3 Continuity

3 **Continuity** 41
3.1 The definition of continuity 41
3.2 Limits of functions 45
3.3 Continuity and compactness 48
3.4 Uniform continuity and compactness 50
3.5 Continuity and connectedness 52

3.1 The definition of continuity

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous at the point $a \in \mathbb{R}$ if for every $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$. Generalizing this definition is straightforward.

Definition 3.1.1. Suppose that (X, d_X) and (Y, d_Y) are metric spaces. The function $f : X \to Y$ is *continuous at the point* $\boldsymbol{a \in X}$ if for every $\epsilon > 0$, there is a $\delta > 0$ such that $d_Y(f(x), f(a)) < \epsilon$ whenever $d_X(x, a) < \delta$.

A function is simply called *continuous* if it is continuous on its entire domain.

Definition 3.1.2. Suppose (X, d_X) and (Y, d_Y) are metric spaces. The function $f : X \to Y$ is *continuous* if and only if it is continuous at every point $a \in X$.

We pause to consider a few examples.

Example 3.1.3. Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$ is a *constant function*, so there is some $y_0 \in Y$ so that $f(x) = y_0$ for all $x \in X$. Then f is continuous.

Proof. For every $x, a \in X$ and every $\epsilon > 0$ we have $d_Y(f(x), f(a)) = d_Y(y_0, y_0) = 0 < \epsilon$. □

Example 3.1.4. Let (X, d_X) be a metric space and fix a point $x_0 \in X$. Define the function $f : X \to \mathbb{R}$ by $f(x) = d_X(x, x_0)$, where the metric on $\mathbb{R}$ is the usual one. Then f is continuous.

Proof. Let $a \in X$ and $\epsilon > 0$ be arbitrary. For every $x \in X$, we have (by the triangle inequality) that

$$|f(x) - f(a)| = |d_X(x, x_0) - d_X(a, x_0)| \leq |d_X(x, a)| = d_X(x, a).$$

Thus if $d_X(x, a) < \epsilon$, then $|f(x) - f(a)| < \epsilon$. Since $\epsilon > 0$ was arbitrary, this proves that f is continuous at a, and since a was arbitrary, that proves that f is continuous on all of X. □

Example 3.1.5. Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$. Suppose that there is a constant $c \in (0, \infty)$ such that for all $x_1, x_2 \in X$, we have

$$d_Y(f(x_1), f(x_2)) \leq c d_X(x_1, x_2).$$

Then f is continuous.

Functions satisfying the conditions of Example 3.1.5 are called *Lipschitz continuous*, and more can and will be said about them.

Proof. If $c = 0$, then f is a constant function and we have already seen that it is continuous. Suppose that $c > 0$ and let $a \in X$ and $\epsilon > 0$ be arbitrary. If $d_X(x, a) < \epsilon / c$, then we have

$$d_Y(f(x), f(a)) \leq c d_X(x, a) < \epsilon,$$

which proves that f is continuous. □

Example 3.1.6. Fix a vector $z \in \mathbb{R}^g$ and define the function $f_z : \mathbb{R}^g \to \mathbb{R}$ by $f_z(x) = \langle x, z \rangle$, where $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{R}^g$. Viewed as a function from $(\mathbb{R}^g, d_2)$ to $\mathbb{R}$ with the usual metric, f_z is continuous.

Proof. Let $x_1, x_2 \in \mathbb{R}^g$. By the linearity of the inner product, we have

$$f(x_1) - f(x_2) = \langle x_1, z \rangle - \langle x_2, z \rangle = \langle x_1 - x_2, z \rangle.$$

Furthermore, by the Cauchy–Schwarz inequality, we have

$$|f(x_1) - f(x_2)| = |\langle x_1 - x_2, z \rangle \leq \|x_1 - x_2\|_2 \, \|z\|_2 = \|z\|_2 \, d_2(x_1, x_2).$$

It follows that f_z satisfies the conditions of our previous example with $c = \|z\|_2$, and thus it is continuous. □

Next we establish two very important results. The first of these gives an alternative, and frequently more useful, definition of continuity: *a function is continuous if and only if the preimage of every open set is open.*

Proposition 3.1.7. *Suppose (X, d_X) and (Y, d_Y) are metric spaces. The function $f : X \to Y$ is continuous if and only if for every open set U in (Y, d_Y), the set*

$$f^{-1}(U) = \{x \in X : f(x) \in U\} \subseteq X$$

is open in (X, d_X).

Proposition 3.1.7 also holds with both instances of the word "open" replaced by "closed"; see Exercise 3.1.4.

The set $f^{-1}(U)$ is called the *preimage* (or *inverse image*) *of* U, but note that its definition does not require the invertibility of f. One should think of the notation "f^{-1}" here as being completely different from the notation "f^{-1}" we use for invertible functions. (And it is even more different than the negative power notation we use when we write "$x^{-1} = 1/x$".)

