces, then continuity of a map $f: X \rightarrow Y$ can now be defined exactly as we did for maps between Euclidean spaces: We say that f is **continuous** if for every $a \in X$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in X$ with $d(x, a) < \delta$, we have $\rho(f(x), f(a)) < \varepsilon$.

Example 3.4. If X is a subset of \mathbb{R}^n with the Euclidean metric, then continuity of a function $f: X \rightarrow \mathbb{R}$ in the metric space sense means exactly the same thing as continuity in the sense of ordinary calculus. Thus all the real-valued functions of one or more variables that you already know to be continuous from real analysis, such as polynomial, rational, trigonometric, exponential, logarithmic, and power functions, and functions obtained from them by composition, are continuous on their appropriate domains. (In this course, you may accept the continuity of all such functions without proof.)

Here are some more types of continuous functions that are frequently useful.

Exercise 3.5. A function $f: X \rightarrow Y$ is said to be a **constant function** if there exists $c \in Y$ such that $f(x) = c$ for all $x \in X$. If X and Y are metric spaces, show that every constant function from X to Y is continuous.

Exercise 3.6. Suppose X is a metric space and $i_X: X \rightarrow X$ is the identity function (see Munkres, Exercise 5, p. 21). Show that i_X is continuous.

Exercise 3.7. Suppose X is a subset of \mathbb{R}^n with the Euclidean metric. Show that the function $f: X \rightarrow \mathbb{R}$ given by $f(x) = \|x\|$ is continuous.

Exercise 3.8. Suppose (X, d) is any metric space, and $f, g: X \rightarrow \mathbb{R}$ are continuous functions. Prove that the function $h: X \rightarrow \mathbb{R}$ defined by $h(x) = \max(f(x), g(x))$ is continuous.

Exercise 3.9. Let (X, δ) be a metric space with the discrete metric.

- (a) Prove that a sequence (x_n) in X converges if and only if it is eventually constant. (It is up to you to figure out what “eventually constant” means.)
- (b) Prove that if (Y, d) is any metric space, and $f: X \rightarrow Y$ any function, then f is continuous! (Later in the course we will be able to show that if $f: \mathbb{R}^n \rightarrow X$ is a continuous function, then f is constant.)

There is a useful relationship between continuity of functions and convergent sequences, given by the following lemma.

Theorem 3.10. *Let (X, d) and (Y, ρ) be metric spaces, and let $f: X \rightarrow Y$ be a function. Then f is continuous if and only if for every convergent sequence $x_n \rightarrow x$ in X , the sequence $f(x_n)$ converges to $f(x)$ in Y .*

Proof: Suppose first that f satisfies the sequence condition: $x_n \rightarrow x$ implies $f(x_n) \rightarrow f(x)$. Let $x \in X$ be arbitrary, and let $\varepsilon > 0$ be given. We wish to show that there exists $\delta > 0$ such that $d(y, x) < \delta$ implies $\rho(f(y), f(x)) < \varepsilon$. Assume for the sake of contradiction that there is no such δ . Then, in particular, for each $n \in \mathbb{Z}^+$, we can take $\delta = 1/n$, and we find that there must exist some point

$x_n \in X$ such that $d(x_n, x) < 1/n$ but $\rho(f(x_n), f(x)) \geq \varepsilon$. The first fact implies that $x_n \rightarrow x$, but the second implies that $f(x_n)$ does not converge to $f(x)$, which contradicts our assumption.

The converse is left as an easy exercise. \square

Exercise 3.11. Complete the proof of the preceding theorem by showing that if f is continuous, then $x_n \rightarrow x$ implies $f(x_n) \rightarrow f(x)$.

5. HOMEOMORPHISMS

The business of topology, roughly speaking, is studying properties of geometric objects that are unchanged by “continuous deformations.” To make this idea precise in the context of metric spaces, we make the following definition: Suppose (X, d) and (Y, ρ) are metric spaces. A map $f: X \rightarrow Y$ is called a **homeomorphism** if it is continuous and bijective, and its inverse map $f^{-1}: Y \rightarrow X$ is also continuous. The idea is that f sets up a 1-to-1 correspondence between the points of X and the points of Y , in such a way that “nearness” in X corresponds to “nearness” in Y . Note that we do *not* require that f preserve distances exactly: It need not be the case that $\rho(f(x), f(y)) = d(x, y)$.

The fundamental idea of topology is that we wish to consider two metric spaces X and Y to be “the same” if there is a homeomorphism between them. If there does exist such a homeomorphism, we say that X and Y are **homeomorphic** (from the Greek for “similar form”) or **topologically equivalent**.

To determine whether a given map f is bijective, it is often easiest to try to find the inverse map explicitly (which can frequently be done by solving the equation $f(x) = y$ for x), instead of proving directly that f is injective and surjective. The following lemma shows that the existence of a (two-sided) inverse map implies bijectivity.

