

1.10: Metric Spaces

Basic Theory

Most of the important topological spaces that occur in applications (like probability) have an additional structure that gives a *distance* between points in the space.

Definitions

A *metric space* consists of a nonempty set S and a function $d : S \times S \rightarrow [0, \infty)$ that satisfies the following axioms: For $x, y, z \in S$,

1. $d(x, y) = 0$ if and only if $x = y$.
2. $d(x, y) = d(y, x)$.
3. $d(x, z) \leq d(x, y) + d(y, z)$.

The function d is known as a *metric* or a *distance function*.

So as the name suggests, $d(x, y)$ is the *distance* between points $x, y \in S$. The axioms are intended to capture the essential properties of distance from geometry. Part (a) is the *positive property*; the distance is strictly positive if and only if the points are distinct. Part (b) is the *symmetric property*; the distance from x to y is the same as the distance from y to x . Part (c) is the *triangle inequality*; going from x to z cannot be longer than going from x to z by way of a third point y .

Note that if (S, d) is a metric space, and A is a nonempty subset of S , then the set A with d restricted to $A \times A$ is also a metric space (known as a *subspace*). The next definitions also come naturally from geometry:

Suppose that (S, d) is a metric space, and that $x \in S$ and $r \in (0, \infty)$.

1. $B(x, r) = \{y \in S : d(x, y) < r\}$ is the *open ball* with center x and radius r .
2. $C(x, r) = \{y \in S : d(x, y) \leq r\}$ is the *closed ball* with center x and radius r .

A metric on a space induces a topology on the space in a natural way.

Suppose that (S, d) is a metric space. By definition, a set $U \subseteq S$ is *open* if for every $x \in U$ there exists $r \in (0, \infty)$ such that $B(x, r) \subseteq U$. The collection \mathcal{S}_d of open subsets of S is a topology.

Proof

1. Trivially S is open and vacuously \emptyset is open.
2. Suppose that A_i is open for i in an arbitrary index set I , and let $A = \bigcup_{i \in I} A_i$. If $x \in A$ then $x \in A_i$ for some $i \in I$. Since A_i is open, there exists $r \in (0, \infty)$ with $B(x, r) \subseteq A_i$. But then $B(x, r) \subseteq A$ so A is open.
3. Suppose that A_i is open for i in a finite index set I , and let $A = \bigcap_{i \in I} A_i$. If $x \in A$ then $x \in A_i$ for every $i \in I$. Hence for each $i \in I$ there exist $r_i \in (0, \infty)$ such that $B(x, r_i) \subseteq A_i$. Let $r = \min\{r_i : i \in I\}$. Since I is finite, $r > 0$ and $B(x, r) \subseteq B(x, r_i) \subseteq A_i$ for each $i \in I$. Hence $B(x, r) \subseteq A$, so A is open.

As the names suggests, an open ball is in fact *open* and a closed ball is in fact *closed*.

Suppose again that (S, d) is a metric space, and that $x \in S$ and $r \in (0, \infty)$. Then

1. $B(x, r)$ is open.
2. $C(x, r)$ is closed.

Proof

1. Let $y \in B(x, r)$, and let $a = d(x, y)$, so that $a < r$. If $z \in B(y, r - a)$ then we have $d(x, y) = a$ and $d(y, z) < r - a$, so by the triangle inequality, $d(x, z) < a + (r - a) = r$. Hence $z \in B(x, r)$. Thus $B(y, r - a) \subseteq B(x, r)$. It follows that $B(x, r)$ is open.
2. We show that $U = [C(x, r)]^c$ is open. Suppose that $y \in U$, and let $a = d(x, y)$, so that $a > r$. Let $z \in B(y, a - r)$ and suppose that $z \in C(x, r)$, so that $d(z, x) \leq r$. By the triangle inequality again,

$$d(x, y) \leq d(x, z) + d(z, y) < r + (a - r) = a \quad (1.10.1)$$

a contradiction. Hence $z \in U$. So $B(y, a-r) \subseteq U$.

