The familiar Intermediate Value Theorem of elementary calculus says that if a real valued function f is continuous on the interval $[a, b] \subseteq \mathbb{R}$ then it takes each value between $f(a)$ and $f(b)$. As our next result shows, the critical fact is that the domain of f, the interval $[a, b]$, is a connected space, for the theorem generalizes to real-valued functions on any connected space.

The Intermediate Value Theorem. *Suppose that $f\colon X \to \mathbb{R}$ is continuous, where X is a nonempty connected space, and let $a, b \in X$. If $y \in \mathbb{R}$ and $f(a) \le y \le f(b)$ then there is an $x \in X$ such that $f(x) = y$.*

Proof. Since X is connected and f is continuous it follows that $f(X)$ is connected (by a result we proved last time). Note that $f(X) \subseteq \mathbb{R}$, since f is a function from X to $\mathbb{R}$, and $f(X)$ is nonempty, since $f(a)$ and $f(b)$ are in $f(X)$. It follows that $f(X)$ is an interval, since (as we proved last time) all nonempty connected subsets of $\mathbb{R}$ are intervals. As we pointed out last time, intervals are characterized by the property that any point lying between two points of an interval is also in the interval. Thus since $f(a) \le y \le f(b)$ and $f(a)$ and $f(b)$ are points of the interval $f(X)$ we conclude that $y \in f(X)$ also; that is, $y = f(x)$ for some $x \in X$, as claimed. □

It is intuitively reasonable that if connected sets overlap then their union ought to be connected. It is quite straightforward to prove this rigorously. A result along these lines was given in Tutorial 11, Exercise 4. For variety, we present here a formulation of the proof that is a little different from the one that appears in the tutorial solutions. We have also slightly strengthened the statement of the result.

Proposition. *Suppose that $(A_i)_{i \in I}$ is a family of connected subsets of a topological space X, and suppose that for all i, $j \in I$ there is a finite sequence $i = i_0, i_1, \ldots, i_n = j$ in I such that $A_{i_{k-1}} \cap A_{i_k} \neq \emptyset$ for all $k \in \{1, 2, \ldots, n\}$. Then the set $\bigcup_{i \in I} A_i$ is connected.*

Proof. Put $A = \bigcup_{i \in I} A_i$, and let $f\colon A \to \{0, 1\}$ be a continuous function (the topology on $\{0, 1\}$ being the discrete topology). Let $i \in I$ be arbitrary. Since A_i is connected and the restriction of f to A_i is a continuous function $A_i \to \{0, 1\}$, this restriction cannot be surjective (for otherwise $A_i \cap f^{-1}(\{0\})$ and $A_i \cap f^{-1}(\{1\})$ would be disjoint nonempty open subsets of A_i whose union is A_i). So there is a c_i, which is either 0 or 1, such that $f(a) = c_i$ for all $a \in A_i$.

Now let i, $j \in I$ be arbitrary. The hypotheses ensure that there exists a finite sequence $i = i_0, i_1, \ldots, i_n = j$ in I such that $A_{i_{k-1}} \cap A_{i_k} \neq \emptyset$ for all $k \in \{1, 2, \ldots, n\}$, and choosing $a \in A_{i_{k-1}} \cap A_{i_k}$ we see that $f(a) = c_{i_{k-1}}$ (since $a \in A_{i_{k-1}}$) and $f(a) = c_{i_k}$ (since $a \in A_{i_k}$). So $c_{i_{k-1}} = c_{i_k}$, and since this holds for all k from 1 to n we have that

$$c_{i_0} = c_{i_1} = \cdots = c_{i_{n-1}} = c_{i_n},$$

whence $c_i = c_j$. Since i and j were arbitrary, we have shown that all the c_i's have the same value; that is, $c_i = c$ for all i, where $c \in \{0, 1\}$ is fixed. If now $a \in A$ is arbitrary, then $a \in A_i$ for some $i \in I$ (since $A = \bigcup_{i \in I} A_i$), and so $f(a) = c_i = c$. So $f(A) = \{c\}$ (either $\{0\}$ or $\{1\}$), and therefore f is not surjective. So there is no surjective continuous function $A \to \{0, 1\}$, and so A is connected. □

Intuitively, points in the closure of a set A should be thought of as having points of A arbitrarily close to them. But are they connected to A? More precisely, if A is connected, is it necessarily true that $\overline{A}$ is connected? We shall prove that the answer to this is yes, for the technical meaning that we have given the word "connected". But whether or not this accords with our (vague) intuitive concept of connectedness is somewhat unclear. It is not true that if A is path-connected (see below for the definition of this) then $\overline{A}$ is necessarily path-connected.