Lemma 3.12. *Suppose X and Y are sets, and $f: X \rightarrow Y$ is a map. If there exists a map $g: Y \rightarrow X$ such that $g \circ f = i_X$ and $f \circ g = i_Y$, then f is bijective and $g = f^{-1}$.*

Proof: This is just the special case of Munkres’s Lemma 2.1 when $g = h$. \square

It is important to observe, though, that continuity of f and existence of the inverse map are not sufficient to conclude that f is a homeomorphism. The continuity of the inverse map is something that needs to be checked separately, because it is entirely possible for a map to be continuous and bijective but to have a discontinuous inverse. Here are two examples.

Exercise 3.13. Let d be the Euclidean metric on \mathbb{R} , and let δ be the discrete metric on \mathbb{R} . Show that the identity map from (\mathbb{R}, δ) to (\mathbb{R}, d) is continuous and bijective, but its inverse is not continuous.

Exercise 3.14. Let S^1 denote the unit circle in the plane:

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

Define a map $F: [0, 2\pi) \rightarrow S^1$ by

$$F(\theta) = (\cos \theta, \sin \theta).$$

Prove that F is continuous and bijective, but is not a homeomorphism.

Here are some examples of homeomorphisms between metric spaces.

Example 3.15 (Homeomorphic Metric Spaces).

- (a) Let D_1 and D_2 be the following disks in the plane:

$$D_1 = \{(x, y) : x^2 + y^2 < 1\},$$

$$D_2 = \{(x, y) : x^2 + y^2 < 4\}.$$

The map $F: D_1 \rightarrow D_2$ given by $F(x, y) = (2x, 2y)$ is obviously continuous and bijective, with inverse map given by $F^{-1}(x, y) = (x/2, y/2)$. Thus D_1 and D_2 are homeomorphic.

- (b) Consider the open interval $(-1, 1) \subset \mathbb{R}$, and define a map $f: (-1, 1) \rightarrow \mathbb{R}$ by

$$f(x) = \frac{x}{1 - |x|}.$$

Then f is continuous because it is formed by composition from the maps $x \mapsto |x|$, $x \mapsto -x$, $x \mapsto 1 + x$, all of which are continuous for all x , and $x \mapsto 1/x$, which is continuous on the set where $x \neq 0$. The map f is bijective because we can construct its inverse map directly. Define $g: \mathbb{R} \rightarrow (-1, 1)$ by

$$g(y) = \frac{y}{1 + |y|}.$$

This is continuous by the same argument, and a straightforward computation shows that $g \circ f(x) = x$ and $f \circ g(y) = y$ (to see this, it is easiest to consider separately the cases $x \geq 0$, $x < 0$, $y \geq 0$, and $y < 0$), so f is bijective and $g = f^{-1}$ by Lemma 3.12. Thus the bounded interval $(-1, 1)$ is homeomorphic to the real line.

- (c) Next let $\mathbb{Z} \subset \mathbb{R}$ be the set of integers, and consider two metrics on \mathbb{Z} : the Euclidean metric d , and the discrete metric δ . Let $i_{\mathbb{Z}}: (\mathbb{Z}, \delta) \rightarrow (\mathbb{Z}, d)$ be the identity map: $i_{\mathbb{Z}}(n) = n$. Then it follows from Exercise 3.9 that $i_{\mathbb{Z}}$ is continuous. Obviously $i_{\mathbb{Z}}$ is bijective (it is its own inverse!). To see that $i_{\mathbb{Z}}^{-1} = i_{\mathbb{Z}}$ is continuous from (\mathbb{Z}, d) to (\mathbb{Z}, δ) , we will work directly from the definition: Let $n \in \mathbb{Z}$ be arbitrary, and let $\varepsilon > 0$ be given. If $m \in \mathbb{Z}$ and $d(m, n) < 1$, then $m = n$, and so certainly $d(m, n) < 1$ implies $\delta(i_{\mathbb{Z}}(m), i_{\mathbb{Z}}(n)) < \varepsilon$.
- (d) Finally, let $S^1 \subset \mathbb{R}^2$ denote the unit circle, and let $C \subset \mathbb{R}^2$ be the following square:

$$C = \{(x, y) \in \mathbb{R}^2 : \max(|x|, |y|) = 1\}.$$

Let $F: C \rightarrow S^1$ be the map that sends each point on C to the unit vector pointing in the same direction:

$$F(x, y) = \frac{(x, y)}{\sqrt{x^2 + y^2}}.$$

Geometrically, F projects C radially inward to the circle. This map is continuous on C by the usual arguments of elementary analysis (notice that the denominator is always nonzero on C). The next exercise shows that F is a homeomorphism.