Recall that for a general topological space, a *neighborhood* of a point $x \in S$ is a set $A \subseteq S$ with the property that there exists an open set U with $x \in U \subseteq A$. It follows that in a metric space, $A \subseteq S$ is a neighborhood of x if and only if there exists $r > 0$ such that $B(x, r) \subseteq A$. In words, a neighborhood of a point must contain an open ball about that point.

It's easy to construct new metrics from ones that we already have. Here's one such result.

Suppose that S is a nonempty set, and that d, e are metrics on S , and $c \in (0, \infty)$. Then the following are also metrics on S :

1. cd
2. $d+e$

Proof

1. Recall that cd is the function defined by $(cd)(x, y) = cd(x, y)$ for $(x, y) \in S^2$. Since $c > 0$, it's easy to see that the axioms are satisfied.
2. Recall that $d+e$ is the function defined by $(d+e)(x, y) = d(x, y) + e(x, y)$ for $(x, y) \in S^2$. Again, it's easy to see that the axioms are satisfied.

Since a metric space produces a topological space, all of the definitions for general topological spaces apply to metric spaces as well. In particular, in a metric space, distinct points can always be separated.

A metric space (S, d) is a Hausdorff space.

Proof

Let x, y be distinct points in S . Then $r = d(x, y) > 0$. The sets $B(x, r/2)$ and $B(y, r/2)$ are open, and contain x and y , respectively. Suppose that $z \in B(x, r/2) \cap B(y, r/2)$. By the triangle inequality,

$$d(x, y) \leq d(x, z) + d(z, y) < \frac{r}{2} + \frac{r}{2} = r \quad (1.10.2)$$

a contradiction. Hence $B(x, r/2)$ and $B(y, r/2)$ are disjoint.

Metrizable Spaces

Again, every metric space is a topological space, but not conversely. A non-Hausdorff space, for example, cannot correspond to a metric space. We know there are such spaces; a set S with more than one point, and with the trivial topology $\mathcal{S} = \{S, \emptyset\}$ is non-Hausdorff.

Suppose that (S, \mathcal{S}) is a topological space. If there exists a metric d on S such that $\mathcal{S} = \mathcal{S}_d$, then (S, \mathcal{S}) is said to be *metrizable*.

It's easy to see that different metrics can induce the same topology. For example, if d is a metric and $c \in (0, \infty)$, then the metrics d and cd induce the same topology.

Let S be a nonempty set. Metrics d and e on S are *equivalent*, and we write $d \equiv e$, if $\mathcal{S}_d = \mathcal{S}_e$. The relation \equiv is an equivalence relation on the collection of metrics on S . That is, for metrics d, e, f on S ,

1. $d \equiv d$, the *reflexive property*.
2. If $d \equiv e$ then $e \equiv d$, the *symmetric property*.
3. If $d \equiv e$ and $e \equiv f$ then $d \equiv f$, the *transitive property*.

There is a simple condition that characterizes when the topology of one metric is finer than the topology of another metric, and then this in turn leads to a condition for equivalence of metrics.

Suppose again that S is a nonempty set and that d, e are metrics on S . Then \mathcal{S}_e is finer than \mathcal{S}_d if and only if every open ball relative to d contains an open ball relative to e .

Proof

Suppose that $\mathcal{S}_d \subseteq \mathcal{S}_e$ so that \mathcal{S}_e is finer than \mathcal{S}_d . If $x \in S$ and $a \in (0, \infty)$, then the open ball $B_d(x, a)$ centered at x of radius a for the metric d is in \mathcal{S}_d and hence in \mathcal{S}_e . Thus there exists $b \in (0, \infty)$ such that $B_e(x, b) \subseteq B_d(x, a)$. Conversely, suppose that the condition in the theorem holds and suppose that $U \in \mathcal{S}_d$. If $x \in U$ there exists $a \in (0, \infty)$ such that $B_d(x, a) \subseteq U$. Hence there exists $b \in (0, \infty)$ such that $B_e(x, b) \subseteq B_d(x, a) \subseteq U$. So $U \in \mathcal{S}_e$.

It follows that metrics d and e on S are equivalent if and only if every open ball relative to one of the metrics contains an open ball relative to the other metric.

