Connected Sets

Math 331, Handout #4

You probably have some intuitive idea of what it means for a metric space to be "connected." For example, the real number line, $\mathbb{R}$, seems to be connected, but if you remove a point from it, it becomes "disconnected."

However, it is not really clear how to define connected metric spaces in general. Furthermore, we want to say what it means for a subset of a metric space to be connected, and that will require us to look more closely at subsets of metric spaces than we have so far. We start with a definition of connected metric space. Note that, similarly to compactness and continuity, connectedness is actually a topological property rather than a metric property, since it can be defined entirely in terms of open sets.

Definition 1. Let (M, d) be a metric space. We say that (M, d) is a **connected** metric space if and only if M **cannot** be written as a disjoint union $N = X \cup Y$ where X and Y are both non-empty open subsets of M. ("Disjoint union" means that $M = X \cup Y$ and $X \cap Y = \varnothing$.) A metric space that is not connected is said to be **disconnnected**.

Theorem 1. A metric space (M, d) is connected if and only if the only subsets of M that are both open and closed are M and $\varnothing$. Equivalently, (M, d) is disconnected if and only if it has a non-empty, proper subset that is both open and closed.

Proof. Suppose (M, d) is a connected metric space. We must show that the only subsets of M that are both open and closed are M and $\varnothing$. Suppose X is a subset of M that is both open and closed. Let $Y = M \smallsetminus X$. Since a set is open if and only if its complement is closed, it follows that Y is also both open and closed. Furthermore, M is a disjoint union of X and Y. Then, by the definition of connectedness, X and Y cannot both be non-empty. So, one of X and Y is the empty set, and the other is M. This means that X is either $\varnothing$ or M.

Conversely, suppose that the only subsets of M that are both open and closed are M and $\varnothing$. We must show that M is connected. If M were not connected, then it could be written as a disjoint union of two non-empty open subsets, X and Y. But then X would be a non-empty, proper subset of M that is both open and closed, contradicting the hypothesis. So, M must be connected. □

To understand what it means for a *subset* of a metric space to be connected, we note that we can make a subset of a metric space into a metric space in its own right, using the same measure of distance on the subset that is used on the full metric space. That is, given a metric space (M, d) and a subset $X \subseteq M$, we get the metric space (X, d'), where $d'(x, y) = d(x, y)$ for x and y in X. When we do this, we will refer to X as a sub*space* of M rather than simply a sub*set*. (In practice, we would usually just write d and not d', but in this handout, we will use d' to make it clear whether we are talking about the original metric space or a subspace.) We can now define connected subset:

Definition 2. Let (M, d) be a metric space, and suppose that X is a subset of M. We say that X is a **connected subset** or **connected subspace** of M if and only if the subspace (X, d') is a connected metric space.

It is important to understand what it means for a set to be open or closed in a subspace. As an example, consider the closed interval $[1, 2]$ in $(\mathbb{R}, d)$ (where d is the usual metric). In the metric space $([1,2],d')$, we have $B^{d'}_{1/2}(1) = [1, \frac{3}{2})$. So $[1, \frac{3}{2})]$ is an open subset of the subspace $([1,2],d')$, even though it is not open in $\mathbb{R}$. What exactly do open sets in $[1,2]$ look like? Note that $[1, \frac{3}{2}]$ is the intersection of $[1,2)$ with the open subset $(0, \frac{3}{2})$ of $\mathbb{R}$. In fact, this characterizes open sets in a subspace: A set is open in a subspace if and only if it is the intersection of the subspace with an open subset of the full space. A similar result holds for closed subsets. We will prove that fact in the next theorem, but let's assume that it's true. We can then show that the subspace $[1,2)\cup(2,3]$ of $\mathbb{R}$ is disconnected. In fact, The set $[1,2)$ is the intersection of the open set $(-\infty,2)$ with the subspace, so $[0,1)$ is open in the subspace; similarly, $(2,3]$ is the intersection of the open set $(2,\infty)$ with the subspace, so $(2,3]$ is open in the subspace. So the subspace $[1,2)\cup(2,3]$ is a disjoint union of two non-empty open sets, which means that it is not connected. A similar argument shows that if a connected subset X of $\mathbb{R}$ contains two distinct points, then it must also contain every point that lies between those two points. This means any connected subset of $\mathbb{R}$ must be an interval. (Note however that we have not yet proved that intervals are, in fact, connected.)

