Part IA Metric and Topological Spaces, first set of notes

Burt Totaro

1 May 2004

1 Week 1: Metric spaces

1.1 Day 1: Thursday 23 April 2004

Recall the definition of a continuous function on the real line.

Definition 1.1 *A function $f : \mathbf{R} \to \mathbf{R}$ is continuous at $a \in \mathbf{R}$ if, for every $\epsilon > 0$, there is a $\delta > 0$ such that for all $x \in \mathbf{R}$ with $|x - a| < \delta$, we have $|f(x) - f(a)| < \epsilon$.*

We can think of this definition as being based on the notion of the *distance* between two real numbers, defined as $d(x, y) = |x - y|$. Informally, two real numbers are "close" to each other if the distance between them is small. The definition of continuity is a precise formulation of the idea that f of a point "close" to a is a point "close" to $f(a)$.

Now there are other geometric situations (not only the real line) where there is a natural notion of distance between two points. As a result, we will be able to talk about continuity in these more general situations.

For example, Euclidean geometry is based on the notion of distance between two points in the plane $\mathbf{R}^2 = \{(x_1, x_2) : x_1 \in \mathbf{R}, x_2 \in \mathbf{R}\}$. Namely, using the Pythagorean theorem, one can check that the distance from a point $x = (x_1, x_2)$ in the plane to $y = (y_1, y_2)$ is:

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}.$$

More generally, we can talk about the distance between two points in space $\mathbf{R}^3$, or more generally in n-dimensional space $\mathbf{R}^n$. (For a positive integer n, we define $\mathbf{R}^n$ to be the product of n copies of the set $\mathbf{R}$ of real numbers; that is, a point in $\mathbf{R}^n$ is a sequence $(x_1, \ldots, x_n)$ of real numbers.) Geometrically, the distance from one point to another in $\mathbf{R}^n$ is the length of the line segment between the two points; algebraically, we can define this distance as follows.

Definition 1.2 *The distance between two points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbf{R}^n$ is defined to be:*

$$d(x, y) = [\sum_{i=1}^{n} (y_i - x_i)^2]^{1/2}.$$

This definition involves the square root of a nonnegative real number: we mean the nonnegative square root.

Using this notion of distance on $\mathbf{R}^n$, we can define what it means for a function $f : \mathbf{R}^n \to \mathbf{R}$ to be continuous.

Definition 1.3 *A function $f : \mathbf{R}^n \to \mathbf{R}$ is continuous at $a \in \mathbf{R}^n$ if, for every $\epsilon > 0$, there is a $\delta > 0$ such that for all $x \in \mathbf{R}^n$ with $d(x, a) < \delta$, we have $d(f(x), f(a)) < \epsilon$.*

Notice that $d(f(x), f(a))$ in this definition just means the distance between two real numbers, that is, $|f(x) - f(a)|$.

We now make a vast generalization. Rather than only talking about the distance between two points of $\mathbf{R}^n$, we try to imagine the distance between two points of any set. For this idea to be of any use, the distance must satisfy several properties which we now list.

Definition 1.4 *(M. Fréchet, 1906) A metric space X (or (X, d)) is a set X together with a function*

$$d : X \times X \to \mathbf{R}$$

which satisfies the following properties.

(0) $d(x, y) \geq 0$ for all $x, y \in X$,

(1) $d(x, y) = 0$ if and only if $x = y$,

(2) $d(x, y) = d(y, x)$ for all $x, y \in X$,

(3) (The triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$, for all $x, y, z \in X$.

The most important point about metric spaces is that it makes sense to ask whether a function from one metric space to another is continuous.

Definition 1.5 *Let X and Y be metric spaces. A function $f : X \to Y$ is continuous at $a \in X$ if, for every $\epsilon > 0$, there is a $\delta > 0$ such that for all $x \in X$ with $d(x, a) < \delta$, we have $d(f(x), f(a)) < \epsilon$. We say that $f : X \to Y$ is continuous if it is continuous at every point of X.*

Examples of metric spaces:

(1) Euclidean space $\mathbf{R}^n$ (with the real line $\mathbf{R}$ as a special case). We defined the distance between two points of $\mathbf{R}^n$ in Definition 1.2. To check that $\mathbf{R}^n$ is a metric space, we have to check the properties (0) to (3). Here you can easily check properties (0), (1), (2). Let us check property (3), the triangle inequality. This is geometrically obvious, but since we gave an algebraic definition of distance on $\mathbf{R}^n$, it seems proper to give an algebraic proof of the triangle inequality.

Given 3 points x, y, z in $\mathbf{R}^n$, we need to show that

$$[\sum_{i=1}^{n}(z_i - x_i)^2]^{1/2} \leq [\sum_{i=1}^{n}(y_i - x_i)^2]^{1/2} + [\sum_{i=1}^{n}(z_i - y_i)^2]^{1/2}.$$

To simplify these formulas, define real numbers $a_i = y_i - x_i$ and $b_i = z_i - x_i$, for $i = 1, \ldots, n$. Then $z_i - x_i = a_i + b_i$, and so we need to show:

$$[\sum(a_i + b_i)^2]^{1/2} \leq [\sum a_i^2]^{1/2} + [\sum b_i^2]^{1/2}.$$

(Here and in what follows, all the sums run from $i = 1$ to n.) Since both sides of this inequality are nonnegative, it suffices to prove the inequality obtained by squaring both sides. That is, we want to show:

$$\sum(a_i + b_i)^2 \leq \sum a_i^2 + 2[\sum a_i^2]^{1/2}[\sum b_i^2]^{1/2} + \sum b_i^2.$$

We can expand the left side of this inequality as

$$\sum a_i^2 + 2\sum a_i b_i + \sum b_i^2.$$

So the inequality we want follows if we can prove:

$$\sum a_i b_i \leq [\sum a_i^2]^{1/2}[\sum b_i^2]^{1/2}.$$

This is called the Cauchy-Schwarz inequality.

To prove the Cauchy-Schwarz inequality, first note that it is clearly true if $a_1 = \cdots = a_n = 0$, or if $b_1 = \cdots = b_n = 0$. So we can assume that at least one a_i is nonzero and at least one b_i is nonzero. Therefore, $[\sum a_i^2]^{1/2}$ and $[\sum b_i^2]^{1/2}$ are positive real numbers (not zero). Notice that if we multiply all the numbers $a_1, \ldots, a_n$ by the same positive number c, then both sides of the inequality are multiplied by c. Choose c in such a way that (after multiplying) $[\sum a_i^2]^{1/2}$ is equal to 1; then we have shown that it suffices to prove the Cauchy-Schwarz inequality in the special case where $\sum a_i^2 = 1$. Likewise, we can scale the numbers $b_1, \ldots, b_n$ by any positive constant. The result is that it suffices to prove the following special case of the Cauchy-Schwarz inequality: if $\sum a_i^2 = 1$ and $\sum b_i^2 = 1$, then

$$\sum a_i b_i \leq 1.$$

To prove that, we argue as follows. For any real numbers a and b, we have

$$(a - b)^2 \geq 0,$$

since the square of any real number is nonnegative. Expanding this expression out and dividing by 2, we find that

$$ab \leq \frac{a^2}{2} + \frac{b^2}{2}$$

for all real numbers a and b. Applying this to the numbers $a = a_i$ and $b = b_i$, for each $i = 1, \ldots n$, we find that

$$\begin{aligned}\sum a_i b_i &\leq \frac{1}{2} \sum a_i^2 + \frac{1}{2} \sum b_i^2 \\ &= \frac{1}{2} + \frac{1}{2} \\ &= 1,\end{aligned}$$

as we want. Thus we have proved the Cauchy-Schwarz inequality, and hence that $\mathbf{R}^n$ is a metric space.

1.2 Day 2: Saturday 25 April

(2) Discrete metric spaces. For any set X, we define a metric on X by: the distance from a point to itself is 0, and the distance between any two distinct points is 1. You can check that this is a metric space (that is, that properties (0) to (3) are satisfied).

(3) The l^1, l^2, and l^∞ metrics on $\mathbf{R}^2$. We define three different metrics on $\mathbf{R}^2$ by:

$$\begin{aligned}d_1(x, y) &= |y_1 - x_1| + |y_2 - x_2| \\ d_2(x, y) &= [|y_1 - x_1|^2 + |y_2 - x_2|^2]^{1/2} \\ d_\infty(x, y) &= \max(|y_1 - x_1|, |y_2 - x_2|).\end{aligned}$$

Here d_2 is the standard metric. You can check that d_1 and d_∞ are metrics; it is rather easier than the case of the standard metric d_2 (proved in (1), above). For example, let us check the triangle inequality for d_∞. We have to show that for x, y, z in $\mathbf{R}^2$,

$$\max(|z_1 - x_1|, |z_2 - x_2|) \leq \max(|y_1 - x_1|, |y_2 - x_2|) + \max(|z_1 - y_1|, |z_2 - y_2|).$$

To prove this, it suffices to show that both $|z_1 - x_1|$ and $|z_2 - x_2|$ are $\leq$ the right side of the inequality; the proof will be the same in both cases, so let us consider $|z_1 - x_1|$. We have:

$$\begin{aligned}|z_1 - x_1| &\leq |y_1 - x_1| + |z_1 - y_1| \\ &\leq \max(|y_1 - x_1|, |y_2 - x_2|) + \max(|z_1 - y_1|, |z_2 - y_2|),\end{aligned}$$

as we want.

We can define the l^1, l^2, and l^∞ metrics on $\mathbf{R}^n$ for any n, by the analogous formulas.

(4) The standard metric on $\mathbf{C}$. Recall the definition of the absolute value of a complex number: for $z = x + iy$, with x and y real numbers, we define

$$|z| = \sqrt{x^2 + y^2}.$$

This is a nonnegative real number. Using this, we define the standard metric on the complex numbers $\mathbf{C}$ by:

$$d(z_1, z_2) = |z_2 - z_1|.$$

You can check that this is indeed a metric; in particular, the triangle inequality follows from the property $|z + w| \leq |z| + |w|$ of the absolute value for complex numbers. This metric on $\mathbf{C}$ is really the same thing as the standard metric on $\mathbf{R}^2$, if you identify a complex number $x + iy$ with the point (x, y) in $\mathbf{R}^2$.

(5) Subspace metrics. Let X be a metric space, and let A be any subset of X. Then we define a metric on A by defining

$$d_A(x, y) = d_X(x, y)$$

for any $x, y \in A$. You can check that this makes A into a metric space.

As a result, there is a colossal supply of examples of metric spaces: any subset of $\mathbf{R}$ or $\mathbf{R}^2$ or $\mathbf{R}^n$ is a metric space. Thus, any kind of geometrical shape – a curve, a surface, a region, or a stranger shape – can be viewed as a metric space.

Definition 1.6 *Let X be a metric space. For any point a in X and any positive real number r, define the open ball $B_r(a)$ to be the set of points x in X with $d(x, a) < r$.*

Examples. In the metric space $\mathbf{R}$, the open ball $B_r(a)$ is the open interval $(a - r, a + r)$. In $\mathbf{R}^2$, $B_r(a)$ is the region strictly inside a circle of radius r centered at a (not including the circle itself). In $\mathbf{R}^3$, likewise, $B_r(a)$ is the region strictly inside a sphere of radius r centered at a. (Here, when we talk about $\mathbf{R}$ or $\mathbf{R}^n$ as a metric space, we always have in mind the standard metric unless otherwise stated.)

