The concept of a connected set is very easy in $\mathbb{R}$, it is an interval. There are a couple of ways we will look at connected. One that would immediately make sense is that any two points in the set could be “connected” somehow. For example, in the plane, we could say that if the line segment from one point to the other lies in the set, then the points are connected. But this is too restrictive. A slightly more general notion is: A set $C \subset \mathbb{R}^n$ is **polygonally connected** if for each pair of points $p, q$ in $C$, there is a finite sequence of points $p = p_1, p_2, \ldots, p_m = q$ so that the polygon

$$\{p_{i-1} + t(p_i - p_{i-1}) : 0 \leq t \leq 1, \quad i = 1, \ldots, m\}$$

is a subset of $C$. More generally, a set $C$ is **arc wise connected** if, given any pair of points $p, q$ in $C$, there is a continuous function $f : [0, 1] \to C$ such that $f(0) = p$ and $f(1) = q$. To make sense of this last definition, we need to understand what it means for a function $f : [0, 1] \to C \subseteq S$ to be continuous.

**Definition:** A function $f$ defined on a subset $E$ of a metric space $(S_1, d_1)$ with values in a metric space $(S_2, d_2)$ is an assignment of a unique element from $S_2$ to each point of $E$. The function is **continuous** at a point $p \in E$ if, given any ball $B_{(S_2, d_2)}(f(p), \epsilon)$ of positive radius $\epsilon$ centered on $f(p) \in S_2$, there is a ball $B_{(S_1, d_1)}(p, \delta)$ centered on $p$ of some radius $\delta = \delta_\epsilon$, such that

$$y \in B_{(S_1, d_1)}(p, \delta) \cap E \implies f(p) \in B_{(S_2, d_2)}(f(p), \epsilon).$$

**Examples:** The Figure shows some examples of sets $C$. In figure (a), the set is the interior of a planar region with a curve boundary that does not belong to the set and a polygonal boundary piece that is part of the set. The set in (a) is both polygonally and arc wise connected. In (b), the region is enclosed by a curve which is “pinched off”, but the one piece of the boundary that does belong to the set is an arc of a circle. Hence, to go from a point in one side of the region to the other, one must travel on the arc. Thus,
this set is arc wise connected, but not polygonally connected. An easier example of a set in \( \mathbb{R}^2 \) which is arc wise connected but not polygonally connected is the graph of a circle. Finally, the set in (c) is a rectangular figure with rounded corners and two elliptical holes cut out of it. Such a region will be both polygonally and arc wise connected.

These definitions still do not capture what will be the essence of “connectedness”. The opposite notion will be that the set can be “separated” in some sense. On the line, given any set \( C \) and a point \( p \) in \( C \), we see that the omission of \( p \) will “separate” the set \( C \setminus \{p\} \) in the following sense: there are two opens sets, namely \((−\infty, p)\) and \((p, +\infty)\) so that \( C \setminus \{p\} \) is contained in the union of the two disjoint open sets. It is that property, given entirely in terms of open sets, that is carried up to arbitrary metric spaces.

**Definition:** Let \((S, d)\) be a metric space, and let \( E \) be a subset of \( S \). The set \( E \) is said to be **disconnected** if there exists two open set \( U_1, U_2 \) in \( S \) so that

\[
E \cap U_1 \neq \emptyset, \quad E \cap U_2 \neq \emptyset, \quad U_1 \cap U_2 = \emptyset \quad \text{and} \quad E \subset U_1 \cup U_2. \quad (4.1)
\]

The set \( E \) is called **connected** if it is not disconnected. In other words, a set \( E \) is connected if every pair of open sets both of which meet \( E \) (have a point from \( E \)) and that cover \( E \) must have at least one point in common.

**Examples:** We again look to \( \mathbb{R}^n \).
1. The interval \([0, 1]\) is connected. Suppose not, then there are disjoint open sets \(U_1\) and \(U_2\) so that \([0, 1] \subseteq U_1 \cup U_2\), \(U_1 \cap [0, 1] \neq \emptyset\), and \(U_2 \cap [0, 1] \neq \emptyset\). Without loss of generality, we may assume that \(1 \notin U_1\) (if it is, switch names of the sets). Let \(y = \sup \{t : t \in [0, 1] \cap U_1\}\). We claim that \(y\) cannot be in \(U_1\) or \(U_2\), because any ball (interval \((y - \delta, y + \delta)\)) centered at \(y\) must contain points of \(U_2\) (e.g. \(y\) itself by definition of sup and the fact that \(U_1\) is open), and points of \(U_1\) by definition of sup. This would contradict the fact that \(U_1\) and \(U_2\) are disjoint, so it is impossible to find such open sets.