Proof. Assume that the function f is continuous and that $U \subseteq Y$ is open. Let $a \in f^{-1}(U)$ be arbitrary. Since U is open and $f(a) \in U$, there is an $\epsilon > 0$ so that $N_\epsilon(f(a)) \subseteq U$. Since f is continuous at a, there is a $\delta > 0$ such that if $x \in N_\delta(a)$, then $f(x) \in N_\epsilon(f(a))$. This shows that $N_\delta(a) \subseteq f^{-1}(U)$, completing the proof of this direction of the result.

Conversely, suppose that $f^{-1}(U)$ is open in (X, d_X) whenever U is open in (Y, d_Y). Let $a \in X$ and $\epsilon > 0$ be arbitrary. The set $U = N_\epsilon(f(a))$ is open in (Y, d_Y), so $f^{-1}(U)$ must be open in (X, d_X). Since $a \in f^{-1}(U)$, this means that there is a $\delta > 0$ such that $N_\delta(a) \subseteq f^{-1}(U)$. This means that if $d_X(x, a) < \delta$, then $d_Y(f(x), f(a)) < \epsilon$. Since $\epsilon > 0$ was arbitrary, this verifies that f is continuous at a, and since $a \in X$ was arbitrary, this proves that f is continuous. □

This definition of continuity in terms of open sets allows one to extend the definition to arbitrary topological spaces, although we will content ourselves with metric spaces.

Notice how easily the viewpoint of Proposition 3.1.7 allows us to prove the following.

Proposition 3.1.8. *Suppose (X, d_X), (Y, d_Y), and (Z, d_Z) are all metric spaces, and that $f : X \to Y$ and $g : Y \to Z$. If f and g are both continuous, then so is their composition $h = g \circ f : X \to Z$.*

Proof. Suppose that $U \subseteq Z$ is open in (Z, d_Z). Since g is continuous, $g^{-1}(U)$ is open in (Y, d_Y). Then since f is continuous, $f^{-1}(g^{-1}(U))$ is open in (X, d_X). Therefore $h^{-1}(U) = f^{-1}(g^{-1}(U))$ is open in (X, d_X), proving the result. □

There is also a local version of this result. We leave the proof to the reader.

Proposition 3.1.9. *Suppose (X, d_X), (Y, d_Y), and (Z, d_Z) are all metric spaces, and that $f : X \to Y$ and $g : Y \to Z$. If f is continuous at the point $a \in X$ and g is continuous at the point $f(a) \in Y$, then $g \circ f$ is continuous at the point a.*

Exercises

Exercise 3.1.1. Let (X, d_{disc}) be a discrete metric space and let (Y, d_Y) be any metric space. Determine all continuous functions $f : X \to Y$.

Exercise 3.1.2. Let $(\mathbb{R}, d)$ be real line with the standard metric and let (Y, d_{disc}) be a discrete metric space. Determine all continuous functions $f : \mathbb{R} \to Y$.

Exercise 3.1.3. Suppose (X, d_X) and (Y, d_Y) are metric spaces. Prove that the function $f : X \to Y$ is continuous at the point $a \in X$ if and only if for every open set $U \subseteq Y$ containing $f(a)$, there is an open set $V \subseteq X$ containing a such that $V \subseteq f^{-1}(U)$.

Exercise 3.1.3 is a "local version" of the open sets definition of continuity from Proposition 3.1.7.

Exercise 3.1.4. Suppose that (X, d_X) and (Y, d_Y) are metric spaces. Prove that the function $f : X \to Y$ is continuous if and only if for every *closed* set C in (Y, d_Y), the set

$$f^{-1}(C) = \{x \in X : f(x) \in C\} \subseteq X$$

is *closed* in (X, d_X).

Exercise 3.1.5. Let (X, d) be a metric space and $A \subseteq X$ be nonempty. Define $f : X \to [0, \infty)$ by $f(x) = \inf\{d(x, a) : a \in A\}$. Prove that f is continuous.

3.2 Limits of functions

IN THIS SECTION WE DEFINE what it means for a general function to have a limit at a point. Recall from our study of functions on the real line that we are only "allowed" to consider limits at certain points. We begin by (re)defining those.

Definition 3.2.1. Suppose (X, d_X) is a metric space and $A \subseteq X$. The point $a \in X$ is a *limit point of* A if for every $\delta > 0$, the set $A \cap N_\delta(a)$ is infinite.

Note that limit points of a set need not lie in the set themselves. We look a few examples of this definition before moving on to consider limits themselves.

Example 3.2.2. The set $\{1/n : n \in \mathbb{N}^+\}$ in the metric space $(\mathbb{R}, d)$ has only one limit point: 0.