We can also consider the open ball $B_r(a)$ with respect to different metrics on $\mathbf{R}^2$: you should draw what it looks like for the l^1, l^2, and l^∞ metrics on $\mathbf{R}^2$, defined at the start of this lecture.

In a discrete metric space X, as defined at the start of this lecture, the open ball $B_r(a)$ is just the point a for $r \leq 1$, whereas it is the whole space X for $r > 1$.

In a subspace S of a metric space X, the open ball $B_r(a)$ in the metric space S is the intersection of $B_r(a)$ in X with S. For example, in the metric space $[0, 1]$ (defined as a subset of the metric space $\mathbf{R}$), the open ball $B_{1/2}(0)$ is $[0, 1/2) \subset [0, 1]$.

Using the notion of open balls, we can restate the definition of continuity in a more geometric way, without explicitly mentioning ϵ and δ.

Lemma 1.7 *Let X and Y be metric spaces, and let $f : X \to Y$ be a function. Then f is continuous at a point a in X if and only if, for every open ball C around $f(a)$, there is an open ball B around a such that $f(B) \subset C$.*

Here $f(B)$ means the subset $\{f(x) : x \in B\}$ of Y. The lemma is easy to prove: just write $C = B_\epsilon(f(a))$ and $B = B_\delta(a)$. Then the lemma is simply a restatement of the definition of continuity (from day 1).

Definition 1.8 *A subset U of a metric space X is open if, for every point x in U, there is an open ball around x (in X) which is contained in U.*

You can convince yourself, for example, that in the real line, the open interval $(0, 1)$ is open, whereas $(0, 1]$ and $[0, 1]$ are not open in $\mathbf{R}$. In $\mathbf{R}^2$, the open ball $\{(x, y) : x^2 + y^2 < 1\}$ is open, for example, but $\{(x, y) : x^2 + y^2 \leq 1\}$ is not open. Also, you can check that every subset of a discrete metric space is open.

1.3 Day 3: Thursday 28 April

Lemma 1.9 *Every open ball in a metric space is an open set.*

Proof. We have to show that for every point x in a metric space X and every $r > 0$, the open ball $B_r(x)$ is an open subset of X. That is, we have to show that for every point y in $B_r(x)$, there is an open ball around y which is contained in $B_r(x)$. Since y is in $B_r(x)$, we have $d(x, y) < r$. Let $a = r - d(x, y)$, which is a positive real number; we will show that the open ball $B_a(y)$ is contained in $B_r(x)$. (To see why this is the right choice of a, draw a picture.)

Thus, we have to show that for every point z in $B_a(y)$, z is also in $B_r(x)$. Equivalently, we have to show that for any point z in X with $d(y, z) < a$, we also have $d(x, z) < r$. This follows from the triangle inequality:

$$\begin{aligned} d(x,z) &\leq d(x,y) + d(y,z) \\ &= (r-a) + d(y,z) \\ &< (r-a) + a = r, \end{aligned}$$

as we want. QED

Thus, we have some examples of open sets in any metric space, namely all open balls. Some other examples: in any metric space X, the empty set $\emptyset$ and the whole space X are open in X, as you can check from the definition of open subsets.

When considering subspaces of metric spaces, it is often helpful to say "open in X", not just "open", to avoid ambiguity. For example, the open interval $(0, 1) \times \{0\}$ is not open in the plane $\mathbf{R}^2$, but it is open in the subspace $\mathbf{R} \times \{0\}$ of $\mathbf{R}^2$. (Draw a picture.)

The following theorem, the main result of the first week, gives yet another characterization of continuous functions, using open sets instead of open balls. This is the most abstract way to describe continuity: the epsilons and deltas are hidden in the definition of open sets. But this makes it also the most convenient definition of continuity for many purposes.

Theorem 1.10 *Let X and Y be metric spaces. A function $f : X \to Y$ is continuous if and only if, for every open subset U of Y, the inverse image $f^{-1}(U)$ is open in X.*

Here the inverse image $f^{-1}(U)$ means the set of points x in X such that $f(x)$ is in U. Notice that this makes sense even though there may not be an inverse *function* f^{-1} from Y to X. (The inverse image of a point in Y may be several points in X, or no points; but we can still talk about the *set* $f^{-1}(U)$ in X, for every subset U of Y.)

Proof. First suppose that $f : X \to Y$ is continuous. We want to show that for every open subset U of Y, the inverse image $f^{-1}(U)$ is open in X. By definition of open subsets in a metric space, we have to show that for every point x in $f^{-1}(U)$, there is an open ball around x which is contained in $f^{-1}(U)$. To say that x is in $f^{-1}(U)$ means exactly that $f(x)$ is in U. We then use that U is open in Y; it follows that there is an open ball C around $f(x)$ in Y which is contained in U. Since f is continuous, Lemma 1.7 shows that there is an open ball B around x such that $f(B)$ is contained in C. Therefore, $f(B)$ is also contained in U. Equivalently, B is contained in the inverse image $f^{-1}(U)$. Thus we have shown that $f^{-1}(U)$ is open, as we want.

Conversely, let $f : X \to Y$ be a function such that the inverse image of every open set is open. We want to show that f is continuous. By Lemma 1.7, it suffices to show that for every point x in X and every open ball C around $f(x)$ in Y, there is an open ball B around x such that $f(B)$ is contained in C. To prove this, we use Lemma 1.9, which tells us that the open ball C is an open subset of Y. Therefore, by our assumption on f, the inverse image $f^{-1}(C)$ is open in X. Since $f(x)$ belongs to the ball C, the point x is in $f^{-1}(C)$. Since $f^{-1}(C)$ is open, there is a ball B around x in X which is contained in $f^{-1}(C)$. Equivalently, $f(B)$ is contained in C, which is the conclusion we want. That is, f is continuous. QED

Beware that continuity does not imply that the image (as opposed to the inverse image) of an open subset is open. You see this in simple examples. For example, take the continuous map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = 0$ for all x. Then the image $f((-1, 1))$ of the open subset $(-1, 1)$ in $\mathbf{R}$ is just the point $\{0\}$, which is not open in $\mathbf{R}$. Or, for a slightly more complex example, take the continuous map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^2$. Then the image of the open subset $(-1, 1)$ is the interval $[0, 1)$, which again is not open in $\mathbf{R}$.

To get an idea of how Theorem 1.10 can be useful, consider the following definition.

Definition 1.11 *Let X be a set with two metrics d_1 and d_2. We say that d_1 and d_2 are topologically equivalent if the d_1-open subsets of X are the same as the d_2-open subsets of X.*

By Theorem 1.10, whether a function $f : X \to Y$ is continuous does not change if you replace the given metric on X by a topologically equivalent metric. Likewise, it does not change if you replace the metric on Y by a topologically equivalent metric. This can be very convenient, as the following example suggests.

Example. The l^1, l^2, and l^∞ metrics on $\mathbf{R}^2$ may seem quite different. In particular, the open balls for these three metrics are all different. (For the l^1 and l^∞ metrics, open balls are (the interiors of) squares, tilted by 45 degrees in the l^1 case.) But the open *subsets* of $\mathbf{R}^2$ are the same for all three of these metrics; in other words, these three metrics are all topologically equivalent, as you can check. (The key point is that each open ball for one of the metrics contains an open ball for each of the other metrics; possibly of smaller radius, but that's good enough.)

As a result, suppose you want to check whether a given function on the plane $\mathbf{R}^2$ is continuous. (Talking about $\mathbf{R}^2$ without specifying a metric, we always mean the standard metric unless stated otherwise.) Then Theorem 1.10 shows that it is equivalent to check continuity of the given function with respect to the l^∞ metric on $\mathbf{R}^2$, since it is topologically equivalent to the standard one. This is actually useful, because calculations tend to be easier for the l^∞ metric than for the standard metric on $\mathbf{R}^2$. (It involves the maximum of two numbers, rather than the square root of the sum of two squares.)

For example, you can check the following important fact. It seems easiest to check by switching from the standard metric to the l^∞ metric on $\mathbf{R}^2$, using the above arguments.

Example. The functions $s : \mathbf{R}^2 \to \mathbf{R}$, $s(x, y) = x+y$, and $p : \mathbf{R}^2 \to \mathbf{R}$, $p(x, y) = xy$, are continuous.

Finally, we make one more geometric definition for metric spaces.

Definition 1.12 *A subset A of a metric space X is closed in X if its complement $X - A = \{x \in X : x \notin A\}$ is open in X.*

Notice that "closed" is not the opposite of "open": some subsets of a metric space are both open and closed, while others are neither. For example, in the real line, the empty set and the whole line are both open and closed; the interval $[0, 1]$ is closed but not open; and the interval $[0, 1)$, like the subset of rational numbers $\mathbf{Q} \subset \mathbf{R}$, is neither open nor closed. In a discrete metric space, you can check that every subset is both open and closed.

We make some last definitions about metric spaces.

Definition 1.13 *A subset of a metric space X is bounded if it is either empty, or contained in an open ball $B_r(x)$ for some point x and some positive real number r.*

For example, the interval $[0, 1]$ in the real line, or any subset of it, is bounded, while the whole real line is unbounded.

Definition 1.14 *The diameter of a bounded subset A in a metric space X is 0 if A is empty, or the real number*

$$\sup_{x,y \in A} d(x,y)$$

if A is not empty.

Here the supremum (or least upper bound) is a real number associated to any nonempty set of real numbers which has an upper bound, by the completeness property of the real numbers. The set of real numbers that comes up in the definition of diameter, the set of distances $d(x,y)$ for $x, y \in A$, is clearly nonempty since A is nonempty, and it has an upper bound since A is bounded. Indeed, if A is contained in the open ball $B_r(x)$, then the distance between any two points of A is less than $2r$, as one checks immediately from the triangle inequality. So the definition of diameter makes sense.

Example. In the metric space $\mathbf{R}^2$, the circle $S^1 = \{(x,y) : x^2 + y^2 = 1\}$, the open ball $\{(x,y) : x^2 + y^2 < 1\}$, and the closed ball $\{(x,y) : x^2 + y^2 \leq 1\}$ all have diameter 2. The supremum in the definition of diameter is actually a maximum in the first and third cases, but not in the second.

Finally, we state some general properties of open subsets in a metric space.

Lemma 1.15 *Let X be a metric space. Then:*

(1) The empty set $\emptyset$ and the whole space X are open in X.

(2) For any open subsets U and V in X, the intersection $U \cap V$ is open in X.

(3) The union of any collection (possibly infinite) of open subsets of X is open in X.

Proof. (1) To show that the empty set is open in X, we have to show that for every element x of the empty set, there is an open ball around x which is contained in the empty set. This is true because the empty set has no elements. (So any statement of the form 'for every element of the empty set, some property holds' is true.) To show that the whole space X is open, we have to show that for every x in X, there is an open ball around x which is contained in X. This is clear, since we can take any open ball, for example the open ball $B_1(x)$.

(2) Given open subsets U and V in X, let us show that the intersection $U \cap V$ is open. That is, for every point x in $U \cap V$, we have to find an open ball around x in X which is contained in $U \cap V$. To say that x is in $U \cap V$ means that x is in U and also in V. Since U and V are open, there are open balls $B_r(x)$ and $B_s(x)$ in X which are contained in U and V, respectively. Let t be the minimum of r and s. This is a positive real number since r and s are positive real numbers. Clearly the open ball $B_t(x)$ is contained both in U and in V, and so it is contained in $U \cap V$, as we want.