Proposition. *Suppose that A is a connected subset of a topological space X. Then every set B such that $A \subseteq B \subseteq \overline{A}$ is also connected.*

Proof. Suppose that $A \subseteq B \subseteq \overline{A}$, and suppose that there exist open sets U_1 and U_2 with $B \subseteq U_1 \cup U_2$ and $B \cap U_1 \cap U_2 = \emptyset$. We shall prove that either $B \cap U_1 = \emptyset$ or else $B \cap U_2 = \emptyset$; this will show that B is connected, for if it were disconnected there would exist such sets U_1 and U_2 with $B \cap U_1$ and $B \cap U_2$ both nonempty.

Since $A \subseteq B \subseteq U_1 \cup U_2$ and $A \cap U_1 \cap U_2 \subseteq B \cap U_1 \cap U_2 = \emptyset$ we see that $A \subseteq U_1 \cup U_2$ and $A \cap U_1 \cap U_2 = \emptyset$. Since A is connected it follows that either $A \cap U_1 = \emptyset$ or $A \cap U_2 = \emptyset$. Suppose first that $A \cap U_1 = \emptyset$. Then $A \subseteq (X \setminus U_1)$; but $X \setminus U_1$ is closed, and, by the definition of the closure of a set (see Lecture 5), $\overline{A}$ is contained in all closed sets containing A. So $\overline{A} \subseteq (X \setminus U_1)$; that is, $\overline{A} \cap U_1 = \emptyset$. Since $B \subseteq \overline{A}$ it follows that $B \cap U_1 = \emptyset$ also. In the alternative case, $A \cap U_2 = \emptyset$, a completely analogous argument yields $B \cap U_2 = \emptyset$; hence either $B \cap U_1 = \emptyset$ or $B \cap U_2 = \emptyset$, as required. □

Corollary. *If A and B are connected subsets of a topological space and A contains a point of $\overline{B}$ then $A \cup B$ is connected.*

Proof. Suppose that $x \in \overline{B} \cap A$. Since $B \subseteq \{x\} \cup B \subseteq \overline{B}$, and B is connected, it follows that $\{x\} \cup B$ is connected. Furthermore, since $A = A \cup \{x\}$, we see that $A \cup B = (A \cup \{x\}) \cup B = A \cup (\{x\} \cup B)$. So $A \cup B$ is the union of the sets A and $\{x\} \cup B$, which are both connected, and which have nonempty intersection (since both sets contain x). So $A \cup B$ is connected. □

Path-connected spaces

Let X be a topological space. A *path* in X is a continuous function $\gamma\colon [0,1] \to X$. We say that the path goes from the point $\gamma(0)$ to the point $\gamma(1)$; we can intuitively think of $\gamma(t)$ as the position, at time t, of a particle that is moving around in the space X.

The image of a path $\gamma\colon [0,1] \to X$ is called a *curve* in X. That is, the curve determined by γ is the set $\gamma([0,1]) \subseteq X$. Whereas a path is a function, a curve is a set. Note that since we require paths to be continuous functions, and since the continuous image of a connected set is necessarily connected, it follows that a curve in a topological space X is always a connected subset of X.

Definition. A space X is said to be *path-connected* if for every a, $b \in X$ there is a path from a to b.

That is, X is path connected if for every a and b in X there is a continuous function γ from $[0,1]$ to X with $\gamma(0) = a$ and $\gamma(1) = b$. Intuitively, this means that a particle can move continuously from any point in X to any other point in X without leaving X.

Theorem. *Every path-connected space is connected.*

Proof. Suppose, for a contradiction, that X is path-connected but not connected. Since it is not connected, there exists a continuous surjective map f from X to the discrete two-element space $\{0,1\}$. Now let a, $b \in X$ with $f(a) = 0$ and $f(b) = 1$. Since X is path-connected, there exists a path $\gamma\colon [0,1] \to X$ with $\gamma(0) = a$ and $\gamma(1) = b$. Since composites of continuous functions are continuous, it follows that $f \circ \gamma\colon [0,1] \to \{0,1\}$ is a continuous function, and it is also surjective since $(f \circ \gamma)(0) = f(\gamma(0)) = f(a) = 0$ and

$(f \circ \gamma)(1) = f(\gamma(1)) = f(b) = 1$. But this contradicts the fact that the interval $[0,1]$ is connected. $\square$