Example 3.2.3. If E is an open set in $(\mathbb{R}^n, d_2)$, then every point of E is a limit point of E, but E may have more limit points.

Example 3.2.4. Sets in a discrete metric space never have any limit points.

Note that continuity comes "for free" at non-limit points, as recorded in the following result.

Proposition 3.2.5. *Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that the point $a \in X$ is not a limit point of X. Then every function $f : X \to Y$ is continuous at a.*

We now define limits.

Definition 3.2.6. Let (X, d_X) and (Y, d_Y) be metric spaces and $A \subseteq X$. Suppose that $a \in X$ is a limit point of A and that $b \in Y$. The function $f : A \to Y$ has *limit b at the point a*, written

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

if for every $\epsilon > 0$, there is a $\delta > 0$ such that $d_Y(f(x), b) < \epsilon$ whenever $0 < d_X(x, a) < \delta$.

Proposition 3.2.7. *If a function has a limit at a point, then that limit is unique.*

We leave the proof of Proposition 3.2.7 to the reader (Exercise 3.2.3).

The reader may recognize our next result as saying that we may "fill in removable singularities" to get a continuous function.

Proposition 3.2.8. *Let (X, d_X) and (Y, d_Y) be metric spaces, let $A \subseteq X$, and suppose that $a \in X$ is a limit point of A. If $f : A \to Y$ has a limit at a, then the function $g : A \cup \{a\} \to Y$ defined by*

$$g(x) = \begin{cases} f(x) & \text{if } x \neq a, \\ \lim_{x \to a} f(x) & \text{if } x = a, \end{cases}$$

is continuous at a.

Note in Proposition 3.2.8 that $A \cup \{a\}$ may well equal A, and g may well equal f, in which case the result is just saying that f is continuous at a.

The following result gives a sequential formulation of limit.

Proposition 3.2.9 (Sequential criterion for limits)**.** *Let (X, d_X) and (Y, d_Y) be metric spaces, let $A \subseteq X$, and suppose that $a \in X$ is a limit point of A. Then*

$$\lim_{x \to a} f(x)$$

exists and equals $b \in Y$ if and only if for every sequence (a_n) from $A \setminus \{a\}$ that converges to a, the sequence $(f(a_n))$ converges to b.

In the statement of Proposition 3.2.9, note that we can have $A \setminus \{a\} = A$ if we are considering a limit at a point outside the domain of f.

Proof. Suppose that the limit of f at a is b and that (a_n) is a sequence from $A \setminus \{a\}$ that converges to a. To see that $(f(a_n))$ converges to b, let $\epsilon > 0$ be arbitrary. There is a $\delta > 0$ such that if $0 < d_X(x, a) < \delta$, then $d_Y(f(x), b) < \epsilon$. There is also an $N \in \mathbb{N}$ such that if $n \geq N$, then $0 < d_X(a_n, a) < \delta$. Hence, if $n \geq N$, then $d_Y(f(a_n), b) < \epsilon$, which proves that $(f(a_n))$ converges to b.

Conversely, suppose that the limit of f at a does not equal b (so either it exists and doesn't equal b, or it simply doesn't exist). Then there is some $\epsilon > 0$ such that for every $n \in \mathbb{N}^+$, we can find a point within distance $1/n$ of a where the value of f is at least distance ϵ away from b. In other words, we can construct a sequence $(a_n)_{n=1}^{\infty}$ in $A \setminus \{a\}$ such that for every $n \in \mathbb{N}^+$, $0 < d_X(a_n, a) < 1/n$ and $d_Y(f(a_n), b) > \epsilon$. This completes the proof, by showing that if the limit of f at a is not b, then we can find a sequence (a_n) from $A \setminus \{a\}$ for which $(f(a_n))$ does not converge to b. □

From the sequential criterion for limits, we quickly obtain a sequential criterion for continuity.

Proposition 3.2.10 (Sequential criterion for continuity)**.** *Let (X, d_X) and (Y, d_Y) be metric spaces, let $A \subseteq X$, and suppose that the point $a \in A$ is a limit point of A. Then the function $f : A \to Y$ is continuous at a if and only if for every sequence (a_n) from A that converges to a, the sequence $(f(a_n))$ converges to $f(a)$.*

Note that in Proposition 3.2.10, we require the point a to lie in the domain of f. By the very definition, functions cannot be continuous at points outside their domains.

Proof. This follows immediately from the sequential criterion for limits and Proposition 3.2.8. □

We conclude with a result about limits of compositions of functions.