Theorem 2. Let (M,d) be a metric space and X a subset of M. A subset U of X is open in the subspace (X,d') if and only if there is an open subset $\mathcal{O}$ in M such that $U = X\cap\mathcal{O}$. A subset F of X is closed in the subspace (X,d) if and only if there is a closed subset $\mathcal{C}$ in M such that $F = X\cap\mathcal{C}$.

Proof. Note first that for a point $x\in X$ and $\epsilon>0$, the open ball of radius ϵ about x in X is is the intersection of X with the open ball of radius ϵ about x in M. That is, $B^{d'}_{\epsilon}(x) = X\cap B^{d}_{\epsilon}(x)$.

Suppose that U is an open subset in the subspace (X,d'). For every $u\in U$, there is an $\epsilon_u>0$ such that $B^{d'}_{\epsilon_u}(u)\subseteq U$. Let $\mathcal{O} = \bigcup_{u\in U} B^{d}_{\epsilon}(u)$. (Note that $\mathcal{O}$ is a union of open balls in M.) $\mathcal{O}$ is an open subset of M, and

$$\begin{aligned} X\cap\mathcal{O} &= X\cap\left(\bigcup_{u\in U} B^{d}_{\epsilon}(u)\right)\\ &= \bigcup_{u\in U}\left(X\cap B^{d}_{\epsilon}(u)\right)\\ &= \bigcup_{u\in U} B^{d'}_{\epsilon}(u)\\ &= U \end{aligned}$$

The last equality follows since $B^{d'}_{\epsilon}(u)\subseteq U$ for all u, and and their union clearly includes every $u\in U$.

Conversely, Suppose that U is a subset of X, and $U = X\cap\mathcal{O}$, where $\mathcal{O}$ is an open subset of M. We must show that U is open in (X,d'). Let $u\in U$. We need to find $\epsilon>0$ such that $B^{d'}_{\epsilon}(u)\subseteq U$. Since $u\in\mathcal{O}$ and $\mathcal{O}$ is open in M, there is an $\epsilon>0$ such that $B^{d}_{\epsilon}(u)\subseteq\mathcal{O}$. But then $B^{d'}_{\epsilon} = X\cap B^{d}_{\epsilon}(u)\subseteq X\cap\mathcal{O} = U$. This shows U is open in (X,d').

Turning to the case of closed subsets, note first that for a subset $\mathcal{O}$ of M and a subset F of X, $X\cap\mathcal{O} = X\setminus F$ if and only if $X\cap(M\setminus\mathcal{O}) = F$. So, applying the first part of the theorem,

$$\begin{aligned} F \text{ is closed in } X &\iff X\setminus F \text{ is open in } X\\ &\iff X\setminus F = X\cap\mathcal{O}, \text{ for some open } \mathcal{O}\subseteq M\\ &\iff F = X\cap(M\setminus\mathcal{O}), \text{ for some open } \mathcal{O}\subseteq M\\ &\iff F = X\cap\mathcal{C}, \text{ for some closed } \mathcal{C}\subseteq M \end{aligned}$$

and that completes the proof. □

It is easy to show that connectedness, like compactness, is preserved by continuous functions. That is, the continuous image of a connected metric space is connected.

Theorem 3. Let (A,ρ) and (B,τ) be metric spaces, and suppose that $f\colon A\to B$ be a continuous function from A to B. If A is connected, then its image $f(A)$ is also connected.

Proof. We prove the contrapositive, that is, if $f(A)$ is disconnected, then A is disconnected. Suppose that $f(A)$ is not connected. Since $f(A)$ is not connected, there are open subsets $\mathcal{O}_1$ and $\mathcal{O}_2$ of B such that $f(A)$ is a disjoint union of $\mathcal{O}_1\cap f(A)$ and $\mathcal{O}_2\cap f(A)$, which are open subsets of the subspace $f(A)$, and neither $\mathcal{O}_1\cap f(A)$ nor $\mathcal{O}_2\cap f(A)$ is empty.