(3) Given any collection of open subsets U_i of X, where i runs through any indexing set I, we want to show that the union $\cup_i U_i$ is open in X. So pick any point x in $\cup_i U_i$. To say that x is an element of this union means that there is some $j \in I$ such that x is in U_j. Since U_j is open, there is an open ball $B_r(x)$ which is contained in U_j. Therefore this open ball is also contained in the union $\cup_i U_i$. That is, we have shown that the union is an open subset of X. QED

We will discuss further the different behaviour of unions and intersections of open sets, early in the next lecture.

2 Week 2: Topological spaces

2.1 Day 4: Thursday 29 April

We have shown that whether a function between metric spaces is continuous only depends on which subsets are open (Theorem 1.10). This fact suggests that we might be able to forget about distances, and only keep track of which subsets are open. That thought leads to the following generalization of metric spaces, which we study for the rest of the course. The notion of topological spaces was perhaps the most influential idea in 20th-century mathematics. It offers a huge generalization of the ancient idea of geometry, which turns out to be useful in essentially all areas of mathematics and physics.

Definition 2.1 *(F. Hausdorff, 1914) A topological space X is a set together with a collection of subsets of X, called open subsets, such that the following conditions hold.*

(1) The empty set $\emptyset$ and the whole space X are open in X.

(2) For any open subsets U and V in X, the intersection $U \cap V$ is open in X.

(3) The union of any collection (possibly infinite) of open subsets of X is open in X.

When we speak of a 'space' without further comment, we mean a topological space.

Example. Every metric space is a topological space, by Lemma 1.15. So we have many examples of topological spaces: the real line $\mathbf{R}$, Euclidean space $\mathbf{R}^n$ for any n, any subset of $\mathbf{R}^n$, and so on. When we talk about the real line or $\mathbf{R}^n$ as a topological space without further comment, we have in mind the standard topology on these sets, which means the topology associated to the standard metric.

Remark. In any topological space, the intersection of any finite collection of open subsets, $U_1 \cap \cdots \cap U_n$, is open. This follows by induction on n from property (2) of topological spaces. But the intersection of infinitely many open subsets need not be open, even in the most fundamental examples of topological spaces. For example, in the real line,

$$\cap_{i \geq 1}(-1/i, 1/i) = \{0\},$$

which shows that the intersection of infinitely many open subsets of $\mathbf{R}$ need not be open. This contrasts with the fact that the union of an infinite collection of open subsets is open, in the real line as in any other topological space.

As with metric spaces, the most important point about topological spaces is that they allow us to talk about continuity, as follows.

Definition 2.2 *A function $f : X \to Y$ from one topological space to another is continuous if, for every open subset U of Y, the inverse image $f^{-1}(U)$ is open in X.*

By Theorem 1.10, this definition agrees with our original definition of continuity when both apply, that is, when X and Y are metric spaces. For example, we can ask whether a function $f : \mathbf{R} \to \mathbf{R}$ is continuous, and it does not matter whether we think of the real line as a metric space or a topological space.

As we already saw in the special case of metric spaces, we should realize that the image of an open subset under a continuous mapping need not be open (you should have some examples in mind).

Example. Discrete topological spaces. For any set X, we define the *discrete* topology on X by declaring that every subset of X is open. You can easily check that this is indeed a topology on X, meaning that the properties (1)-(3) are satisfied. Also, you can check that for a discrete topological space X and any topological space Y, every function $f : X \to Y$ is continuous.

It might be useful to point out that every discrete topological space X comes from a metric space, namely from the discrete metric space (X, d) that we have defined.

Example. Two topologically equivalent metrics on a set X determine the same topology on X, clearly.

Example. Indiscrete spaces X. For any set X, we define the *indiscrete* topology on X by declaring that the empty set and the whole set X are the only open subsets of X. You can easily check that this is indeed a topology on X. If the set X has at least two elements, then the indiscrete space X is not *metrizable*, that is, it does not come from any metric on the set X. This is easy to check (using Question 5 of Examples Sheet 1), as we will discuss in more detail later.

We now prove a first general property of continuous functions on topological spaces. This is convenient when you actually have to show that a given function is continuous, by building up from simpler functions which you know are continuous.

Lemma 2.3 *Let $f : X \to Y$ and $g : Y \to Z$ be continuous mappings between topological spaces. Then the composition $g \circ f : X \to Z$ is continuous.*

Proof. Here the composition $g \circ f$ is defined by $(g \circ f)(x) = g(f(x))$. We have to show that for every open subset U of Z, the inverse image $(g \circ f)^{-1}(U)$ is open. You can check that $(g \circ f)^{-1}(U)$ is equal to $f^{-1}(g^{-1}(U))$. Here $g^{-1}(U)$ is open, since g is continuous. Therefore $f^{-1}(g^{-1}(U))$ is open, since f is continuous. QED

2.2 Day 5: Saturday 1 May

We begin with some simple examples of continuous maps which work for any topological space.

Lemma 2.4 *(1) For any topological space X, the identity map $1_X : X \to X$ is continuous.*

(2) For any topological spaces X and Y, any constant map $f : X \to Y$ is continuous.

Proof. (1) For each open subset U of X, its inverse image under the identity map of X is equal to U, which is open. So the identity map is continuous.

(2) To say that $f : X \to Y$ is constant means that there is a point a in Y such that $f(x) = a$ for all x in X. Let U be any open subset of Y. Then $f^{-1}(U)$ is the empty set if a is not in U, while it is the whole space X if a is in U. Both of these are open subsets of X. So f is continuous. QED

Definition 2.5 *Let X be a topological space, and let A be any subset of the set X. We define the subspace topology on A by saying that a subset of A is open in A if and only if it can be written as the intersection $A \cap U$ for some open set U in X.*

You can check that the subspace topology is indeed a topology on A.

If X is a metric space and A is a subset of X, then one can think of two different ways to define a topology on A. First, we can consider the subspace metric on A, and then consider the associated topology. Or, we can consider the topology on X associated to the metric, and then consider the subspace topology on A using the above definition. Fortunately, these two constructions give the same topology on A.

So we have a well-defined topology on any subset of Euclidean space, defined by either of the above methods. For example, we have a well-defined topology on the interval $[0, 1]$, the circle $S^1 = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$, more general curves, and so on.

The following are the basic properties of the subspace topology. We omit the proofs, which are straightforward.

Lemma 2.6 *Let X be a space, $A \subset X$ a subset, with the subspace topology.*

(1) The inclusion $i : A \to X$ is continuous. As a result, for any continuous map $f : X \to Y$, the restriction $f|_A : A \to Y$ is continuous, since it is the composition $f \circ i$.

(2) For any function $g : Z \to A$, g is continuous if and only if $i \circ g : Z \to X$ is continuous.

We now give another important construction of topological spaces, which generalizes the process of going from the real line $\mathbf{R}$ to the plane $\mathbf{R}^2$ or more generally $\mathbf{R}^n$.

Definition 2.7 *Given topological spaces X_1 and X_2, we define the product topology on the set $X_1 \times X_2 = \{(x_1, x_2) : x_1 \in X_1, x_2 \in X_2\}$, as follows. A subset U of $X_1 \times X_2$ is open if and only if it is a union of some collection (possibly infinite) of sets $U_1 \times U_2$, where U_1 is an open subset of X_1 and U_2 is an open subset of X_2.*

We omit the proof that this is indeed a topology on $X_1 \times X_2$, which is not difficult. Also, you can check that the product topology on $\mathbf{R} \times \mathbf{R}$ coincides with the standard topology on $\mathbf{R}^2$. More generally, the product topology on the product of n copies of the real line gives the standard topology on $\mathbf{R}^n$.

Notice that an open subset of $X_1 \times X_2$ need not be of the form $U_1 \times U_2$, or even a finite union of such subsets, even in simple examples. For example, the open ball $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 < 1\}$ is not a finite union of subsets $U_1 \times U_2$, which are roughly open rectangles.

Notice an equivalent formulation: A subset U of $X_1 \times X_2$ is open if and only if, for every point x in U, there are open subsets U_1 in X_1 and U_2 in X_2 such that $x \in U_1 \times U_2$ and $U_1 \times U_2 \subset U$.

We now give the basic examples of continuous functions on product spaces.

Lemma 2.8 *For any spaces X_1 and X_2, the projections $p_1 : X_1 \times X_2 \to X_1$, $p_1(x_1, x_2) = x_1$, and $p_2 : X_1 \times X_2 \to X_2$, $p_2(x_1, x_2) = x_2$, are continuous.*

Proof. The proofs in the two cases are the same, and so we prove that the first projection p_1 is continuous. Let U be any open subset of X_1. Then $p_1^{-1}(U) = U \times X_2$. Since X_2 is an open subset of X_2, this is an open subset of $X_1 \times X_2$, by definition of the product topology. Thus p_1 is continuous. QED

Theorem 2.9 *Let A, X_1, X_2 be topological spaces. Then a function $f : A \to X_1 \times X_2$ is continuous if and only if the two projections $p_1 \circ f : A \to X_1$ and $p_2 \circ f : A \to X_2$ are continuous.*

Proof. First, suppose that $f : A \to X_1 \times X_2$ is continuous. Since the projections p_1 and p_2 are continuous, the compositions $p_1 \circ f$ and $p_2 \circ f$ are continuous.

Conversely, suppose that $p_1 \circ f : A \to X_1$ and $p_2 \circ f : A \to X_2$ are continuous. We change notation by writing $f(a) = (f_1(a), f_2(a))$; then we are assuming that $f_1 : A \to X_1$ and $f_2 : A \to X_2$ are continuous. We now show that $f : A \to X_1 \times X_2$ is continuous. So let U be any open subset of $X_1 \times X_2$. By definition of the product topology, U is a union of some collection of subsets $U_1 \times U_2$ where U_1 is open in X_1 and U_2 is open in X_2. We note that $f^{-1}(\cup_i S_i) = \cup_i f^{-1}(S_i)$ for any collection of subsets S_i. Since the union of any collection of open subsets in A is open, it follows that $f^{-1}(U)$ is open if we can show that $f^{-1}(U_1 \times U_2)$ is open for every open subsets U_1 in X_1 and U_2 in X_2.

Since $f(a) = (f_1(a), f_2(a))$ for all $a \in A$, we can rewrite the set $f^{-1}(U_1 \times U_2)$ as the intersection $f_1^{-1}(U_1) \cap f_2^{-1}(U_2)$. Since f_1 and f_2 are continuous, the subsets $f_1^{-1}(U_1)$ and $f_2^{-1}(U_2)$ are open in A. So the intersection of these two sets is open in A, as we want. QED

Theorem 2.9 makes product spaces very convenient to work with, as we see in the following example.

Example. Show that $f : \mathbf{R} \to \mathbf{R}^2$ defined by $f(x) = (e^x, \sin x)$ is continuous.

Answer: By Theorem 2.9, f is continuous if and only if the two functions $\mathbf{R} \to \mathbf{R}$ defined by e^x and $\sin x$ are continuous. This is true, as we know from earlier calculus courses.

2.3 Day 6: Tuesday 4 May

Here is a fundamental fact, not involving product spaces in the statement, which admits an easy proof using product spaces.