Proposition 3.2.11. *Let (X, d_X), (Y, d_Y), and (Z, d_Z) all be metric spaces, let $A \subseteq X$, and suppose that $a \in X$ is a limit point of A. If $\lim_{x \to a} f(x) = b$ and if $g : Y \to Z$ is continuous at b, then $\lim_{x \to a} g \circ f(x) = g(b)$.*

Proof. The function $h : X \to Y$ defined by $h(x) = f(x)$ if $x \neq a$ and $h(a) = b$ is continuous at a by Proposition 3.2.8. Hence the function $g \circ h$ is continuous at a by Proposition 3.1.9. It follows from Proposition 3.2.8 that

$$\lim_{x \to a} g \circ f(x) = \lim_{x \to a} g \circ h(x) = g(h(a)) = g(b),$$

completing the proof. □

Exercises

Exercise 3.2.1. Suppose (X, d_X) is a metric space and $A \subseteq X$. Prove that the point $a \in X$ is a limit point of A if and only if for every $\delta > 0$, the set $A \cap N_\delta(a)$ contains at least one point other than a.

Since Exercises 3.2.1 and 3.2.2 are both if and only if statements, they provide alternative definitions of limit points.

Exercise 3.2.2. Suppose (X, d_X) is a metric space and $A \subseteq X$. Prove that the point $a \in X$ is a limit point of A if and only if there is a sequence (a_n) from $A \setminus \{a\}$ that converges to a.

Exercise 3.2.3. Prove Proposition 3.2.7: if (X, d_X) and (Y, d_Y) are metric spaces, $A \subseteq X$, and the function f has limit b and b' at the limit point $a \in A$, then $b = b'$.

3.3 Continuity and compactness

CONTINUITY AND COMPACTNESS interact in a very convenient way, as we see in this section, with the following theorem and two of its corollaries.

Theorem 3.3.1. *Let (X, d_X) and (Y, d_Y) be metric spaces. If the function $f : X \to Y$ is continuous and the set $K \subseteq X$ is compact, then the image*

$$f(K) = \{f(x) : x \in K\}$$

is also compact.

The continuous image of a compact set is compact.

Proof. Let $\mathcal{W}$ be an arbitrary open cover of $f(K)$. The family

$$\{f^{-1}(W) : W \in \mathcal{W}\}$$

is a cover of K by definition (every point $x \in K$ has an image in $f(K)$, so it lies in $f^{-1}(W)$ for some set $W \in \mathcal{W}$), and this family is an open cover of K because f is continuous (so the preimage of an open set is open). Since K is compact, this open cover has a finite subcover. Thus there is some finite subfamily $\mathcal{F} \subseteq \mathcal{W}$ so that

$$K \subseteq \bigcup_{W \in \mathcal{F}} f^{-1}(W).$$

Our goal is to prove that $\mathcal{F}$ covers $f(K)$.

For any set $W \subseteq Y$, we have $W \supseteq f(f^{-1}(W))$ because if $y \in Y$ has a preimage, then y is (obviously) the image of that preimage. We therefore have

$$\bigcup_{W \in \mathcal{F}} W \supseteq \bigcup_{W \in \mathcal{F}} f(f^{-1}(W)).$$

This union of images is the same the image of the union, so we can rewrite this as

$$\bigcup_{W \in \mathcal{F}} f(f^{-1}(W)) = f\left(\bigcup_{W \in \mathcal{F}} f^{-1}(W)\right).$$

We have assumed that $\{f^{-1}(W) : W \in \mathcal{F}\}$ covers K, and thus

$$f\left(\bigcup_{W \in \mathcal{F}} f^{-1}(W)\right) \supseteq f(K).$$

Therefore $\mathcal{F}$ is a finite subcover of the given open cover of $f(K)$, proving the result. □

We prove Theorem 3.3.1 using the definition of compactness here, but it can also be proved from the definition of sequential compactnessness. Let $(f(x_n))$ be a sequence in $f(K)$. Since K is sequentially compact, there is a subsequence (x_{n_k}) that converges to some point $x \in K$. Then since f is continuous, the subsequence $(f(x_{n_k}))$ converges to $f(x)$, proving that $f(K)$ is also sequentially compact.

Corollary 3.3.2 (Extreme value theorem)**.** *Let (X, d_X) be a nonempty compact metric space. If the function $f : X \to \mathbb{R}$ is continuous, then there is a point $x_{max} \in X$ such that $f(x_{max}) \geq f(x)$ for all $x \in X$.*

Here $\mathbb{R}$ is considered with the standard metric.

The extreme value theorem says that f attains its maximum. By a similar argument, or by applying the same result to $-f$, we see that f also attains its minimum.