Let $U_1=f^{-1}(\mathcal{O}_1)=f^{-1}(\mathcal{O}_1\cap f(A))$, and let $U_2=f^{-1}(\mathcal{O}_2)=f^{-1}(\mathcal{O}_2\cap f(A))$. The fact that $\mathcal{O}_1\cap f(A)$ and $\mathcal{O}_2\cap f(A)$ are disjoint implies that U_1 and U_2 are disjoint. Since f is continuous, U_1 and U_2 are open subsets of A. Since $\mathcal{O}_1\cup\mathcal{O}_2$ includes all of $f(A)$, it follows that $U_1\cup U_2$ covers all of A. Furthermore, U_i is non-empty: Since $\mathcal{O}_i\cap f(A)\neq\varnothing$, there is some $a\in A$ such that $f(a)\in\mathcal{O}_i$. But then $a\in f^{-1}(\mathcal{O}_i)$. Since U_1 and U_2 are disjoint, open, non-empty subsets of A whose union is A, then by definition A is not connected. □

It is certainly not true that the union of connected sets is connected. (Just consider $[1,2]\cup[3,4]$.) However, if a collection of connected sets have a non-empty intersection, then the union is connected.

Theorem 4. Let (M,d) be a metric space and suppose that $\{X_\alpha\,|\,\alpha\in\mathscr{A}\}$ is a collection of subsets of M. If each X_α is connected and $\bigcap_{\alpha\in\mathscr{A}}X_\alpha\neq\varnothing$, then the union $\bigcup_{\alpha\in\mathscr{A}}X_\alpha$ is connected.

Proof. Suppose that the hypotheses of the theorem hold. Let C be the union $C=\bigcup_{\alpha\in\mathscr{A}}X_\alpha$. We need to show C is connected. Let U_1 and U_2 be open sets in M such that C is the disjoint union of $C\cap U_1$ and $C\cap U_2$. Note that $C\subseteq U_1\cup U_2$. To prove that M is connected, it suffices to show that C is actually contained entirely is one of U_1 or U_2. Let α be some element of $\mathscr{A}$, the facts that $X_\alpha\subseteq C$, and C is a disjoint union of $C\cap U_1$ and $C\cap U_2$ imply that X_α is a disjoint union of $X_\alpha\cap U_1$ and $X_\alpha\cap U_2$. Since X_α is connected, it must be entirely contained either in U_1 or in U_2. Without loss of generality, say $X_\alpha\subseteq U_1$, and $X_\alpha\cap U_2=\varnothing$. Now, for any other $\beta\in\mathscr{A}$, it is similarly true that one of $X_\beta\cap U_1$ or $X_\beta\cap U_2$ is empty. But since $X_\alpha\cap X_\beta\neq\varnothing$, and $X_\alpha\cap U_2=\varnothing$, it must be $X_\beta\cap U_2$ that is empty. Since $X_\beta\subseteq U_1$ for all $\beta\in\mathscr{A}$, then C, which is the union of all the X_β, is a subset of U_1. □

We finish with the claim that opened this handout, a complete characterization of the connected subsets of $\mathbb{R}$: A subset of $\mathbb{R}$ is connected if and only if it is an interval. The term "interval" includes bounded intervals of the form $[a,b]$, (a,b), $[a,b)$, or $(a,b]$, as well as infinite intervals of the form $(-\infty,a]$, $(-\infty,a)$, (a,∞), $[a,\infty)$ or $(-\infty,\infty)$. The empty set is also considered to be an interval.

Intervals are characterized by the property that whenever an interval includes two points, it includes all the points that lie between those two points. The fact that sets that have this property are exactly the intervals of the form listed above is then a theorem. I will give only an outline of a proof: For a bounded set, I, that satisfies the interval property, let $a=\text{glb}(I)$ and $b=\text{lub}(I)$. It can be shown that I must include all the points between a and b and is therefor of one of the forms $[a,b]$, $[a,b)$, $(a,b]$, or (a,b): By the definitions of greatest lower bound and least upper bound, if $a<x<b$, then there must be an $a'\geq a$ and a $b'\leq b$ such that a' and b' are in I, and $a'<x<b'$, and $x\in I$ then follows from the interval property. For an unbounded interval, the greatest lower bound is replaced by $-\infty$, or the least upper bound is replaced by ∞, or both.

Theorem 5. A subset of $\mathbb{R}$ is connected if and only if it is an interval.

Proof. Suppose that C is a connected subset of $\mathbb{R}$. To show that C is an interval, suppose $a,b\in C$ with $a<b$, and let x satisfy $a<x<b$. If x were not in C, then C would be a disjoint union of the non-empty sets $(-\infty,x)\cap C$ and $(x,\infty)\cap C$, contradicting the fact that C is connected.