So every metrizable topology on S corresponds to an equivalence class of metrics that produce that topology. Sometimes we want to know that a topological space is metrizable, because of the nice properties that it will have, but we don't really need to use a specific metric that generates the topology. At any rate, it's important to have conditions that are sufficient for a topological space to be metrizable. The most famous such result is the *Urysohn metrization theorem*, named for the Russian mathematician Pavel Uryshon:

Suppose that (S, \mathcal{S}) is a regular, second-countable, Hausdorff space. Then (S, \mathcal{S}) is metrizable.

Review of the terms

Recall that *regular* means that every closed set and point not in the set can be separated by disjoint open sets. As discussed earlier, *Hausdorff* means that any two distinct points can be separated by disjoint open sets. Finally, *second-countable* means that there is a countable base for the topology, that is, there is a countable collection of open sets with the property that every other open set is a union of sets in the collection.

Convergence

With a distance function, the convergence of a sequence can be characterized in a manner that is just like calculus. Recall that for a general topological space (S, \mathcal{S}) , if $(x_n : n \in \mathbb{N}_+)$ is a sequence of points in S and $x \in S$, then $x_n \rightarrow x$ as $n \rightarrow \infty$ means that for every neighborhood U of x , there exists $m \in \mathbb{N}_+$ such that $x_n \in U$ for $n > m$.

Suppose that (S, d) is a metric space, and that $(x_n : n \in \mathbb{N}_+)$ is a sequence of points in S and $x \in S$. Then $x_n \rightarrow x$ as $n \rightarrow \infty$ if and only if for every $\epsilon > 0$ there exists $m \in \mathbb{N}_+$ such that if $n > m$ then $d(x_n, x) < \epsilon$. Equivalently, $x_n \rightarrow x$ as $n \rightarrow \infty$ if and only if $d(x_n, x) \rightarrow 0$ as $n \rightarrow \infty$ (in the usual calculus sense).

Proof

Suppose that $x_n \rightarrow x$ as $n \rightarrow \infty$, and let $\epsilon > 0$. Then $B(x, \epsilon)$ is a neighborhood of x , so there exists $m \in \mathbb{N}_+$ such that $x_n \in B(x, \epsilon)$ for $n > m$, which is the condition in the theorem. Conversely, suppose that condition in the theorem holds, and let U be a neighborhood of x . Then there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$. By assumption, there exists $m \in \mathbb{N}_+$ such that if $n > m$ then $x_n \in B(x, \epsilon) \subseteq U$.

So, no matter how tiny $\epsilon > 0$ may be, all but finitely many terms of the sequence are within ϵ distance of x . As one might hope, limits are unique.

Suppose again that (S, d) is a metric space. Suppose also that $(x_n : n \in \mathbb{N}_+)$ is a sequence of points in S and that $x, y \in S$. If $x_n \rightarrow x$ as $n \rightarrow \infty$ and $x_n \rightarrow y$ as $n \rightarrow \infty$ then $x = y$.

Proof

This follows immediately since a metric space is a Hausdorff space, and the limit of a sequence in a Hausdorff space is unique. Here's a direct proof: Let $\epsilon > 0$. Then there exists $k \in \mathbb{N}_+$ such that $d(x_n, x) < \epsilon/2$ for $n > k$, and there exists $m \in \mathbb{N}_+$ such that $d(x_n, y) < \epsilon/2$ for $n > m$. Let $n > \max\{k, m\}$. By the triangle inequality,

$$d(x, y) \leq d(x, x_n) + d(x_n, y) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad (1.10.3)$$

So we have $d(x, y) < \epsilon$ for every $\epsilon > 0$ and hence $d(x, y) = 0$ and thus $x = y$.

Convergence of a sequence is a topological property, and so is preserved under equivalence of metrics.

Suppose that d, e are equivalent metrics on S , and that $(x_n : n \in \mathbb{N}_+)$ is a sequence of points in S and $x \in S$. Then $x_n \rightarrow x$ as $n \rightarrow \infty$ relative to d if and only if $x_n \rightarrow x$ as $n \rightarrow \infty$ relative to e .