Proof. Because the continuous image of a compact set is compact, the set $f(X) \subseteq \mathbb{R}$ is compact. Since all compact sets (in any metric space) are closed and bounded, $f(X)$ is closed and bounded, and since we have assumed that X is nonempty, $f(X)$ is also nonempty. Finally, since nonempty closed and bounded subsets of $\mathbb{R}$ contain their maximums, there is some $M \in f(X)$ such that $M \geq f(x)$ for all $x \in X$, and since $M \in f(X)$, we must have $M = f(x_{\max})$ for some point $x_{\max} \in X$. □

Corollary 3.3.3. *Let (X, d_X) be a compact metric space and (Y, d_Y) be any metric space. If $f : X \to Y$ is a continuous bijection, then its inverse f^{-1} is also continuous.*

Proof. Let U be an arbitrary open set in (X, d_X). To prove that f^{-1} is continuous, we want to show that $(f^{-1})^{-1}(U)$ is open in (Y, d_Y). Furthermore, since f is injective (one-to-one), $(f^{-1})^{-1}(U) = f(U)$, so it suffices to prove that f maps open sets to open sets.

Since U is open, $X \setminus U$ is a closed subset of a compact space, so it is compact itself (by Proposition 2.1.5). The continuous image of a compact set is compact, so $f(X \setminus U)$ is compact in (Y, d_Y). Since all compact sets are closed, $f(X \setminus U)$ is closed. Finally, since f is surjective (onto), $f(X \setminus U) = Y \setminus f(U)$. Finally, because $Y \setminus f(U)$ is closed, $f(U)$ is open in (Y, d_Y), as desired. □

The proof of Corollary 3.3.3 could be shortened by appealing to Exercise 3.1.4. Instead, we essentially solve that exercise here.

Functions that map open sets to open sets are called *open maps*. We leave it to the reader to construct examples showing that functions may be continuous without being open maps, and conversely, that functions may be open maps without being continuous. However, as this result shows, continuous bijections on compact domains are always open maps.

Exercises

Exercise 3.3.1. Suppose that (X, d_X) and (Y, d_Y) are metric spaces and define the metric space (Z, d_Z) where $Z = X \times Y$ and the metric d_Z is given by

$$d_Z((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}.$$

Prove that if the function $f : X \to Y$ is continuous, then the function $F : X \to Z$ defined by $F(x) = (x, f(x))$ is also continuous.

Exercise 3.3.2. Assuming the same hypotheses as the previous exercise, prove that if f is continuous and X is compact, then the *graph* of f,

$$\text{graph}(f) = \{(x, f(x)) : x \in X\} \subseteq Z$$

is compact.

Exercise 3.3.3. A function $f : X \to Y$ is a *homeomorphism* if it is a continuous bijection and its inverse f^{-1} is also continuous. Suppose that $f : X \to Y$ is a homeomorphism. Show that if $Z \subseteq X$, then $f|_Z : Z \to f(Z)$ is also a homeomorphism.

3.4 Uniform continuity and compactness

The definition of uniform continuity for arbitrary metric spaces is (like our initial definition of continuity) a straightforward generalization from its definition on the real line.

Definition 3.4.1. Suppose (X, d_X) and (Y, d_Y) are metric spaces. The function $f : X \to Y$ is *uniformly continuous* if for every $\epsilon > 0$, there is a $\delta > 0$ such that if the points $a, x \in X$ satisfy $d_X(x, a) < \delta$, then $d_Y(f(x), f(a)) < \epsilon$.

We have given a purely topological definition of continuity: *a function is continuous if and only if the preimage of every open set is open*. However, there is no purely topological definition of uniform continuity.

As on the real line, there is no notion of "uniform continuity at a point".

Our next result states that *every continuous function with compact domain is uniformly continuous*. Thus, for example, the function (on the real line) defined by $f(x) = x^2$ is not uniformly continuous, but when restricted to any closed and bounded interval, it is uniformly continuous. It was essentially this fact that we used when we proved that continuous functions are Reimann integrable.

We remind the reader why the function $f(x) = x^2$ is not uniformly continuous (on the real line with the usual metric). Choose $\epsilon = 1$. Given $\delta > 0$, let $a = 2/\delta$ and $x = 2/\delta + \delta/2$. Then $|x - a| = \delta/2 < \delta$, but

$$\begin{aligned} &|f(x) - f(a)| \\ &\quad = \left| \left(\frac{4}{\delta^2} \right) - \left(\frac{4}{\delta^2} + 2 + \frac{\delta^2}{4} \right) \right| \\ &\quad = 2 + \frac{\delta^2}{4} \\ &\quad \geq \epsilon. \end{aligned}$$

Theorem 3.4.2. *Suppose (X, d_X) and (Y, d_Y) are metric spaces. If the function $f : X \to Y$ is continuous and the domain X is compact, then f is uniformly continuous.*

Proof. Let $\epsilon > 0$ be given and suppose that $f : X \to Y$ is continuous and X is compact. For each point $a \in X$ there is a quantity $\delta_a > 0$ (depending on a) such that if $d_X(x, a) < \delta_a$, then $d_Y(f(x), f(a)) < \epsilon/2$.