2. Any (open, closed, mixed) \(n\)-cell \(E\) is connected. We first note that \(E\) is polygonally connected. Indeed, if \(x = (x_1, \ldots, x_n)\) and \(y = (y_1, \ldots, y_n)\), then the polygonal path is can be taken along the “diagonal”:

\[
\{x + t(y - x) : 0 \leq t \leq 1\} = \{ty + (1 - t)x : 0 \leq t \leq 1\}.
\]

Notice that the \(j\)th coordinate of \(ty + (1 - t)x\) satisfies

\[
\min\{x_j, y_j\} \leq (ty + (1 - t)x)_j \leq \max\{x_j, y_j\}.
\]

We shall prove that any polygonally connected set is connected.

3. \(\mathbb{R}^n\) is connected. Again this will follow if we know that polygonally connected implies connected.

4. The following set will be connected:

\[
\{(x, y, z) : x^2 + y^2 = 1, z = 0\} \cup \{(\cos(\pi/t), \sin(\pi/t), t) : t > 0\}.
\]

This is a spiral wrapping around the cylinder of radius 1 centered on the \(z\)-axis which wraps infinitely often as the \(xy\)-plan is approached. This set is not arc wise connected, because there is no way to go from the circle to the spiral along a curve. Yet any open set that contains even one point on the circle \(x^2 + y^2 = 1, z = 0\), will contain a ball centered on the point and such a ball, no matter how small the radius, will contain some points on the spiral.

**Theorem S4.1.** If the set \(E\) in the metric space \((S, d)\) is arc wise connected, then \(E\) is connected.
Proof. Suppose that $E$ is arc wise connected, but not connected. Since $E$ is disconnected, there are two open sets $U_1$ and $U_2$ such that (4.1) holds. Take any point $p \in E \cap U_1$ and any point $q \in E \cap U_2$. Since $E$ is arc wise connected, there exists a continuous function $f : [0, 1] \to E$ such that $f(0) = p$ and $f(1) = q$. Consider the two sets

$$A = f^{-1}(U_1) := \{ t \in [0, 1] : f(t) \in U_1 \} \quad \text{and} \quad B = f^{-1}(U_2) := \{ t \in [0, 1] : f(t) \in U_2 \}. \quad \text{(4.1)}$$

Now $0 \in A$ and $1 \in B$, so $A$ and $B$ are nonempty. Also, $A \cap B = \emptyset$, for if there were a common point $t_0$, then $f(t_0)$ would be in both $U_1$ and $U_2$, which is impossible. Finally, $A \cup B = [0, 1]$, because $f([0, 1]) \subset E \subseteq U_1 \cup U_2$. Now the following Proposition will tell us that $A$ and $B$ are open sets relative to $[0, 1]$. Therefore, $A$ and $B$ disconnect $[0, 1]$, which contradicts the connectedness of $[0, 1]$ shown in Example 1 above. \hfill \square

**Proposition S4.2.** For a continuous function from a subset $E$ of the metric space $(S_1, d_1)$ into a metric space $(S_2, d_2)$, the inverse image of an open set $U \subset S_2$ is a (relatively) open set in $E$. That is, the set $f^{-1}(U_1)$ is open in $(E, d_1)$.

**Proof.** Let $p \in f^{-1}(U_1)$. Then $f(p) \in U_1$, so there is a ball $B_{(S_2, d_2)}(f(p), \epsilon) \subset U_1$. From the definition of $f$ being continuous at $p$, there is a ball $B_{(S_1, d_1)}(p, \delta)$ so that

$$f \left( B_{(S_1, d_1)}(p, \delta) \cap E \right) \subset B_{(S_2, d_2)}(f(p), \epsilon).$$

Hence, $B_{(S_1, d_1)}(p, \delta) \cap E \subseteq f^{-1}(U_1)$. That means $f^{-1}(U_1)$ is open relative to $E$. \hfill \square