Closed subsets of a metric space have a simple characterization in terms of convergent sequences, and this characterization is more intuitive than the abstract axioms in a general topological space.

Suppose again that (S, d) is a metric space. Then $A \subseteq S$ is closed if and only if whenever a sequence of points in A converges, the limit is also in A .

Proof

Suppose that A is closed and that $(x_n : n \in \mathbb{N}_+)$ is a sequence of points in A with $x_n \rightarrow x \in S$ as $n \rightarrow \infty$. Suppose that $x \in A^c$. Since A^c is open, $x_n \in A^c$ for n sufficiently large, a contradiction. Hence $x \in A$. Conversely, suppose that A has the sequential closure property, but that A is not closed. Then A^c is not open. This means that there exists $x \in A^c$ with the property that every neighborhood of x has points in A . Specifically, for each $n \in \mathbb{N}_+$ there exists $x_n \in B(x, 1/n)$ with $x_n \in A$. But clearly $x_n \rightarrow x$ as $n \rightarrow \infty$, again a contradiction.

The following definition also shows up in standard calculus. The idea is to have a criterion for convergence of a sequence that does not require knowing *a-priori* the limit. But for metric spaces, this definition takes on added importance.

Suppose again that (S, d) is a metric space. A sequence of points $(x_n : n \in \mathbb{N}_+)$ in S is a *Cauchy sequence* if for every $\epsilon > 0$ there exist $k \in \mathbb{N}_+$ such that if $m, n \in \mathbb{N}_+$ with $m > k$ and $n > k$ then $d(x_m, x_n) < \epsilon$.

Cauchy sequences are named for the ubiquitous Augustin Cauchy. So for a Cauchy sequence, no matter how tiny $\epsilon > 0$ may be, all but finitely many terms of the sequence will be within ϵ distance of each other. A convergent sequence is always Cauchy.

Suppose again that (S, d) is a metric space. If a sequence of points $(x_n : n \in \mathbb{N}_+)$ in S converges, then the sequence is Cauchy.

Proof

By assumption, there exists $x \in S$ such that $x_n \rightarrow x$ as $n \rightarrow \infty$. Let $\epsilon > 0$. There exists $k \in \mathbb{N}_+$ such that if $n \in \mathbb{N}_+$ and $n > k$ then $d(x_n, x) < \epsilon/2$. Hence if $m, n \in \mathbb{N}_+$ with $m > k$ and $n > k$ then by the triangle inequality,

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad (1.10.4)$$

So the sequence is Cauchy.

Conversely, one might think that a Cauchy sequence should converge, but it's relatively trivial to create a situation where this is false. Suppose that (S, d) is a metric space, and that there is a point $x \in S$ that is the limit of a sequence of points in S that are all distinct from x . Then the space $T = S - \{x\}$ with the metric d restricted to $T \times T$ has a Cauchy sequence that does not converge. Essentially, we have created a “convergence hole”. So our next definition is very natural and very important.

Suppose again that (S, d) is metric space and that $A \subseteq S$. Then A is *complete* if every Cauchy sequence in A converges to a point in A .

Of course, completeness can be applied to the entire space S . Trivially, a complete set must be closed.

Suppose again that (S, d) is a metric space, and that $A \subseteq S$. If A is complete, then A is closed.

Proof

Suppose that $\mathbf{x} = (x_n : n \in \mathbb{N})$ is a sequence of points in A and that $x_n \rightarrow x \in S$ as $n \rightarrow \infty$. Then \mathbf{x} is a Cauchy sequence, and so by completeness, $x \in A$. Hence A is closed by (12).

Completeness is such a crucial property that it is often imposed as an assumption on metric spaces that occur in applications. Even though a Cauchy sequence may not converge, here is a partial result that will be useful later: if a Cauchy sequence has a convergent subsequence, then the sequence itself converges.

Suppose again the (S, d) is a metric space, and that $(x_n : n \in \mathbb{N}_+)$ is a Cauchy sequence in S . If there exists a subsequence $(x_{n_k} : k \in \mathbb{N}_+)$ such that $x_{n_k} \rightarrow x \in S$ as $k \rightarrow \infty$, then $x_n \rightarrow x$ as $n \rightarrow \infty$.