Lemma 2.10 *Let X be a topological space, f and g continuous real-valued functions on X. (As usual, we use the standard topology on the real line here.) Then $f + g$ and fg are also continuous real-valued functions on X.*

Proof. We can view $f+g$ as the composition of the two maps

$$X \overset{\rightarrow}{(f,g)}\, \mathbf{R}^2 \overset{\rightarrow}{+} \mathbf{R}.$$

Here $(f,g) : X \to \mathbf{R}^2$ is continuous by Theorem 2.9, and we have mentioned earlier that the function $+ : \mathbf{R}^2 \to \mathbf{R}$ are continuous. Therefore the composition, $f+g$, is continuous. The same argument works for fg, using that the function $\cdot : \mathbf{R}^2 \to \mathbf{R}$ is continuous. QED

Example. Show that the function $f : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $f(x,y) = (x^2+y^2, e^{xy})$ is continuous.

Answer: First, by Theorem 2.9, it suffices to show that the two functions $\mathbf{R}^2 \to \mathbf{R}$ defined by $(x,y) \mapsto x^2+y^2$ and $(x,y) \mapsto e^{xy}$ are continuous. By Lemma 2.8, we know that the functions $\mathbf{R}^2 \to \mathbf{R}$ defined by $(x,y) \to x$ and $(x,y) \mapsto y$ are continuous. By repeatedly taking sums and products of these functions, it follows that $(x,y) \mapsto x^2+y^2$ is continuous. Also, by taking products of these functions, we know that $(x,y) \mapsto xy$ is continuous. By composing this with the continuous function $\mathbf{R} \to \mathbf{R}$ defined by $x \mapsto e^x$, it follows that e^{xy} is also continuous.

Definition 2.11 *We define a subset A of a topological space X to be closed in X if its complement $X - A$ is open in X.*

It would be possible to base the definition of topological spaces on closed sets rather than open sets. The definition would involve the basic properties of closed sets listed in Examples Sheet 1, Question 14. It is straightforward to check that continuity can be defined using closed sets, as follows.

Lemma 2.12 *Let X and Y be topological spaces. A function $f : X \to Y$ is continuous if and only if the inverse image of every closed subset of Y is closed in X.*

Example. Show that $B = \{(x,y,z) : x^2+y^2+z^2 \leq 1\}$ is closed in $\mathbf{R}^3$.

One answer: Define $f : \mathbf{R}^3 \to \mathbf{R}$ by $f(x,y,z) = x^2+y^2+z^2$. This is continuous, since it is built from the coordinate projections $\mathbf{R}^3 \to \mathbf{R}$ by sums and products. We can think of B as the inverse image of $(-\infty, 1]$ under the function f. Here $(-\infty, 1]$ is closed in $\mathbf{R}$, and so B is closed in $\mathbf{R}^3$ by Lemma 2.12.

Definition 2.13 *Let X and Y be topological spaces. A homeomorphism is a bijective continuous function $f : X \to Y$ such that the inverse function $f^{-1} : Y \to X$ is also continuous. We say that spaces X and Y are homeomorphic if there is a homeomorphism $f : X \to Y$.*

To say that a function $f : X \to Y$ is bijective, also called a 'one-to-one correspondence', means that f is both one-to-one and onto. Equivalently, for every point y in Y, there is a unique point x in X such that $f(x) = y$. For a bijective function f, there is an inverse function $f^{-1} : Y \to X$, mapping a point y of Y to the unique point x in X such that $f(x) = y$.

The following statement, easy to prove, clarifies the meaning of homeomorphisms.

Lemma 2.14 *Let X and Y be topological spaces. A function $f : X \to Y$ is a homeomorphism if and only if: f is bijective and a subset of X is open if and only if its image in Y is open.*

This means that two homeomorphic spaces can be thought of as the same topological space, with different names for the points. As a result, any property of a topological space which can be defined using only its topology (for example, which functions on it are continuous) is equivalent to a corresponding property for any homeomorphic space.

Examples. The closed intervals $[0, 1]$ and $[0, 2]$ are homeomorphic. Explicitly, $f : [0, 1] \to [0, 2]$ defined by $f(x) = 2x$ is a homeomorphism, since it is bijective and continuous, and the inverse function is $x \mapsto x/2$, which is also continuous.

The open intervals $(0, 1)$, $(0, \infty)$, and the real line $\mathbf{R}$ are all homeomorphic. For example, the function $\tan x : (-\pi/2, \pi/2) \to \mathbf{R}$ is a homeomorphism, and there is an easy homeomorphism from $(0, 1)$ to $(-\pi/2, \pi/2)$ of the form $x \mapsto ax + b$. To show that $(0, \infty)$ is homeomorphic to these other spaces, we can use, for example, that $x \mapsto 1/x$ is a homeomorphism from $(0, 1)$ to $(1, \infty)$, and the latter space is homeomorphic to $(0, \infty)$ via $x \mapsto x - 1$.

The circle $S^1 = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$ is homeomorphic to a square, a triangle, or (informally speaking) any other closed curve which does not cross itself. But, for example, the interval $[0, 1]$ is not homeomorphic to the circle, as we will prove later in the course. To give more examples, without proofs: the capital letters A and B (viewed as closed subsets of $\mathbf{R}^2$) are not homeomorphic since A has 'one hole' and B has 'two holes'. Also, the letters A and O are not homeomorphic even though both have 'one hole', because A has 4 'special points' and O has none. On the other hand, the letters Q and R are homeomorphic.

Definition 2.15 *A topological space X is metrizable if there is a metric d on X such that the open subsets of X are exactly the d-open sets.*

Most of the interesting topological spaces are metrizable, for example $\mathbf{R}^n$ and all its subsets. But not all topological spaces are metrizable, as we will check using the following notion.

Definition 2.16 *A topological space X is Hausdorff if, for any $x, y \in X$ with $x \neq y$, there are open sets U and V with $x \in U$, $y \in V$, and $U \cap V = \emptyset$.*

Lemma 2.17 *Every metrizable topological space is Hausdorff.*

Proof. Let X be a metrizable topological space. So there is a metric d on X which induces the given topology on X. To show that X is Hausdorff, pick any two distinct points x and y in X. Then the distance from x to y is positive (not zero). By Examples Sheet 1, question 5, there are open balls around x and y which are disjoint from each other. QED

Example. An indiscrete topological space X with at least two points is not Hausdorff. Indeed, for any two distinct points x and y in X, any open subset U of X containing x must be the whole space X, and any open subset V of X containing y is also the whole space. So U and V cannot be disjoint; that is, X is not Hausdorff.

By Lemma 2.17, it follows that an indiscrete topological space with at least two points is not metrizable.

Part IA Metric and Topological Spaces, second set of notes

Burt Totaro

11 May 2004

3 Week 3: Connected topological spaces

3.1 Day 7: Thursday 6 May 2004

We now make precise the idea of a topological space being 'all in one piece'.

Definition 3.1 *A topological space X is connected if X is nonempty and the only subsets of X that are both open and closed are the empty set and the whole space X.*

Since a subset is closed if and only if its complement is open, we can rephrase this definition just in terms of open sets. Namely, a space X is connected if and only if it is nonempty and it is impossible to write X as the union of two disjoint nonempty open subsets. Intuitively, this means that X is connected if it is impossible to divide X into two pieces which are 'separated' from each other.

Another equivalent condition: a space X is connected if and only if it is nonempty and there is no continuous function from X onto the subset $\{0, 1\}$ of $\mathbf{R}$. Indeed, given such a function, X would be the union of the two disjoint open subsets $f^{-1}(0)$ and $f^{-1}(1)$, which are nonempty because f is onto, and so X would not be connected. Conversely, if X is not connected, then we can write $X = A \cup B$ with A and B disjoint nonempty open subsets of X. In that case, we can define a function f from X onto $\{0, 1\}$ by mapping A to 0 and B to 1; this is continuous because A and B are both open.

Example. The subspace $X = [0, 1] \cup [2, 3]$ of the real line is not connected, because the subsets $[0, 1]$ and $[2, 3]$ are both open in X.

Although the definition of connected spaces is fairly simple, it is nontrivial to show that even simple spaces like the real line $\mathbf{R}$ or the unit interval $[0, 1]$ are connected. In fact, to prove this, we will need the completeness property of the real numbers, for essentially the first time in the course. (It came up in the definition of diameter, Definition 1.14, but we made no further use of that.) Indeed, until now, all our results about the topological space $\mathbf{R}$ would apply equally well to the topological space $\mathbf{Q}$ of rational numbers; but the space $\mathbf{Q}$ is not connected. For example, we can write it as the union of the following two disjoint nonempty open subsets:

$$\mathbf{Q} = \{x \in \mathbf{Q} : x^2 > 2\} \cup \{x \in \mathbf{Q} : x^2 < 2\},$$

because the square root of 2 is irrational.

Define an *interval* in the real line to be a subset of the form $[a, b]$ or $[a, b)$ or $(a, b]$ or (a, b) for $a < b$, or $[a, \infty)$ or (a, ∞), or $(-\infty, a]$ or $(-\infty, a)$, or the whole line $\mathbf{R}$, or a single point $\{a\}$. (In some situations, one might not want to call a single point an interval, but for now I will do so.)

Theorem 3.2 *Every interval $I \subset \mathbf{R}$ is connected.*

Notice that, when we say that a given subset of a topological space is connected, we are talking about the subspace topology on that subset.

Proof. Clearly I is nonempty, as we check from the definition of an interval. So we just need to show that it is impossible to write I as the union of two disjoint nonempty open subsets A and B in I. Suppose that we can write I in that way; we will derive a contradiction.

Since A and B are nonempty, we can pick a point a in A and a point b in B. Since A and B are disjoint, a is not equal to b. We can assume that $a < b$; if not, switch the two sets A and B.

Since a and b are in I and I is an interval, we see (by inspection of the various types of interval) that I contains the whole closed interval $[a, b]$. To avoid the inconvenience of dealing with the various types of interval, we now replace I by $[a, b]$, A by $A \cap [a, b]$, and B by $B \cap [a, b]$. Here $A \cap [a, b]$ and $B \cap [a, b]$ are both open subsets of $[a, b]$, by definition of the subspace topology on $[a, b]$. They are disjoint since A and B are disjoint. Finally, we know that a is in $A \cap [a, b]$ and that b is in $B \cap [a, b]$.

Thus (writing A instead of $A \cap [a, b]$ and B instead of $B \cap [a, b]$), we now have the following situation. The closed interval $[a, b]$ is a union of two disjoint open subsets A and B with $a \in A$ and $b \in B$. We want to derive a contradiction from this.

The key idea is to consider the real number $c = \sup A$, which exists by the completeness property of the real numbers. Indeed, the set A is nonempty since it contains a, and it has an upper bound, namely b; so it has a least upper bound, namely c. Since A is open in $[a, b]$ and contains a, A must contain $[a, a + \epsilon)$ for some $\epsilon > 0$; therefore, $c = \sup A$ must be greater than a. Also, since B is open in $[a, b]$ and contains b, B contains $(b - \epsilon, b]$ for some $\epsilon > 0$. Therefore, $c = \sup A$ is less than b. Thus, c is in the open interval (a, b).