To prove the converse, we need to show show that any interval is connected. We show here that a closed bounded interval $[a,b]$ is connected. The remainder of the proof is left as an exercise.

Suppose for the sake of contradiction that the closed, bounded interval $[a, b]$ is not connected. Then $[a, b]$ can be written as a disjoint union, $A = U \cup V$, where U and V are non-empty subsets of the subspace $[a, b]$ that are both open and closed in that subspace. Since U is closed in $[a, b]$, then $U = F \cap [a, b]$ for some subset of $\mathbb{R}$ that is closed in $\mathbb{R}$. But since $[a, b]$ is itself closed in $\mathbb{R}$, and the intersection of closed sets is closed, we see that U itself is a closed subset of $\mathbb{R}$.

Since $[a, b] \subseteq U \cup V$ and $V \neq \varnothing$, we can assume, without loss of generality, that $b \in V$. Let c be the least upper bound of U. Because U is a closed subset of $\mathbb{R}$, Exercise 8 from Handout 1 implies that $c \in U$. Since b is an upper bound for U, $c \leq b$. Since V is open in $[a, b]$ and $b \in V$, V must contain an interval $(b - \epsilon, b]$ for some $\epsilon > 0$. Since U and V are disjoint, that interval cannot intersect U. This implies that in fact, $c < b$.

Now, the facts that U is open in $[a, b]$, and $c < b$ imply that U contains an interval $[c, c + \delta)$ for some $\delta > 0$. But then the fact that $c + \frac{1}{2}\delta \in U$ contradicts the fact that c is an upper bound for U. This contradiction proves that the assumption that $[a, b]$ is not connected is false. $\square$

Exercises

Exercise 1. Consider the metric space $(\mathbb{Q}, d)$, using the usual metric that $\mathbb{Q}$ inherits from $\mathbb{R}$. Suppose that C is a non-empty, connected subset of $\mathbb{Q}$. Show that C consists of a single point. (Hint: Suppose that C contains two distinct points. Let λ be an irrational number that lies between those two points.) A metric space in which every connected subset consists of a single point is said to be **totally disconnected**, so this exercise shows that $\mathbb{Q}$ is totally disconnected.

Exercise 2. Suppose $f\colon \mathbb{R} \to \mathbb{Q}$ is a continuous function, where $\mathbb{R}$ and $\mathbb{Q}$ have their usual metrics. Show that f is a constant function. (This is trivial, using the previous exercise and a theorem from this handout.) Show that the function $g\colon \mathbb{Q} \to \mathbb{R}$ defined by $g(x) = \begin{cases} 1 & \text{if } x < \sqrt{2} \\ 2 & \text{if } x > \sqrt{2} \end{cases}$ is continuous, in spite of the jump.

Exercise 3. Suppose that (M, d) is a metric space and A is a subset of M. Show that the function $i\colon A \to M$ given by $i(x) = x$ is continuous, when A is considered as a subspace of M. Next, suppose that (N, e) is another metric space and that $f\colon M \to N$ is a continuous function. Show that the restriction of f to A is a continuous function from A to N. Is every continuous function from A to N the restriction of a continuous function from M to N? (Hint: See the previous exercise.)

Exercise 4. Let (M, d) be a metric space, and let K be a subset of M. Show that K is a compact subset of M if and only if the metric space (K, d') is a compact metric space (where d', as usual, is the distance function d restricted to K).

Exercise 5. Finish the proof of Theorem 5 by showing that any interval is a connected subset of $\mathbb{R}$. (Hint: Verify that every non-empty interval is an union of a sequence of nested bounded closed intervals, and apply Theorem 4.)

Exercise 6. Use Theorem 3 to prove the Intermediate Value Theorem: If f is a continuous function, $f\colon [a, c] \to \mathbb{R}$, and if y lies between $f(a)$ and $f(b)$, then there is a $b \in [a, c]$ such that $f(c) = y$.

Exercise 7. Take it as given that every straight line in $\mathbb{R}^n$ is the image of a continuous function from $\mathbb{R}$ to $\mathbb{R}^n$. Prove that $\mathbb{R}^n$ is connected.

Exercise 8. Is the intersection of connected sets connected? This is true in $\mathbb{R}$ since an intersection of intervals is an interval, but is it true more generally? (Hint: You can take it as given that the function $f\colon \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = (\cos(t), \sin(t))$ is a continuous function.)