Exercise 3.16. Show that the map $F: C \rightarrow S^1$ is a homeomorphism by showing that its inverse can be written

$$F^{-1}(x, y) = \frac{(x, y)}{\max(|x|, |y|)}.$$

The preceding examples illustrate that many geometric properties, such as lengths, areas, boundedness, and corners, are not preserved by homeomorphisms, and thus are not topological properties. It might not be clear at this point exactly what topological properties *are*, let alone how one might rigorously define and compute them. Figuring out how to do so will be the main goal of this course.

6. ANOTHER FORMULATION OF CONTINUITY

Before we start to study topological properties, however, let's think again about the role of the metric. The examples above show that, for the purposes of determining topological properties, a metric contains lots of irrelevant information, because homeomorphisms do not need to preserve distances. All we really need from the metric is a sense of "nearness," so that we can define what it means for functions to be continuous. In this section, we reformulate the notion of continuity in a way that makes the metric recede to the background.

If (X, d) is a metric space and x is a point of X , for any $\varepsilon > 0$ we define the **ε -ball centered at x** to be the set

$$B(x, \varepsilon) = \{y \in X : d(x, y) < \varepsilon\}.$$

If it is important to specify which metric we are using to define the ball, then we will use the notation $B_d(x, \varepsilon)$. The definition of continuity can be immediately reformulated in terms of ε -balls as follows. Suppose X and Y are metric spaces. Then

(3.3) a map $f: X \rightarrow Y$ is continuous if and only if for every $x \in X$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that $f(B(x, \delta)) \subset B(f(x), \varepsilon)$.

Or, to state the inclusion another way in terms of preimages,

(3.4) a map $f: X \rightarrow Y$ is continuous if and only if for every $x \in X$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that $B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon))$.

We have not yet eliminated the metric from our definition, because it is used explicitly in the definition of balls. So we need yet another reformulation that pushes the metric even farther into the background.

Given a metric space (X, d) , we say that a subset $A \subset X$ is **open** if for every $a \in A$ there exists $\varepsilon > 0$ such that $B(a, \varepsilon) \subset A$.

Exercise 3.17. Let (X, d) be any metric space. Prove that open sets in X have the following properties.

- The empty set is open.
- X itself is open.
- For any $\varepsilon > 0$ and any $x \in X$, the ball $B(x, \varepsilon)$ is open.
- If U_1, \dots, U_n are finitely many open subsets of X , then their intersection $U_1 \cap \dots \cap U_n$ is open.

- (e) If $\{U_\alpha\}_{\alpha \in J}$ is any collection (finite or infinite) of open sets, then their union $\bigcup_{\alpha \in J} U_\alpha$ is open.

(When you do these proofs, be sure that you use only the properties that are given by the axioms for a metric and the definition of balls. As a general rule, pictures can guide your intuition, and can and should be used to illustrate your proofs, but be careful, because not all of the intuition you acquired in the Euclidean world carries over to general metric spaces. For example, when X has the discrete metric, $B(x, \varepsilon)$ is either $\{x\}$ or the entire space X !)

Now comes the payoff: We can reformulate continuity purely in terms of open sets.

Theorem 3.18. *Let X and Y be metric spaces, and let $f: X \rightarrow Y$ be a map. Then f is continuous if and only if for every open set $U \subset Y$, the preimage $f^{-1}(U)$ is open in X .*

Proof: First suppose that f is continuous, and let U be an open subset of Y . We must show that $f^{-1}(U)$ is open in X , i.e., that for every $x \in f^{-1}(U)$ we can find $\delta > 0$ such that $B(x, \delta) \subset f^{-1}(U)$. Now since U is open in Y , we can find $\varepsilon > 0$ such that $B(f(x), \varepsilon) \subset U$. Since f is continuous, the characterization (3.4) of continuity guarantees that we can find $\delta > 0$ such that $B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon))$. Then

$$B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon)) \subset f^{-1}(U),$$

as required.

Conversely, suppose that for every open set $U \subset Y$, the preimage $f^{-1}(U)$ is open in X . Let $x \in X$ and suppose we are given $\varepsilon > 0$. Then we must find $\delta > 0$ such that $B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon))$. But $B(f(x), \varepsilon)$ is an open set by Exercise 3.17, and hence $f^{-1}(B(f(x), \varepsilon))$ is open by assumption. Therefore, the required δ exists by the definition of an open set. \square

The moral of this story is that to detect continuity, and therefore to detect which metric spaces are homeomorphic, all we really need to know is which subsets are open. Of course, we used the metric to define the open sets. But if we could find other reasonable ways to decide which sets are open, without reference to any metric, we might be able to develop a *qualitative* theory of space and “nearness” without any irrelevant *quantitative* metric data. This brings us to the concept of a *topological space*, and we may return to the text, Chapter 2, p. 75.