We ask whether c is in A or B; we will get a contradiction either way. Suppose c is in B. In that case, since B is open in $[a, b]$, B contains the open interval $(c - \epsilon, c + \epsilon)$ for some $\epsilon > 0$. This contradicts the fact that c is the supremum of A. On the other hand, suppose c is in A. Since A is open in $[a, b]$, A contains the open interval $(c - \epsilon, c + \epsilon)$ for some $\epsilon > 0$. But then $c = \sup A$ would be at least $c + \epsilon$, which is a contradiction. This completes the proof that every interval in the real line is connected. QED

To show the depth of Theorem 3.2, we can use it to prove that certain topological spaces are not homeomorphic to each other, something which is in general very hard.

Example. Show that the spaces $[0, 1]$ and $[0, 1] \cup [2, 3]$ are not homeomorphic. Answer: $[0, 1] \cup [2, 3]$ is not connected (as we showed easily, above), whereas $[0, 1]$ is connected (as we showed with difficulty in Theorem 3.2). Therefore these two spaces cannot be homeomorphic, since two homeomorphic spaces have all the same topological properties.

We can complete Theorem 3.2 by showing that this theorem gives all the connected subspaces of the real line, as follows.

Lemma 3.3 *Any connected subspace of* $\mathbf{R}$ *is an interval.*

Proof. We use the following characterization of intervals: a subset S of the real line is an interval if and only if it is nonempty and, for all real numbers $a < b < c$ such that a and c are in S, b is also in S. I omit the proof of this, which is easy but a bit messy because of the various types of intervals.

Now let S be a connected subspace of $\mathbf{R}$. We know that S is nonempty. It remains to prove that for all real numbers $a < b < c$ with a and c in S, b must also be in S. So suppose that b is not in S. Then we can write

$$S = (S \cap (-\infty, b)) \cup (S \cap (b, \infty)).$$

This expresses S as the union of two disjoint open subsets. They are both nonempty, because a is in the first subset and c is in the second subset. This contradicts the fact that S is connected. So in fact S must be an interval. QED

Lemma 3.4 *Let $f : X \to Y$ be a continuous mapping between topological spaces. If X is connected, then the image $f(X) \subset Y$ is connected.*

Here, to talk about connectedness of $f(X) \subset Y$, we view it as a topological space using the subspace topology.

Proof. By Lemma 2.6 (2), since $f : X \to Y$ is continuous and maps into the subset $f(X)$ of Y, the function $f : X \to f(X)$ is also continuous, where we give $f(X)$ the subspace topology. Let us replace Y by the subspace $f(X)$ of Y. Then the situation is: we have a continuous map f from X onto Y, with X connected, and we want to show that Y is connected.

Clearly Y is nonempty since X is nonempty. To show that Y is connected, we have to show that we cannot write Y as the union of two disjoint nonempty open subsets A and B. Suppose we can. Then X is the union of the inverse images $f^{-1}(A)$ and $f^{-1}(B)$. These are open in X because f is continuous. They are disjoint, because A and B are disjoint. Finally, $f^{-1}(A)$ and $f^{-1}(B)$ are both nonempty, because A and B are nonempty and f is onto. This contradicts the fact that X is connected. Thus we have shown that the image of f is connected. QED

Corollary 3.5 *(Intermediate Value Theorem) For any continuous function $f : [a, b] \to \mathbf{R}$, the image $f([a, b])$ is an interval in $\mathbf{R}$. In particular, for any real number c between $f(a)$ and $f(b)$, there is a point x in $[a, b]$ with $f(x) = c$.*

Proof. The interval $[a, b]$ is connected by Theorem 3.2. So its image $f([a, b]) \subset \mathbf{R}$ is connected by Lemma 3.4. By Lemma 3.3, this means that $f([a, b])$ is an interval in $\mathbf{R}$. The second statement then follows from the properties of intervals I in the real line: if $u < c < v$ with u and v in I, then c is also in I. QED

3.2 Day 8: Saturday 8 May

The familiar Intermediate Value Theorem for real-valued functions on an interval (Corollary 3.5) generalizes to real-valued functions on any connected space, with the same proof. Since we there are many more examples of connected spaces, as we will soon see, this generalization shows the value of topological ideas.

Corollary 3.6 *(Intermediate Value Theorem for connected spaces) For any continuous real-valued function f on a connected topological space X, the image $f(X)$ is an interval. Therefore, if there are points x_1 and x_2 in X with $f(x_1) = a_1$ and $f(x_2) = a_2$ with $a_1 < a_2$, then for any real number b in the interval $[a_1, a_2]$, there is a point x in X with $f(x) = b$.*

Proof. Since X is connected, the image $f(X) \subset \mathbf{R}$ is connected, by Lemma 3.4. This connected subset of the real line must be an interval, by Lemma 3.3. This proves the first statement of the corollary; the second statement follows from the properties of intervals in the real line. QED

To give more examples of connected spaces, one can use the following results, to be proved in Examples Sheet 2.

Lemma 3.7 *Let X be a topological space. Suppose that X is the union of two connected subspaces A and B whose intersection is nonempty. Then X is connected.*

Lemma 3.8 *For any connected topological spaces X and Y, the product space $X \times Y$ is connected.*

So, for example, Euclidean space $\mathbf{R}^n$ is connected for any n, by induction using Lemma 3.8. Likewise, the n-dimensional cube $[0, 1]^n$ is connected for any n.

Example. Show that the letter 'T' is connected; we could define this precisely as the subspace $X = ([0,1] \times \{0\}) \cup (\{1/2\} \times [-1,0])$ of $\mathbf{R}^2$. One answer: The space X is the union of two subspaces which are homeomorphic to the closed interval and whose intersection is nonempty; so X is connected by Lemma 3.7.

We now turn to a different notion of connectedness, based upon the idea of a path.

Definition 3.9 *A path in a topological space X is a continuous function $f : [0,1] \to X$. We say that f is a path from $f(0)$ to $f(1)$ in X.*

Definition 3.10 *A topological space X is path-connected if it is nonempty and, for all points x and y in X, there is a path from x to y.*

Example. The real line is path-connected. To prove this, let x and y be any real numbers. We can define a path from x to y, $f : [0,1] \to \mathbf{R}$, by:

$$f(t) = (1-t)x + ty.$$

This is a continuous function from $[0,1]$ to $\mathbf{R}$, with $f(0) = x$ and $f(1) = y$. So we have shown that $\mathbf{R}$ is path-connected.

More generally, every interval I in the real line is path-connected. To prove this, let x and y be any points in I. Then we can define a path from x to y, $f : [0,1] \to I$, by the same formula. This function clearly has all the right properties if we can show that it does map into I (which would not be true for an arbitary subset of $\mathbf{R}$). This is true because, for x and y in an interval I, and for any $t \in [0,1]$, the real number $(1-t)x + ty$ is between x and y, and therefore also belongs to the interval I.

Also, Euclidean space $\mathbf{R}^n$ is also path-connected. Indeed, for any points x and y in $\mathbf{R}^n$, there is a path from x to y, going along the line segment from x to y. We can define this by the same formula:

$$f(t) = (1-t)x + ty$$

for $t \in [0,1]$. One just has to interpret this formula correctly: for a real number t and a point $x = (x_1, \ldots, x_n)$ in $\mathbf{R}^n$, tx means the point $(tx_1, \ldots, tx_n)$ in $\mathbf{R}^n$. You can check that $f : [0,1] \to \mathbf{R}^n$ is continuous and goes from x to y, so that $\mathbf{R}^n$ is path-connected.

We mention one last generalization: any nonempty convex subset of $\mathbf{R}^n$ is path-connected. By definition, a subset of $\mathbf{R}^n$ is convex if it contains the line segment between any two points in the subset. Then, for a convex subset X in $\mathbf{R}^n$, there is an obvious path between any two points x and y in X, namely the straight-line path defined by the same formula as above. So X is path-connected.

The following fact makes the notion of path-connectedness easier to work with.

Lemma 3.11 *Let X be a topological space. If there is a path from x to y in X, and a path from y to z, then there is a path from x to z.*

Proof. Let $f : [0,1] \to X$ and $g : [0,1] \to X$ be paths from x to y and from y to z. Define $h : [0,1] \to X$ by:

$$h(t) = \begin{cases} f(2t) & \text{if } t \in [0,1/2] \\ g(2t-1) & \text{if } t \in [1/2,1]. \end{cases}$$

(This is the natural way to define a path which goes along the path f from x to y, and then along the path g from y to z.) Clearly h is a well-defined function from $[0,1]$ to X, because the two parts of the definition agree when $t = 1/2$, using that $f(1) = y = g(0)$. Also, it is clear that $h(0) = x$ and $h(1) = z$. So we are done if we can prove that h is continuous.

Clearly the restrictions of h to the intervals $[0,1/2]$ and $[1/2,1]$ are continuous, since they are defined as composites of continuous functions. We want to show that h is continuous on the whole interval $[0,1]$. So let U be any open subset of X; we have to show that $h^{-1}(U)$ is open in $[0,1]$. Since the restrictions of h to the two half-intervals are continuous, we know that $h^{-1}(U) \cap [0,1/2]$ and $h^{-1}(U) \cap [1/2,1]$ are open in $[0,1/2]$ and in $[1/2,1]$, respectively.

To show that $h^{-1}(U)$ is open in $[0,1]$, pick any point t in $h^{-1}(U)$; we have to show that the subset $h^{-1}(U)$ contains an open ball around t in $[0,1]$. This is clear if t is not equal to $1/2$, using that $h^{-1}(U) \cap [0,1/2]$ is open in $[0,1/2]$ and $h^{-1}(U) \cap [1/2,1]$ is open in $[1/2,1]$. So suppose that $t = 1/2$ is in $h^{-1}(U)$. From the openness of $h^{-1}(U)$ in the two half-intervals, we know that $h^{-1}(U)$ contains $(1/2-\epsilon, 1/2]$ for some $\epsilon > 0$, and also that it contains $[1/2, 1/2+\epsilon)$ for some $\epsilon > 0$. Therefore (taking the smaller of these two positive numbers ϵ), $h^{-1}(U)$ contains $(1/2-\epsilon, 1/2+\epsilon)$. This completes the proof that $h^{-1}(U)$ is open in $[0,1]$. QED

We now relate our two notions of connectedness.

Lemma 3.12 *Every path-connected space is connected.*

Proof. Let X be a path-connected space. We know that X is nonempty. To show that X is connected, we have to show that we cannot write X as the union of two disjoint nonempty open subsets A and B. Suppose we can; we will derive a contradiction.

Since A and B are nonempty, we can pick a point a in A and a point b in B. Since X is path-connected, there is a path $f : [0,1] \to X$ from a to b. We see that the interval $[0,1]$ is the union of the two subsets $f^{-1}(A)$ and $f^{-1}(B)$. They are disjoint, since A and B are disjoint. Both subsets are open in $[0,1]$, since f is continuous and A and B are open in X. Finally, both subsets of $[0,1]$ are nonempty, since 0 is in $f^{-1}(A)$ and 1 is in $f^{-1}(B)$. This contradicts the connectedness of $[0,1]$. QED

For subsets of the real line, the converse holds: a connected subspace of the real line is an interval (Lemma 3.3) and hence is also path-connected. But for general topological spaces, even subspaces of the plane, connectedness does not imply path-connectedness.

(*) **Example.** Let $X \subset \mathbf{R}^2$ be the union of $A = \{0\} \times [-1,1]$ and $B = \{(x, \sin(1/x)) : x \in (0,1]\}$. Then X is connected but not path-connected. (Draw a picture.) The proof of these properties of X is complicated, and is in any case non-examinable.