We shall show that path-connectedness is a stronger condition than connectedness. For this purpose it is sufficient to give an example of a connected space that is not path-connected. In $\mathbb{R}^2$, let $A = \{ (x, \sin(1/x)) \mid 0 < x \leq (1/\pi) \}$, and let $B = A \cup \{(0,0)\}$. Since the interval $(0, 1/\pi)$ is a connected set, and since $x \to (x, \sin(1/x))$ is a continuous function that maps this interval onto the set A, it follows that A is connected (since continuous images of connected sets are connected). If we can show that $(0,0)$ is in the closure of A then it will follow that $A \subseteq B \subseteq \overline{A}$, and hence that B is connected. To show that $(0,0) \in \overline{A}$ it suffices to find a sequence $(a_n)_{n=1}^{\infty}$ of points in A such that $\lim_{n\to\infty} a_n = (0,0)$. But if $x = 1/(n\pi)$, where $n \in \mathbb{Z}^+$, then $\sin(1/x) = \sin(n\pi) = 0$, and so the point $a_n = (1/(n\pi), 0)$ is in the set A; moreover, it is clear that $a_n \to (0,0)$ as $n \to \infty$, as required.

We have shown that the set B is connected. It turns out that it is not path connected. The proof presented here was not done in the lecture (though it may be done as an example in Lecture 25 or 26.

It is understood (and was implicitly assumed above) that we are dealing with the usual topologies for subsets of $\mathbb{R}$ and $\mathbb{R}^2$. We shall use d to denote the Euclidean metrics on both $\mathbb{R}$ and $\mathbb{R}^2$; it will always be clear from the context which one is meant.

Suppose, for a contradiction, that a continuous function $\gamma\colon [0,1] \to B$ exists satisfying $\gamma(0) = (0,0)$ and $\gamma(1) = ((1/\pi), 0)$. Let $T = \gamma^{-1}(\{(0,0)\}) = \{ t \in [0,1] \mid \gamma(t) = (0,0) \}$. Note that $0 \in T$, since $\gamma(0) = (0,0)$ by hypothesis. The set $\{(0,0)\}$ is a closed subset of B, since it is true in every metric space that single-element subsets are closed. Since γ is continuous, and continuous preimages of closed sets are closed, it follows that T is a closed subset of $[0,1]$. Since $[0,1]$ itself is a closed subset of $\mathbb{R}$ it follows that T is a closed subset of $\mathbb{R}$.† We have already observed that T is nonempty, and it is bounded above, since 1 is an upper bound for any subset of $[0,1]$. So T has a supremum. Since T is closed we know that $\sup T \in T$. (See the solutions to Tutorial 11 (Question 2) for a proof of this.)

Define $t_0 = \sup T$. Since $t_0 \in T$ we know that $\gamma(t_0) = (0,0)$. Since $\gamma(1) = (1/\pi, 0)$ (by hypothesis), we know that $t_0 \neq 1$, and so $0 \leq t_0 < 1$. Furthermore, if $t_0 < t \leq 1$ then $t \notin T$, since t_0 is an upper bound for T, and so $\gamma(t) \neq (0,0)$. So for all $t \in (t_0, 1]$ we have $\gamma(t) \in B \setminus \{(0,0)\} = \{ (x, \sin(1/x)) \mid 0 < x \leq (1/\pi) \}$.

For all $t \in [0,1]$, write $\gamma(t) = (X(t), Y(t))$, so that X and Y are real valued functions on the interval $[0,1]$. The projection map $\eta\colon \mathbb{R}^2 \to \mathbb{R}$ given by $(x,y) \mapsto x$ (for all x and y) is continuous, and since the function X is the composite $\eta_1 \circ \gamma$, and γ is also continuous, it follows that X is continuous. Since $\gamma(t_0) = (0,0)$ we have $X(t_0) = 0$ and $Y(t_0) = 0$, and since $\gamma(t) \in \{ (x, \sin(1/x)) \mid 0 < x \leq (1/\pi) \}$ for all $t \in (t_0, 1]$ we see that $X(t) \in (0, 1/\pi]$ and $Y(t) = \sin(1/X(t))$ for all $t \in (t_0, 1]$.