Proof

Recall that in the construction of a subsequence, the indices $(n_k : k \in \mathbb{N}_+)$ must be a strictly increasing sequence in \mathbb{N}_+ . In particular, $n_k \rightarrow \infty$ as $k \rightarrow \infty$. So let $\epsilon > 0$. From the hypotheses, there exists $j \in \mathbb{N}_+$ such that if $k > j$ then $d(x_{n_k}, x) < \epsilon/2$. There exists $N \in \mathbb{N}_+$ such that if $m > N$ and $p > N$ then $d(x_m, x_p) < \epsilon/2$. Now let $m > N$. Pick $k \in \mathbb{N}_+$ such that $k > j$ and $n_k > N$. By the triangle inequality,

$$d(x_m, x) \leq d(x_m, x_{n_k}) + d(x_{n_k}, x) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad (1.10.5)$$

Continuity

In metric spaces, continuity of functions also has simple characterizations in terms of that are familiar from calculus. We start with local continuity. Recall that the general topological definition is that $f : S \rightarrow T$ is continuous at $x \in S$ if $f^{-1}(V)$ is a neighborhood of x in S for every open set V of $f(x)$ in T .

Suppose that (S, d) and (T, e) are metric spaces, and that $f : S \rightarrow T$. The continuity of f at $x \in S$ is equivalent to each of the following conditions:

1. If $(x_n : n \in \mathbb{N}_+)$ is a sequence in S with $x_n \rightarrow x$ as $n \rightarrow \infty$ then $f(x_n) \rightarrow f(x)$ as $n \rightarrow \infty$.
2. For every $\epsilon > 0$, there exists $\delta > 0$ such that if $y \in S$ and $d(x, y) < \delta$ then $e[f(y) - f(x)] < \epsilon$.

Proof

1. This condition is *sequential continuity* at x . Continuity at x implies sequential continuity at x for general topological spaces, and hence for metric spaces. Conversely, suppose that sequential continuity holds at $x \in S$, and let V be a neighborhood of $f(x)$ in T . If $U = f^{-1}(V)$ is not a neighborhood of x in S , then for every $n \in \mathbb{N}_+$, there exists $x_n \in B(x, 1/n)$ with $x_n \notin U$. But then clearly $x_n \rightarrow x$ as $n \rightarrow \infty$ but $f(x_n)$ does not converge to $f(x)$ as $n \rightarrow \infty$, a contradiction.
2. Suppose that f is continuous at x . For $\epsilon > 0$, $B_T[f(x), \epsilon]$ is a neighborhood of $f(x)$, and hence $U = f^{-1}(B_T[f(x), \epsilon])$ is a neighborhood of x . Hence there exists $\delta > 0$ such that $B_S(x, \delta) \subseteq U$. But this means that if $d(y, x) < \delta$ then $e[f(y), f(x)] < \epsilon$. Conversely suppose that the condition in (b) holds, and suppose that V is a neighborhood of $f(x)$. Then there exists $\epsilon > 0$ such that $B_T[f(x), \epsilon] \subseteq V$. By assumption, there exists $\delta > 0$ such that if $y \in B_S(x, \delta)$ then $f(y) \in B_T[f(x), \epsilon] \subseteq V$. This means that $f^{-1}(V)$ is a neighborhood of x .

More generally, recall that f continuous on $A \subseteq S$ means that f is continuous at each $x \in A$, and that f continuous means that f is continuous on S . So general continuity can be characterized in terms of sequential continuity and the ϵ - δ condition.

On a metric space, there are stronger versions of continuity.

Suppose again that (S, d) and (T, e) are metric spaces and that $f : S \rightarrow T$. Then f is *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that if $x, y \in S$ with $d(x, y) < \delta$ then $e[f(x), f(y)] < \epsilon$.

In the ϵ - δ formulation of ordinary point-wise continuity above, δ depends on the point x in addition to ϵ . With uniform continuity, there exists a δ depending only on ϵ that works *uniformly* in $x \in S$.