Although not every connected space is path-connected, any connected space which one encounters in daily life is likely to be path-connected. One might consider connectedness to be the more fundamental notion, while path-connectedness has a clearer geometric meaning and often seems easier to check in examples.

Example. Show that $\mathbf{R}^2 - \{0\}$ is connected. (Here 0 denotes the point $(0,0)$ in $\mathbf{R}^2$.)

One could prove this from the general properties of connected spaces (notably, Lemma 3.7). But it seems more intuitively appealing to show that $\mathbf{R}^2 - \{0\}$ is path-connected; of course that implies that it is connected. Notice that this is not completely trivial, in that $\mathbf{R}^2 - \{0\}$ is not convex.

To show that $\mathbf{R}^2 - \{0\}$ is path-connected, we first observe that many points in X can be connected to the point $(1,0)$ (for example) by a single straight-line path. Namely, this works for all the points in $\mathbf{R}^2 - \{0\}$ except those in the ray $(-\infty, 0) \times \{0\}$. But all the points in that ray can be connected to the point $(-1,1)$ by a single straight-line path, and then a second straight-line path goes from $(-1,1)$ to $(1,0)$. (I am using Lemma 3.11 tacitly here.) So $\mathbf{R}^2 - \{0\}$ is path-connected.

Remark. Any connected *open* subset U of Euclidean space is path-connected.

Proof. Choose a point x in U. Let A be the set of points in U which can be reached by a path starting at x, and let B be the complement $A - U$. I claim that A and B are both open in $\mathbf{R}^n$. Indeed, for any point a in A, U contains some open ball around a in $\mathbf{R}^n$, and this is a convex set. So all the points in this open ball can be reached by a path from a, hence by a path from x (using Lemma 3.11); that is, all the points in this open ball are in A. So A is open in U.

Likewise, let b be any point in B. Then U contains some open ball around b in $\mathbf{R}^n$. If any of the points in this ball were in A, that is, if any of those points could be reached by a path from x, then so would b be (using a straight line path from the given point to b, plus Lemma 3.11 again), which is a contradiction. So B contains an open ball around b in $\mathbf{R}^n$. This shows that B is open in U. Also, A is nonempty, since it contains the point x. By connectedness of U, it follows that B must be empty. This means exactly that U is path-connected. QED

3.3 Day 9: Tuesday 11 May

Today, we finish our work on connectedness, and then make a series of definitions about closures in topological spaces and completeness in metric spaces.

For any topological space X, and any two points x and y in X, say that $x \sim y$ if there is a connected subspace of X which contains both x and y. This is an equivalence relation, by Lemma 3.7.

Definition 3.13 *The connected components of a topological space X are the equivalence classes for the above equivalence relation on X.*

One can check that the connected components of X are, in fact, connected. Moreover, any connected subspace of X is contained in a connected component.

Examples. For X connected, X has just one connected component, the whole space. For $X = [0,1] - \{1/2\}$, the connected components are $[0, 1/2)$ and $(1/2, 1]$. For $X = \mathbf{Q}$, with the subspace topology from the real line, the connected components are just the points of $\mathbf{Q}$, because any connected subspace of $\mathbf{Q}$ is a point. This follows, for example, from the fact that the only connected subspaces of the real line are intervals (Lemma 3.3).

Remark. If two spaces are homeomorphic, then they have the same number of connected components, clearly. This gives another possible way to show that two given spaces are not homeomorphic.

Example. Show that the closed interval $X = [0,1]$ and the letter 'T', defined as the subspace

$$Y = ([0,1] \times \{0\}) \cup (\{1/2\} \times [-1,0])$$

of $\mathbf{R}^2$ are not homeomorphic.

Answer: Both X and Y are connected (in fact, path-connected), and so we cannot distinguish them that way. But suppose that there is a homeomorphism $f : X \to Y$. Then, for every point $y \in Y$, f restricts to a homeomorphism from $X - \{f^{-1}(y)\}$ to $Y - \{y\}$. In particular, $X - \{f^{-1}(y)\}$ must have the same number of connected components as $Y - \{y\}$. For the point $y = (1/2, 0)$ in Y, however, $Y - \{y\}$ has 3 connected components. By contrast, $X - \{x\}$ has 1 or 2 connected components for all $x \in X = [0, 1]$. So in fact X and Y are not homeomorphic.

The same method can be used to show that $\mathbf{R}$ and $\mathbf{R}^2$ are not homeomorphic, and also that the interval $[0,1]$ and the circle S^1 are not homeomorphic (Examples Sheet 2, Questions 6 and 12).

We now turn away from connectedness to make some general definitions about closures in topological spaces and completeness in metric spaces. These ideas help to prepare for the final section of the course, on compactness.

Definition 3.14 *Let S be a subset of a topological space X. The closure $\overline{S}$ of S in X is the intersection of all closed subsets of X which contain S.*

Since the intersection of any collection of closed subsets is closed, the closure of S is itself a closed subset of X. Clearly, it is the smallest closed subset of X that contains S.

Examples. The closure of $[0, 1]$ in $\mathbf{R}$ is $[0, 1]$, since this is a closed set. The closure of $(0, 1)$ in $\mathbf{R}$ is $[0, 1]$.

Definition 3.15 *Let S be a subset of a topological space X. The interior int S of S in X is the union of all open subsets of X which are contained in S.*

Since the union of any collection of open subsets is open, the interior of S is an open subset of X. It is the largest open subset of X contained in S.

Examples. The interior of the closed ball $\{(x,y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1\}$ in $\mathbf{R}^2$ is the open ball $\{(x,y) \in \mathbf{R}^2 : x^2 + y^2 < 1\}$. The interior of $[0,1] \times \{0\}$ in $\mathbf{R}^2$ is the empty set.

Definition 3.16 *Let X be a topological space. We say that a subset S of X is dense in X if the closure $\overline{S}$ is equal to X.*

Example. Both $\mathbf{Q}$ and $\mathbf{R} - \mathbf{Q}$ (the sets of rational and irrational numbers) are dense in the real line. This follows from the fact that every open interval in the real line contains both a rational and an irrational number. In particular, we see that an uncountable topological space can have a countable dense subset, since $\mathbf{Q}$ is dense in $\mathbf{R}$.

In the special case of metric spaces, we can give an alternative, more explicit description of the closure of a subset, Lemma 3.18 below. First, we need the following basic definition.

Definition 3.17 *Let X be a metric space. We say that a sequence $x_1, x_2, \ldots$ in X converges to a point x in X if, for every open ball U around x, there is an integer N such that x_i is in U for all $i \geq N$.*

The limit of a convergent sequence in a metric space is unique, by Examples Sheet 1, Question 9.

Lemma 3.18 *Let X be a metric space. The closure $\overline{S}$ of a subset S in X is the set of points x in X such that there is a sequence of points in S which converges to x in X.*

Proof. First, let $x_1, x_2, \ldots$ be a sequence of points in S which converges to a point x in X. We want to show that x is in the closure $\overline{S}$ of S in X. Suppose that x is not in $\overline{S}$. Since $\overline{S}$ is closed, its complement $X - \overline{S}$ is open, and we are assuming that it contains the point x. So there is an open ball U around x which is contained in $X - \overline{S}$. This open ball does not contain any of the points x_i, since they are in S and hence in $\overline{S}$. This contradicts the fact that the sequence x_i converges to x. Thus, we have shown that the limit point x must be in the closure $\overline{S}$.

Conversely, let x be any point in the closure $\overline{S}$. Then, for any open ball U around x, $X - U$ is a closed set that does not contain x. If $X - U$ contained the subset S, then it would also contain the closure $\overline{S}$ and hence the point x; so $X - U$ does not contain S. That is, for every open ball U around x, there is a point in $S \cap U$.

As a result, for each positive integer i, we can choose a point x_i in $S \cap B_{1/i}(x)$. Thus we have a sequence $x_1, x_2, \ldots$ in S such that $d(x_i, x) < 1/i$ for all i. Therefore the sequence x_i converges to x. Thus we have found a sequence in S that converges to any given point x in the closure $\overline{S}$. QED

For example, by Lemma 3.18, we can rephrase the fact that $\mathbf{Q}$ is dense in $\mathbf{R}$ by saying that every real number is the limit of some sequence of rational numbers.

We now turn to the notion of completeness for metric spaces.

Definition 3.19 *A sequence $x_1, x_2, \ldots$ in a metric space X is a Cauchy sequence if, for every $\epsilon > 0$ there is an integer N such that $d(x_i, x_j) < \epsilon$ for all $i, j \geq N$.*

Any convergent sequence in a metric space is Cauchy, as one easily checks.

Definition 3.20 *A metric space X is complete if every Cauchy sequence in X is convergent (to some point in X).*

Completeness of a metric space is a remarkable property: it says that you can check whether a sequence has a limit or not just by looking at the distance between different points in the sequence, without knowing what the limit point might be.

Lemma 3.21 *The metric space* **R** *is complete.*

Proof. Let $x_1, x_2, \ldots$ be a Cauchy sequence of the real numbers; we want to show that this sequence converges to some real number. First, using the definition of a Cauchy sequence, we notice that this sequence is bounded. That is, it is contained in some interval $[a, b]$. As a result, using the completeness property of the real numbers, we can define a real number

$$y_N := \sup_{i \geq N} x_i$$

for each positive integer N. Clearly y_N is also in the interval $[a, b]$, and we have $y_1 \geq y_2 \geq \cdots$. Therefore, again using the completeness property of the real numbers, we can define a real number

$$x := \inf_{N \geq 1} y_N.$$

I claim that the sequence x_i converges to x.

To prove this, pick any $\epsilon > 0$. Since x_i is a Cauchy sequence, there is a positive integer M such that $d(x_i, x_j) < \epsilon$ for all $i, j \geq M$. In particular, all the numbers x_i with $i \geq M$ are in the closed interval $[x_M - \epsilon, x_M + \epsilon]$, and so the supremum y_N is also in that interval for all $N \geq M$. Therefore, also, the infimum x is in that same interval. (This uses that, because the sequence y_N is decreasing, the numbers y_N with $N < M$ are irrelevant to the definition of x.) As a result, we have $d(x_i, x) \leq 2\epsilon$ for all $i \geq M$. This shows that x_i converges to x. QED

Examples. The metric space **Q** is not complete. Indeed, any sequence in **Q** which converges to the real number $\sqrt{2}$ is Cauchy, but does not converge to any point in **Q**.

The open interval $(0, 1)$ is not complete: the sequence $x_i := 1/i$ in $(0, 1)$ is Cauchy, but does not converge to any point in $(0, 1)$.

Since **R** and $(0, 1)$ are homeomorphic, but **R** is complete while $(0, 1)$ is not, we see that completeness of metric spaces is not a topological property.

Lemma 3.22 *The metric space* $\mathbf{R}^n$ *is complete.*

Proof. As usual, we refer here to the standard metric on $\mathbf{R}^n$. Let $x^1, x^2, \ldots$ be a Cauchy sequence in $\mathbf{R}^n$. Thus, for every $\epsilon > 0$, there is an integer N such that $d(x^i, x^j) < \epsilon$ for all $i, j \geq N$. We recall the definition of the metric on $\mathbf{R}^n$:

$$d(x, y) = [\sum_{m=1}^{n} (y_m - x_m)^2]^{1/2}.$$

In particular, for each $m = 1, \ldots, n$, the distance between the mth coordinates x_m and y_m is at most the distance between x and y. It follows that, for each $m = 1, \ldots, n$, the mth coordinates of our sequence, $x_m^1, x_m^2, \ldots$, form a Cauchy sequence in the real line.