Define

$$\mathcal{U} = \{N_{\delta_a/2}(a) : a \in X\}.$$

Every set in $\mathcal{U}$ is open, and every point in X lies in at least one of the sets of $\mathcal{U}$ (the neighborhood centered at that point), so $\mathcal{U}$ is an open cover of X. Because X is compact, $\mathcal{U}$ has a finite subcover. Therefore there is a finite subset $F \subseteq X$ such that

$$\mathcal{V} = \{N_{\delta_a/2}(a) : a \in F\} \subseteq \mathcal{U}$$

covers X.

Because F is finite, the set $\{\delta_a : a \in F\}$ has a minimum, which is positive. Define

$$\delta = \min\{\delta_a/2 : a \in F\}.$$

Now suppose that $x, y \in X$ satisfy $d_X(x, y) < \delta$. Because $\mathcal{V}$ covers X, there is a point $a \in F$ such that $y \in N_{\delta_a/2}(a)$. This implies that $f(y)$ is close to $f(a)$. But also, because x is close to y, x is also close to a:

$$d_X(x, a) \leq d_X(x, y) + d_X(y, a) < \delta + \frac{\delta_a}{2} \leq \delta_a.$$

Therefore both x and y are within distance δ_a of the point a, and so we may use the triangle inequality to bound the difference in the function values $f(x)$ and $f(y)$:

$$d_Y(f(x), f(y)) \leq d_Y(f(x), f(a)) + d_Y(f(a), f(y)) < \frac{\epsilon}{2} + \frac{\epsilon}{2},$$

completing the proof. □

Note that the converse to Theorem 3.4.2 is trivial—if a function is uniformly continuous, then it is continuous. Thus we could have stated this result as an "if and only if".

A proof of Theorem 3.4.2 using sequential compactness is outlined in Exercise 3.4.3.

Exercises

Exercise 3.4.1. Let (Y, d_Y) be a metric space, let $L \in Y$, and suppose $f : [0, \infty) \to Y$, where $[0, \infty)$ is considered with the standard metric on $\mathbb{R}$. We say that f *has limit* $L \in Y$ *at infinity*, written

$$\lim_{x \to \infty} f(x) = L,$$

if for every $\epsilon > 0$, there is some $C > 0$ such that if $x > C$, then $d_Y(f(x), L) < \epsilon$. Prove that if $f : [0, \infty) \to Y$ is continuous and has a limit at infinity, then f is uniformly continuous.

Exercise 3.4.2. Let (X, d_X) and (Y, d_Y) be metric spaces. Prove that if $f : X \to Y$ uniformly continuous and (x_n) is a Cauchy sequence from X, then $(f(x_n))$ is Cauchy in Y.

Uniformly continuous functions map Cauchy sequences to Cauchy sequences.

Exercise 3.4.3. Fill in the following outline of an alternate proof of Theorem 3.4.2.

- Suppose to the contrary that f is not uniformly continuous.
- Then there is an $\epsilon > 0$ such that for every $n \in \mathbb{N}$, there are points $x_n, y_n \in X$ such that $d_X(x_n, y_n) < 1/n$, but $d_Y(f(x_n), f(y_n)) \geq \epsilon$.
- There is a choice of indices $n_1 < n_2 < \cdots$ such that both subsequences $(x_{n_k})_k$ and $(y_{n_k})_k$ converge to the same point.
- This last statement contradicts the hypothesis that f is continuous.

3.5 Continuity and connectedness

Here we consider the interplay between continuity and a concept called connectedness. This will allow us to prove a generalization of the intermediate value theorem. Connectedness is often defined by its negation, as below.

Definition 3.5.1 (Connected spaces)**.** The metric space (X, d_X) is *disconnected* if there exist disjoint, nonempty, open sets $U, V \subseteq X$ such that $U \cup V = X$. The metric space (X, d_X) is *connected* if it is not disconnected.

To prove that (X, d_x) is *disconnected*, you just need to exhibit appropriate sets U and V. To prove that (X, d_X) is *connected*, one way is to assume that U and V are disjoint, open sets satisfying $U \cup V = X$, and then show that either U or V must be empty (and thus the other must be all of X).

Most metric spaces we have seen are connected, but a notable exception are the discrete metric spaces, as we consider below.

Example 3.5.2. Let (X, d_{disc}) be a discrete metric space. If $|X| \geq 2$, then X is disconnected, because letting $x \in X$ be arbitrary, we see that $U = \{x\}$ and $V = X \setminus \{x\}$ are disjoint, nonempty, open sets such that $U \cup V = \{x\} \cup (X \setminus \{x\}) = X$.

This example hints at a more "positive" definition of connectedness. The space (X, d_X) is disconnected if and only if there are disjoint, nonempty, open sets $U, V \subseteq X$ such that $U \cup V = X$. In this case $U = X \setminus V$ and $V = X \setminus U$, so since U and V are open by assumption, they are also both closed. In fact, it suffices to have one nontrivial clopen (both closed and open) set: if $U \neq X, \emptyset$ is clopen, then $X \setminus U \neq X, \emptyset$ is also clopen and $X = U \cup (X \setminus U)$. Thus we have proved the following result, which we can take as an alternative definition of connectedness.