Suppose again that (S, d) and (T, e) are metric spaces, and that $f : S \rightarrow T$. If f is uniformly continuous then f is continuous.

Here is an even stronger version of continuity.

Suppose again that (S, d) and (T, e) are metric spaces, and that $f : S \rightarrow T$. Then f is *Hölder continuous* with exponent $\alpha \in (0, \infty)$ if there exists $C \in (0, \infty)$ such that $e[f(x), f(y)] \leq C[d(x, y)]^\alpha$ for all $x, y \in S$.

The definition is named for Otto Höder. The exponent α is more important than the constant C , which generally does not have a name. If $\alpha = 1$, f is said to be *Lipschitz continuous*, named for the German mathematician Rudolf Lipschitz.

Suppose again that (S, d) and (T, e) are metric spaces, and that $f : S \rightarrow T$. If f is Hölder continuous with exponent $\alpha > 0$ then f is uniformly continuous.

The case where $\alpha = 1$ and $C < 1$ is particularly important.

Suppose again that (S, d) and (T, e) are metric spaces. A function $f : S \rightarrow T$ is a *contraction* if there exists $C \in (0, 1)$ such that

$$e[f(x), f(y)] \leq Cd(x, y), \quad x, y \in S \quad (1.10.6)$$

So contractions shrink distance. By the result above, a contraction is uniformly continuous. Part of the importance of contraction maps is due to the famous *Banach fixed-point theorem*, named for Stefan Banach.

Suppose that (S, d) is a complete metric space and that $f : S \rightarrow S$ is a contraction. Then f has a unique *fixed point*. That is, there exists exactly one $x^* \in S$ with $f(x^*) = x^*$. Let $x_0 \in S$, and recursively define $x_n = f(x_{n-1})$ for $n \in \mathbb{N}_+$. Then $x_n \rightarrow x^*$ as $n \rightarrow \infty$.

Functions that preserve distance are particularly important. The term *isometry* means distance-preserving.

Suppose again that (S, d) and (T, e) are metric spaces, and that $f : S \rightarrow T$. Then f is an *isometry* if $e[f(x), f(y)] = d(x, y)$ for every $x, y \in S$.

Suppose again that (S, d) and (T, e) are metric spaces, and that $f : S \rightarrow T$. If f is an isometry, then f is one-to-one and Lipschitz continuous.

Proof

If $x, y \in S$ with $x \neq y$, then $e[f(x), f(y)] = d(x, y) > 0$, so $f(x) \neq f(y)$. Hence f is one-to-one. Directly from the definition, f is Hölder continuous with exponent $\alpha = 1$ and constant multiple $C = 1$.

In particular, an isometry f is uniformly continuous. If one metric space can be mapped isometrically *onto* another metric space, the spaces are essentially the same.

Metric spaces (S, d) and (T, e) are *isometric* if there exists an isometry f that maps S onto T . Isometry is an equivalence relation on metric spaces. That is, for metric spaces (S, d) , (T, e) , and (U, ρ) ,

1. (S, d) is isometric to (S, d) , the *reflexive property*.
2. If (S, d) is isometric to (T, e) then (T, e) is isometric to (S, d) , the *symmetric property*.
3. If (S, d) is isometric to (T, e) and (T, e) is isometric to (U, ρ) , then (S, d) is isometric to (U, ρ) , the *transitive property*.

Proof

1. The identity function $I : S \rightarrow S$ defined by $I(x) = x$ for $x \in S$ is an isometry from (S, d) onto (S, d) .
2. If f is an isometry from (S, d) onto (T, e) then f^{-1} is an isometry from (T, e) onto (S, d) .
3. If f is an isometry from (S, d) onto (T, e) and g is an isometry from (T, e) onto (U, ρ) , then $g \circ f$ is an isometry from (S, d) to (U, ρ) .

In particular, if metric spaces (S, d) and (T, e) are isometric, then as topological spaces, they are homeomorphic.

Compactness and Boundedness

In a metric space, various definitions related to a set being *bounded* are natural, and are related to the general concept of compactness.