Since γ is continuous at t_0, there exists $\delta > 0$ such that $d(\gamma(t), \gamma(t_0)) < 1$ for all $t \in [0,1]$ with $d(t, t_0) < \delta$. Writing $t_1 = \min(t_0 + \delta, 1)$, we have that $X(t_0) = 0$ and $X(t_1) > 0$. By the Intermediate Value Theorem, for every real number x satisfying $X(t_0) < x < X(t_1)$ there is a $t \in (t_0, t_1)$ such that $X(t) = x$. We can choose an integer n

† The topology on $[0,1]$ is its topology as a subspace of $\mathbb{R}$, and so closed subsets of $[0,1]$ have the form $[0,1] \cap F$, where F is closed in $\mathbb{R}$. But the intersection of two closed sets is closed; so any such set is also closed in $\mathbb{R}$.

such that $0 < 2/((4n+1)\pi) < X(t_1)$ (since $2/((4n+1)\pi) \to 0$ as $n \to \infty$), and conclude that there is a $t \in (t_0, t_1)$ with $X(t) = 2/((4n+1)\pi)$. Moreover, this gives

$$Y(t) = \sin(1/X(t)) = \sin(2n\pi + \tfrac{\pi}{2}) = 1,$$

and hence $\gamma(t) = (x, 1)$, where $x = 2/((4n+1)\pi)$. Observe that

$$d(\gamma(t), \gamma(t_0)) = d((0,0),(x,1)) = \sqrt{x^2+1} > 1.$$

But since $t \in (t_0, t_1)$ we have $d(t, t_0) < d(t_1, t_0) \le \delta$; so, by the way δ was chosen, $d(\gamma(t), \gamma(t_0)) < 1$. So we have obtained a contradiction, as desired.

More on compactness

Proposition. *Let X be a topological space, and C a compact subset of X. If A is an infinite subset of C then A has at least one point of accumulation in C.*

Proof. Suppose, for a contradiction, that no point of C is an accumulation point of A. So, if c is any point of C, it is not true that every open neighbourhood of c contains a point of $A \setminus \{c\}$. It follows that for each $c \in C$ we may choose an open set U_c such that $c \in U_c$ and $U_c \cap A \setminus \{c\} = \emptyset$. That is, $U_c \cap A \subseteq \{c\}$.

If $U = \bigcup_{c \in C} U_c$ then for each $c \in C$ we have $c \in U_c \subseteq U$. Thus the family of open sets $(U_c)_{c \in C}$ forms a covering of C, and because C is compact it follows that there is a finite subcovering. So there exists a finite subset $P = \{c_1, c_2, \dots, c_n\}$ of C such that $C \subseteq \bigcup_{c \in P} U_c$. Now $A \subseteq C$, and hence

$$A = A \cap C \subseteq A \cap \bigcup_{c \in P} U_c = \bigcup_{c \in P} (A \cap U_c) \subseteq \bigcup_{c \in P} \{c\} = P.$$

But A was assumed to be infinite, whereas P is finite; so we have obtained a contradiction. □

Corollary. *Let C be a compact subset of a metric space X. Every infinite sequence $(a_n)_{n=1}^{\infty}$ in C has a subsequence that converges to a point of C.*

Proof. Let $A = \{\, a_n \mid n \in \mathbb{Z}^+ \,\}$, the set of points of C that occur as terms of the sequence. The sequence is infinite, but it is possible that the set A is finite; if so, there must be at least one $a \in A$ such that $a_n = a$ for infinitely many values of n. That is, in this case there exists an infinite sequence of positive integers $n_1 < n_2 < \cdots$ such that $a_{n_i} = a$ for all i. Clearly, the subsequence $(a_{n_i})_{i=1}^{\infty}$ of (a_n) is then convergent, its limit being a (which is an element of C).

We are left with the case that the set A is infinite. The proposition then guarantees the existence of a point $c \in C$ that is an accumulation point of A. Every open neighbourhood of c then contains a point of A. We can thus choose an infinite increasing sequence of positive integers as follows: let $n_1 = 1$, and for each $i > 1$ choose n_i so that a_{n_i} is in the open ball with centre x and radius $\frac{1}{2}d(x, a_{n_{i-1}})$. That is, a_{n_i} is any point of A that lies in this ball. A straightforward induction shows that $d(x, a_{n_i}) < 2^{-(i-1)}d(x, a_1)$ for all i, and hence $d(x, a_{n_i}) \to 0$ as $i \to \infty$. That is, the subsequence $(a_{n_i})_{i=1}^{\infty}$ of (a_n) converges to the point $x \in C$. □

A subset S of a metric space X is said to be *sequentially compact* if every infinite sequence in S has a subsequence converging to a point of S. We have shown above that compact implies sequentially compact; in the next lecture we shall prove the converse. Thus in metric spaces compact and sequentially compact are equivalent conditions.