By completeness of **R**, it follows that, for each $m = 1, \ldots, n$, the sequence $x_m^1, x_m^2, \ldots$ converges to some real number x_m. I claim that our sequence $x^1, x^2, \ldots$ in $\mathbf{R}^n$ converges to the point $x := (x_1, \ldots, x_n)$ in $\mathbf{R}^n$. We know that for every $\epsilon > 0$, and every $m = 1, \ldots, n$, there is an N_m such that $d_{\mathbf{R}}(x_m^i, x_m) < \epsilon$ for all $i \geq N_m$. It follows that, for every $i > \max(N_1, \ldots, N_n)$,

$$\begin{aligned} d_{\mathbf{R}^n}(x^i, x) &= [\sum_{m=1}^{n} (x_m^i - x_m)^2]^{1/2} \\ &< n^{1/2}\epsilon. \end{aligned}$$

Here $n^{1/2}$ is just a constant. So this shows that the sequence x^i converges to x in $\mathbf{R}^n$. Thus $\mathbf{R}^n$ is complete. QED

Lemma 3.23 *(1) A subspace X of any metric space Y which is complete (using the subspace metric) must be closed in Y.*

(2) Let Y be a complete metric space. Then a subspace X of Y is complete if and only if it is closed in Y.

Lemma 3.23 makes it easy to check completeness for a large class of metric spaces. For example, this lemma tells us that a subset X of $\mathbf{R}^n$ is complete in the subspace metric if and only if it is closed in $\mathbf{R}^n$.

Proof. (1) Let X be a complete subspace of a metric space Y. By Lemma 3.18, to show that X is closed in Y, it is equivalent to show that for every sequence in X with a limit point in Y, the limit is actually in X. To prove this, note that such a sequence must be a Cauchy sequence. By completeness of X, this sequence has a limit point in X. So the limit point in Y is actually in X (by uniqueness of the limit).

(2) The direction $\Longrightarrow$ follows from (1) (even without knowing that Y is complete). The direction $\Longleftarrow$ is Examples Sheet 1, Question 20. It is proved most easily using the description of closed sets given by Lemma 3.18. Namely, let X be a closed subset of a complete metric space Y. We want to show that X is complete (with the subspace metric). So take any Cauchy sequence in X. Since the bigger space Y is complete, our sequence converges to some point in Y. By Lemma 3.18, since X is closed, every sequence in X that converges to a point in Y actually has its limit point in X. So our Cauchy sequence actually converges to a point in X. That is, X is a complete metric space. QED

Part IA Metric and Topological Spaces, third set of notes

Burt Totaro

18 May 2004

4 Week 4: Compact topological spaces

4.1 Day 10: Thursday 13 May 2004

The last week of the course will be devoted to "compact" topological spaces. A basic example of a compact space will be a closed interval in the real line. One main result about compact spaces will be that any continuous real-valued function on a compact space is bounded, and moreover attains its maximum and minimum. Notice that this does not hold for open intervals (the function $1/x$ on $(0, 1)$ is continuous, but unbounded). So open intervals will not be compact.

Definition 4.1 *A topological space X is compact if, for every collection of open subsets of X whose union is X, X is actually the union of finitely many of these open subsets.*

More briefly: X is compact if every open cover of X has a finite subcover.

Examples. (1) Any finite topological space is compact. Indeed, if the set X is finite, then X has only finitely many open subsets, and so any open cover of X has a finite subcover.

In fact, we can think of compactness as a natural generalization of finiteness. Most of the properties of compact spaces will be generalizations of obvious properties of finite sets. (For example, every real-valued function on a finite set is bounded.)

(2) The real line $\mathbf{R}$ is not compact. For example,

$$\mathbf{R} = \cup_{n\geq 1}(-n, n),$$

which says that the open sets $(-n, n)$ form an open cover of $\mathbf{R}$. This open cover has no finite subcover, since the union of any finite number of these sets $(-n, n)$ will be bounded and hence not all of $\mathbf{R}$.

(3) The open interval $(0, 1)$ is not compact. This follows from the fact that the real line is not compact, since $(0, 1)$ and $\mathbf{R}$ are homeomorphic (and compactness is clearly a topological property). We can also check directly that $(0, 1)$ is not compact: the sets $(1/n, 1)$ for positive integers n form an open cover of $(0, 1)$ which has no finite subcover.

As with connected spaces, it is easy to give examples of non-compact spaces (as above), but hard to prove that any interesting space is compact. We now give the fundamental example of a compact topological space. Just as with the corresponding fact for connectedness (that intervals are connected), the proof is fairly deep, relying upon the completeness property of the real numbers.

Theorem 4.2 *The closed interval $[a, b]$, for real numbers $a < b$, is compact.*

Proof. Suppose that $[a, b]$ is the union of a collection of open sets U_i, where i runs over some indexing set I. We need to show that $[a, b]$ is the union of finitely many of the sets U_i.

The basic idea is to consider the set A of numbers x in $[a, b]$ such that the interval $[a, x]$ is covered by finitely many of the sets U_i. Clearly we will be done if we can show that b belongs to the set A.

Since the open sets U_i cover $[a, b]$, the point a belongs to some set U_i. Therefore a belongs to the set A (because 1 is a finite number). In particular, since A is nonempty and bounded, we can consider the real number $c = \sup A$, using the completeness property of the real numbers.

Let U_i be an open set in the given collection which contains a. Since U_i is open in $[a, b]$, it contains the interval $[a, a + \epsilon)$ for some $\epsilon > 0$. Then it is clear that $[a, a + \epsilon)$ is contained in A. Therefore $c = \sup A$ is greater than a.

Since c belongs to the interval $[a, b]$, c is in one of the open sets U_j in the given collection. We know that c is greater than a; suppose that c is also less than b. Then, since U_j is open in $[a, b]$, U_j must contain $(c - \epsilon, c + \epsilon)$ for some $\epsilon > 0$. Since $c = \sup A$, the definition of A shows that the interval $[a, c - \epsilon/2]$ is covered by finitely many of the open sets U_i. Therefore, using those open sets together with U_j, the interval $[a, c + \epsilon/2]$ is covered by finitely many of the open sets U_i. That is, $c + \epsilon/2$ is in the set A, which contradicts the fact that $c = \sup A$. This contradiction shows that in fact c must be equal to b.

Thus, we have $b = \sup A$. We want to show that b itself is in A, to complete the proof. This is a similar argument to the previous paragraph. Namely, since b is in $[a, b]$, b must belong to some U_j. Since U_j is open in $[a, b]$, it must contain $(b - \epsilon, b]$ for some $\epsilon > 0$. Since $b = \sup A$, the definition of A shows that the interval $[a, b - \epsilon/2]$ is covered by finitely many of the sets U_i. Using those open sets together with U_j, it follows that the whole interval $[a, b]$ is covered by finitely many of the sets U_i. QED

Now that we have the basic example of a compact space, we prove some general properties of compact spaces. These will lead fairly easily to the basic properties of continuous real-valued functions on a compact space.

Lemma 4.3 *Any compact subspace of a metric space is bounded.*

For example, this proves again that the real line is noncompact, since it is an unbounded metric space.

Proof. It suffices to show that every compact metric space X is bounded. This is clear if X is empty. Otherwise, let x be a point in X. Consider the open balls $B_n(x)$, where n runs over the positive integers. They form an open cover of X, since for every point y in X there is a positive integer n such that $d(x, y) < n$. Since we assume X is compact, this open cover has a finite subcover. That is, X is the union of finitely many of the open balls $B_n(x)$. Taking the largest of these integers n, we conclude that X is equal to $B_n(x)$. That is, X is bounded. QED

Lemma 4.4 *Any compact subspace S of a Hausdorff topological space X is closed in X.*

For example, this proves again that the open interval $(0, 1)$ is noncompact, since it is contained in the Hausdorff space $\mathbf{R}$ as a subset which is not closed. More generally, any metric space is Hausdorff. So Lemmas 4.3 and 4.4 imply that any compact subspace of a metric space must be closed and bounded.

Proof of Lemma 4.4. Let S be a compact subspace of a Hausdorff space X. Suppose that S is not closed in X; we want to derive a contradiction. Since S is not closed in X, we can choose a point x in the closure $\overline{S}$ which is not in S.

For every point s in S, we know that s is not equal to x. Since X is Hausdorff, we can choose open subsets U_s and V_s of X such that s is in U_s, x is in V_s, and U_s and V_s are disjoint. Choosing such subsets for every point s in S, we see that the sets $U_s \cap S$ (one for each point s in S) form an open cover of S, using the subspace topology on S. Since we assume S is a compact topological space, it follows that S is the union of the open sets $U_{s_1} \cap S, \ldots, U_{s_n} \cap S$ for some finite collection of points $s_1, \ldots, s_n$ in S. Equivalently, S is covered by the open sets $U_{s_1}, \ldots, U_{s_n}$ in X.

Let V be the intersection of the other open sets $V_{s_1}, \ldots, V_{s_n}$ in X. Since this is a finite intersection of open sets, V is again open in X, and it clearly contains the point x. Also, V is disjoint from each of the open sets $U_{s_1}, \ldots, U_{s_n}$, and so V is disjoint from S. But this contradicts the fact that x is in the closure of S. (Indeed, $X - V$ is a closed subset of X which contains S, and so it contains the closure $\overline{S}$ and hence the point x. This contradicts the fact that x is in V.) So in fact S must be closed. QED

4.2 Day 11: Saturday 15 May 2004

Lemma 4.5 *Let $f : X \to Y$ be a continuous map of topological spaces. If X is compact, then $f(X)$ is compact.*

Proof. Here $f(X)$ is viewed as a topological space, using its topology as a subspace of Y. By Lemma 2.6 (2), the function $f : X \to f(X)$ is continuous. So, replacing Y by $f(X)$, we have a continuous map f from X onto Y with X compact, and we want to deduce that Y is compact.

So let U_i be any open cover of Y, where i runs over some indexing set I. We want to show that Y is covered by finitely many of the sets U_i. Since f is continuous, the sets $f^{-1}(U_i)$ are open. Clearly the union of these sets is all of X (since f of any point in X belongs to some set U_i). Since X is compact, X is in fact the union of only finitely many of these open sets, say $f^{-1}(U_1), \ldots, f^{-1}(U_n)$. That is, for every point x in X, $f(x)$ belongs to one of the sets $U_1, \ldots, U_n$. Since f is onto, this means that Y is covered by the finitely many sets $U_1, \ldots, U_n$. QED

We now have a powerful set of tools for studying continuous functions on compact spaces. In particular, we can draw the following conclusion.

Corollary 4.6 *Let X be a compact topological space. Then any continuous real-valued function $f : X \to \mathbf{R}$ is bounded. If X is nonempty, then f also attains its maximum and minimum on X.*

Proof. Since X is compact, the image $f(X) \subset \mathbf{R}$ is compact (Lemma 4.5). Therefore $f(X)$ is a bounded subset of $\mathbf{R}$ (Lemma 4.3). We rephrase this by saying that f is bounded.