Proposition 3.5.3. *The metric space (X, d_X) is connected if and only if the only clopen (both closed and open) subsets of X are $\emptyset$ and X itself.*

To define connectedness for *subsets* of a metric space, we simply apply the previous definition to the induced subspace, as in the following definition.

Definition 3.5.4 (Connected sets)**.** Let (X, d) be a metric space. The set $Y \subseteq X$ is *connected* (resp., *disconnected*) if the metric space $(Y, d|_{Y \times Y})$ is connected (resp., disconnected).

The definition of connected sets has some subtleties because we only need the sets U and V in the definition to be disjoint, nonempty, and open *in the space* $(Y, d|_{Y \times Y})$. By definition, a set is open in $(Y, d|_{Y \times Y})$ if and only if it is *relatively open with respect to* Y in the space (X, d). We proved (Proposition 1.6.4) that a set is relatively open with respect to Y if and only if it is the intersection of Y with an open set of (X, d). See Exercise 3.5.6.

Note that the definition of connected sets is *intrinsic*—whether a set Y is connected depends only on Y and the metric restricted to Y, not on the ambient space X.

Example 3.5.5. On the real line with the standard metric, the set $[0, 2] \setminus \{1\}$ is disconnected. This is because the sets

$$\begin{aligned} U &= [0,1) = (-1,1) \cap ([0,2] \setminus \{1\}) \quad \text{and} \\ V &= (1,2] = (1,3) \cap ([0,2] \setminus \{1\}) \end{aligned}$$

are both disjoint, nonempty, and relatively open with respect to $[0, 2] \setminus \{1\}$.

Having defined connected sets, our goal in the rest of the section is to generalize the intermediate value theorem. This generalization has two parts. First we show that continuous functions preserve connectedness, and second, we show that the only connected subsets of the real line are the intervals.

Theorem 3.5.6. *Let (X, d_X) and (Y, d_Y) be metric spaces. If the function $f : X \to Y$ is continuous and the space (X, d_X) is connected, then its image $f(X)$ is also connected.*

The continuous image of a connected set is connected.

Proof. Suppose U and V are disjoint open subsets of $f(X)$ such that $f(X) = U \cup V$. The sets $A = f^{-1}(U)$ and $B = f^{-1}(V)$ are both open in (X, d_X) because f is continuous. Moreover, since $f(X) = U \cup V$, we must have $X = A \cup B$, and since U and V are disjoint, we must have $A \cap B = \varnothing$ (if there were a point $x \in A \cap B$, then we would have $f(x) \in U \cap V$).

Thus $A, B \subseteq X$ are disjoint open sets such that $A \cup B = X$. Since (X, d_X) is connected by our hypotheses, it must be the case that one of A or B is empty. This implies that one of U or V is empty, and it follows that $f(X)$ is connected. □

It is perhaps slightly shorter to prove Theorem 3.5.6 using Proposition 3.5.3. Suppose that $U \subseteq f(X)$ is clopen. Because f is continuous and U is open, $f^{-1}(U)$ is open. But also by Exercise 3.1.4, because f is continuous and U is *closed*, $f^{-1}(U)$ is *closed*. Thus $f^{-1}(U)$ is clopen in (X, d_X), and because we have assumed that (X, d_X) is connected, we must have $f^{-1}(U) = X$, and thus we must also have $U = f(X)$. □

With Theorem 3.5.6 in hand, the reader may want to revisit Exercise 3.1.2.

Next we characterize the connected subsets of the real line.

Theorem 3.5.7. *The connected subsets of the real line with the standard metric are precisely the intervals.*

Proof. We want to prove that a subset of $\mathbb{R}$ (with the standard metric) is connected if and only if it is an interval. Recall that an interval is a set that satisfies the betweenness property: if x and z lie in the set and $x < y < z$, then y also lies in the set.

One direction is easy. If the set $S \subseteq \mathbb{R}$ is not an interval, then it fails the betweenness property. This means that there are real numbers x, y, and z satisfying $x < y < z$ such that $x, z \in S$ but $y \notin S$. Then can be expressed as the union

$$S = \big((-\infty, y) \cap S\big) \cup \big((y, \infty) \cap S\big)$$

of disjoint and nonempty sets that are relatively open with respect to S, so S is disconnected.

To prove that all intervals are connected, suppose to the contrary that $I \subseteq \mathbb{R}$ is an interval (so it satisfies the betweenness property) but that there are disjoint and nonempty sets U and V that are both relatively open with respect to I, such that $I = U \cup V$. Since U and V are both nonempty, we can choose points $u \in U$ and $v \in V$. Since U and V are disjoint, $u \neq v$. By swapping the roles of U and V

if necessary, we may assume without loss of generality that $u < v$. Because I is an interval, $[u, v] \subseteq I$.