If X is nonempty, then we can define the greatest lower bound and least upper bound of the nonempty bounded set $f(X)$, $a = \inf f(X)$ and $b = \sup f(X)$, by the completeness property of the real numbers. Also, the compact subspace $f(X)$ is closed in $\mathbf{R}$ by Lemma 4.4. It is a general fact (which you can check) that for any nonempty bounded subset A of the real line, $\inf A$ and $\sup A$ belong to the closure of A. Therefore, since $f(X)$ is closed, a and b belong to $f(X)$. That is, there are points x and y in X with $f(x) = a$ and $f(y) = b$; clearly these are the minimum and maximum values of f. QED

In particular, Corollary 4.6 generalizes the following result from calculus.

Corollary 4.7 *Every continuous function $f : [a, b] \to \mathbf{R}$ is bounded and attains its maximum and minimum.*

We now show how to construct many more examples of compact spaces.

Lemma 4.8 *A closed subset of a compact topological space is compact.*

Proof. Let S be a closed subset of a compact space X. Suppose we are given an open cover of S; we want to show that this cover has a finite subcover. By definition of the subspace topology on S, we can write the open cover of S as $U_i \cap S$, where i runs over some indexing set I and each U_i is an open subset of the bigger space X.

Since S is closed in X, the complement $X - S$ is open in X. Therefore $X - S$ together with the open sets U_i gives an open cover of X. Since X is compact, X is covered by finitely many of these open sets. In particular, S is covered by finitely many of these sets. The set $X - S$ is irrelevant for S, and so we have shown that S is covered by finitely many of the sets U_i. QED

Example. Let X be the subset of $\mathbf{R}$ of numbers $1/n$, where n runs over the positive integers, together with the point 0. Then X is a closed subset of $[0, 1]$, as you can check (for example using Lemma 3.18). Therefore X is a compact topological space by Lemma 4.8. You can imagine even more complicated closed subsets of $[0, 1]$. It seems that there is no hope of giving a simple 'list' of all compact subspaces of the real line, as we were able to do for connected subspaces of the real line (which are all intervals).

At least there is a clear characterization of which subsets of the real line are compact. Namely, a subset $S \subset \mathbf{R}$ is compact if and only if it is closed in $\mathbf{R}$ and bounded. We already know that a compact subset of $\mathbf{R}$ must be closed and bounded (Lemmas 4.3 and 4.4). Conversely, a closed bounded subset of $\mathbf{R}$ is a closed subset of some interval $[a, b]$, and hence is compact.

Theorem 4.9 *The product of two compact topological spaces is compact.*

Proof. Let X and Y be compact spaces. Let U_i, $i \in I$, be an open cover of the product space $X \times Y$; we want to show that this open cover has a finite subcover.

Pick any point x in X. The subspace $\{x\} \times Y$ of $X \times Y$ is homeomorphic to Y, since you can check that the obvious maps from Y to this subspace and back are both continuous. Therefore this subspace is compact.

For every point y in Y, the point (x, y) must belong to some subset U_i. By definition of the product topology, since U_i is open in $X \times Y$, there are open subsets B_y in X and C_y in Y such that x is in B_y, y is in C_y, and $B_y \times C_y$ is contained in U_i. Choose such subsets B_y and C_y for every point y in Y.

Then the open subsets $B_y \times C_y$, as y runs through all the points in Y, form an open covering of $\{x\} \times Y$. Since this subspace is compact, it is in fact covered by the open subsets $B_y \times C_y$ for y running through some finite subset S of Y. Let A_x be the intersection of the finitely many open subsets $B_y \subset X$ with $y \in S$. Then A_x is an open subset of X which contains x (one says: A_x is an *open neighbourhood* of x). We see that $\{x\} \times Y$ is covered by the products $A_x \times C_y$ for y in the finite set S. It follows that the whole 'strip' $A_x \times Y$ is covered by these finitely many products $A_x \times C_y$. But this product is contained in $B_y \times C_y$, which in turn in contained in some set U_i. Thus, we have shown: (*) For every point x in X, there is an open neighbourhood A_x of x such that the strip $A_x \times Y$ is covered by finitely many of the sets U_i.

For each point x in X, choose an open neighbourhood A_x with the property listed in (*). Then these open sets A_x form an open covering of X. Since X is compact, X is covered by only finitely many of the sets A_x. It follows that the whole space $X \times Y$ is covered by finitely many strips $A_x \times Y$, each of which is covered by finitely many of the sets U_i. So $X \times Y$ is covered by finitely many of the sets U_i. QED

By induction on n, it follows that the n-dimensional cube $[0, 1]^n$ is compact, for any positive integer n.

Corollary 4.10 *A subset of $\mathbf{R}^n$ is compact, in the subspace topology, if and only if it is closed in $\mathbf{R}^n$ and bounded.*

Proof. The proof as the same as the one we gave earlier for subsets of the real line. If S is a compact subspace of $\mathbf{R}^n$, then it must be closed and bounded, by Lemmas 4.3 and 4.4 (which hold in any metric space). Conversely, if S is a closed bounded subset of $\mathbf{R}^n$, then it is a closed subset of the cube $[a, b]^n$ for some real numbers $a < b$. So S is a closed subset of a compact space, and hence is compact. QED

Thus we have a huge supply of interesting compact spaces. For example, the surface of the sphere, $\{(x,y,z) \in \mathbf{R}^3 : x^2+y^2+z^2=1\}$, is closed in $\mathbf{R}^3$ and bounded, and so it is compact. This has valuable consequences: any continuous real-valued function on the sphere (for example) is bounded and attains its maximum and minimum.

4.3 Day 12: Tuesday 18 May 2004

Definition 4.11 *A metric space X is sequentially compact if every sequence $x_1, x_2, \ldots$ in X has a convergent subsequence.*

Here a subsequence means the sequence $x_{i_1}, x_{i_2} \ldots$ in X associated to a sequence of positive integers $i_1 < i_2 < \cdots$.

Lemma 4.12 *Every compact metric space X is sequentially compact.*

Thus we know that all the compact subspaces of $\mathbf{R}^n$, constructed at the end of the last section, are sequentially compact. That is, every sequence has a convergent subsequence, which is a useful property of these spaces. In fact, the converse to Lemma 4.12 is also true: a sequentially compact metric space is compact.

Proof of Lemma 4.12. Let $x_1, x_2, \ldots$ be a sequence in a compact metric space X. Suppose that this sequence has no convergent subsequence; we will derive a contradiction. Let S be the set of points x_i, viewed as a subset of X. If S is finite, then the conclusion is easy: there must be infinitely many numbers i such that x_i is equal to some point of S, and this constant subsequence is certainly convergent.

So we can assume that S is infinite. In order to use that X is compact, we have to construct a suitable open cover of X. We note that for every point x in X, there is an $\epsilon > 0$ such that the open ball $B_\epsilon(x)$ contains no points of S, or only the point $\{x\}$ if x happens to be in S. Indeed, if there were no such ϵ, then we could find a subsequence of the sequence x_i converging to x. So use these open balls $B_\epsilon(x)$, for all points of x, as an open covering of X. Since X is compact, it is covered by finitely many of these open balls. Since each of these open balls contains only 0 or 1 point of S, it follows that S is finite, which is a contradiction. So in fact every sequence in X has a convergent subsequence. QED

Finally, we turn to some other remarkable properties of compact topological spaces. First, there is the following inverse function theorem.

Lemma 4.13 *Every bijective continuous map f from a compact space X to a Hausdorff space Y is a homeomorphism. Equivalently, the inverse function $Y \to X$ is also continuous.*

Let me point out how common it is, outside the setting of the lemma, to have bijective continuous maps which are not homeomorphisms (that is, such that the inverse function is not continuous). This is common whenever you consider two different topologies on the same set. For example, there is a continuous bijective map (the identity map) from the set $\mathbf{R}$ with the discrete topology to the set $\mathbf{R}$ with its standard topology. It is not a homeomorphism, because there are many open sets in the first topology which are not open for the second topology. (Or, more geometrically: the inverse function cannot possibly be continuous, because it maps the connected space $\mathbf{R}$ onto the space $(\mathbf{R}, \text{discrete})$ whose connected components are all points.) Similarly, there is a continuous bijective map from the set $\{0,1\}$ with the discrete topology to the same set with the indiscrete topology, and this is not a homeomorphism.

One can also give examples with a more geometric flavour. For example, there is a continuous bijective function from $[0,1] \cup (2,3]$ to $[0,2]$, defined by subtracting 1 on the component $(2,3]$. Again, you can check that the inverse function is not continuous. Lemma 4.13 shows the surprising fact that such things cannot happen when the domain space is compact (and the image space is Hausdorff).

Proof of Lemma 4.13. Let f be a continuous bijective map from a compact space X to a Hausdorff space Y. We use the interpretation of continuity in terms of closed sets: it means that for every closed subset B of Y, the inverse image $f^{-1}(B)$ is closed. To show that f^{-1} is continuous, we have to show that for every closed subset A of X, the image $f(A)$ is closed.

Since X is compact, any closed subset A of X is compact. Therefore the image $f(A)$ is a compact subspace of Y. Since Y is Hausdorff, a compact subspace is closed in Y. QED

Example. Using Lemma 4.13, we can make precise the vague idea that any closed curve in the plane which does not cross itself must be homeomorphic to the circle S^1. Namely, let $f : S^1 \to \mathbf{R}^2$ be any continuous injective function. Then Lemma 4.13 shows that the image $f(S^1)$ is in fact homeomorphic to the circle.

There is a generalization of Lemma 4.13 using the notion of quotient spaces.

Definition 4.14 *An identification map or quotient map is a surjective function $f : X \to Y$ between topological spaces such that a subset U of Y is open if and only if $f^{-1}(U)$ is open in X. We say that Y is a quotient space of X.*

The point is that if Y is a quotient space of X, then the topology on Y is completely determined by the topology of X (as the definition makes clear). Examples Sheet 1, Question 24 shows how to construct a quotient space Y starting from any partition of a topological space X.

Notice that a bijective map is a quotient map if and only if it is a homeomorphism. Therefore, the following result generalizes Lemma 4.13

Lemma 4.15 *Every continuous map f from a compact space X onto a Hausdorff space Y is a quotient map.*

Proof. We have to show that a subset U of Y is open if and only if $f^{-1}(U)$ is open in X. Equivalently, we have to show that a subset A of Y is closed if and only if $f^{-1}(A)$ is closed in X. Since f is continuous, we know that if A is closed in Y, then $f^{-1}(A)$ is closed in X. Conversely, let A be a subset of Y such that $f^{-1}(A)$ is closed in X. Then $f^{-1}(A)$ is compact since X is compact. So the image $f(f^{-1}(A))$ is a compact subspace of Y. Since f is onto, this image is all of A. Finally, since Y is Hausdorff, it follows that A is closed in Y. QED

Example. Show that the space obtained by identifying the points 0 and 1 in $[0, 1]$ is homeomorphic to the circle $S^1 = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$.

Answer: We could check this from the definition of the quotient topology, but it is easier to use Lemma 4.15. Namely, consider the function $f : [0, 1] \to S^1$ defined by

$$f(t) = (\cos 2\pi t, \sin 2\pi t).$$

Clearly f is continuous and surjective. Since $[0, 1]$ is compact and S^1 is Hausdorff, this is a quotient map by Lemma 4.15. Also, this map identifies the two points 0 and 1 and is otherwise injective. So S^1 is the quotient space obtained by identifying 0 and 1 in $[0, 1]$.