At this point in the proof we have narrowed in on the interval $[u, v]$. The points on the left of $[u, v]$ are contained in U, while the points on the right of $[u, v]$ are contained in V. Somewhere in the middle, points must change from being in U to being in V, but we will show that that is impossible. This is why we define u^* as we do; it is a point where this switch occurs.

The set $U \cap [u, v]$ is therefore bounded and nonempty, so it has a supremum; define

$$u^* = \sup(U \cap [u, v]).$$

We have $u^* \in [u, v] \subseteq I$, so u^* must lie in either U or V, but not both (because they are disjoint). In either case, we find a contradiction.

Suppose first that $u^* \in U$. Because $u^* \notin V$, we have $u^* < v$. Since U is open, there is some $\epsilon > 0$ so that $N_\epsilon(u^*) \subseteq U$. But then $[u^*, u^* + \epsilon) \subseteq U \cap [u, v]$, which contradicts the fact that u^* is an upper bound on $U \cap [u, v]$.

Now suppose that $u^* \in V$. Because $u^* \notin U$, we have $u^* > u$. Furthermore, since V is open in $(I, d_{I \times I})$, there is some $\epsilon > 0$ such that $N_\epsilon(u^*) \subseteq V$. But then $(u^* - \epsilon, u^*] \subseteq V \cap [u, v]$, which contradicts the fact that u^* is the least upper bound of $U \cap [u, v]$. □

Combining Theorems 3.5.6 and 3.5.7, we immediately obtain the intermediate value theorem.

Corollary 3.5.8 (Intermediate value theorem)**.** *Let $I \subseteq \mathbb{R}$ be an interval. If $f : I \to \mathbb{R}$ is continuous, then $f(I)$ is also an interval.*

It is frequently easier to establish a stronger property than connectedness.

Definition 3.5.9 (Path-connected sets)**.** Let (X, d) be a metric space. The set $Y \subseteq X$ is *path-connected* if, given any two points $y_0, y_1 \in Y$, there is a continuous function $f : [0, 1] \to Y$ such that $f(0) = y_0$ and $f(1) = y_1$.

Corollary 3.5.10 (Path-connected implies connected)**.** *Let (X, d) be a metric space. If the set $Y \subseteq X$ is path-connected, then it is connected.*

Proof. Suppose that Y is path-connected, let $U, V \subseteq Y$ be nonempty open sets in $(Y, d|_{Y \times Y})$ with $U \cup V = Y$, and take $y_0 \in U$ and $y_1 \in V$. Because Y is connected, there is a continuous function $f : [0, 1] \to Y$ with $f(0) = y_0$ and $f(1) = y_1$. Theorem 3.5.7 tells us that the interval $[0, 1]$ is connected, so Theorem 3.5.6 implies that $f([0, 1])$ is connected. The sets $U \cap f([0, 1])$ and $V \cap f([0, 1])$ are nonempty, open in $(f([0, 1]), d|_{f([0,1]) \times f([0,1])})$, and their union is all of $f([0, 1])$. Since $f([0, 1])$ is connected, this means that $U \cap f([0, 1])$ and $V \cap f([0, 1])$ must not be disjoint, and thus U and V must not be disjoint either. □

Exercises

Exercise 3.5.1. In which metric spaces is the empty set connected? In which spaces is it disconnected?

Exercise 3.5.2. Prove that in every metric space, singleton sets are connected, but finite sets with two or more elements are not.

Exercise 3.5.3. Determine the connected subsets of a discrete metric space.

Exercise 3.5.4. Determine the path-connected subsets of a discrete metric space.

Exercise 3.5.5. Prove that every neighborhood $N_\epsilon(x)$ in every Euclidean space $(\mathbb{R}^n, d_2)$ is path-connected, and thus connected.

For Exercise 3.5.5, it may be useful to observe that these neighborhoods are *convex*: given any two points $y_0, y_1 \in N_\epsilon(x)$, the line segment between y_0 and y_1 also lies in $N_\epsilon(x)$.

Exercise 3.5.6. Let (X, d) be a metric space and $Y \subseteq X$. Prove that Y is disconnected if and only if there are sets $U, V \subseteq X$ such that

(a) U and V are open in (X, d),
(b) $U \cap V \cap Y = \emptyset$, and
(c) $Y \subseteq (U \cup V)$.

Exercise 3.5.7. Prove that if A, B, C are connected sets in the metric space (X, d_X) and $A \cap B \neq \emptyset$ and $B \cap C \neq \emptyset$, then $A \cup B \cup C$ is connected.

Exercise 3.5.8. Must the intersection of two connected sets be connected?