Suppose again that (S, d) is a metric space, and that $A \subseteq S$. Then A is *bounded* if there exists $r \in (0, \infty)$ such that $d(x, y) \leq r$ for all $x, y \in A$. The *diameter* of A is

$$\text{diam}(A) = \inf\{r > 0 : d(x, y) < r \text{ for all } x, y \in A\} \quad (1.10.7)$$

Additional details

Recall that $\inf(\emptyset) = \infty$, so $\text{diam}(A) = \infty$ if A is unbounded. In the bounded case, note that if the distance between points in A is bounded by $r \in (0, \infty)$, then the distance is bounded by any $s \in [r, \infty)$. Hence the diameter definition makes sense.

So A is bounded if and only if $\text{diam}(A) < \infty$. Diameter is an increasing function relative to the subset partial order.

Suppose again that (S, d) is a metric space, and that $A \subseteq B \subseteq S$. Then $\text{diam}(A) \leq \text{diam}(B)$.

Our next definition is stronger, but first let's review some terminology that we used for general topological spaces: If S is a set, A a subset of S , and \mathcal{A} a collection of subsets of S , then \mathcal{A} is said to *cover* A if $A \subseteq \bigcup \mathcal{A}$. So with this terminology, we can talk about open covers, closed covers, finite covers, disjoint covers, and so on.

Suppose again that (S, d) is a metric space, and that $A \subseteq S$. Then A is *totally bounded* if for every $r > 0$ there is a finite cover of A with open balls of radius r .

Recall that for a general topological space, a set A is *compact* if every open cover of A has a finite subcover. So in a metric space, the term *precompact* is sometimes used instead of *totally bounded*: The set A is totally bounded if every cover of A with open balls of radius r has a finite subcover.

Suppose again that (S, d) is a metric space. If $A \subseteq S$ is totally bounded then A is bounded.

Proof

There exists a finite cover of A with open balls of radius 1. Let C denote the set of centers of the balls, and let $c = \max\{d(u, v) : u, v \in C\}$, the maximum distance between two centers. Since C is finite, $c < \infty$. Now let $x, y \in A$. Since the balls cover A , there exist $u, v \in C$ with $x \in B(u, 1)$ and $y \in B(v, 1)$. By the triangle inequality (what else?)

$$d(x, y) \leq d(x, u) + d(u, v) + d(v, y) \leq 2 + c \quad (1.10.8)$$

Hence A is bounded.

Since a metric space is a Hausdorff space, a compact subset of a metric space is closed. Compactness also has a simple characterization in terms of convergence of sequences.

Suppose again that (S, d) is a metric space. A subset $C \subseteq S$ is compact if and only if every sequence of points in C has a subsequence that converges to a point in C .

Proof

The condition in the theorem is known as *sequential compactness*, so we want to show that sequential compactness is equivalent to compactness. The proof is harder than most of the others in this section, but the proof presented here is the nicest I have found, and is due to Anton Schep.

Suppose that C is compact and that $\mathbf{x} = (x_n : n \in \mathbb{N}_+)$ is a sequence of points in C . Let $A = \{x_n : n \in \mathbb{N}_+\} \subseteq C$, the unordered set of distinct points in the sequence. If A is finite, then some element of $a \in A$ must occur infinitely many times in the sequence. In this case, we can construct a subsequence of \mathbf{x} all of whose terms are a , and so this subsequence trivially converges to $a \in C$. Suppose next that A is infinite. Since the space is Hausdorff, C is closed, and therefore $\text{cl}(A) \subseteq C$. Our next claim is that there exists $a \in \text{cl}(A)$ such that for every $r > 0$, the set $A \cap B(a, r)$ is infinite. If the claim is false, then for each $a \in \text{cl}(A)$ there exists $r_a > 0$ such that $A \cap B(a, r_a)$ is finite. It then follows that for each $a \in A$, there exists $\epsilon_a > 0$ such that $A \cap B(a, \epsilon_a) = \{a\}$. But then $\mathcal{U} = \{B(a, \epsilon_a) : a \in \text{cl}(A)\} \cup \{\text{cl}(A)^c\}$ is an open cover of C that has no finite subcover, a contradiction. So the claim is true and for some $a \in \text{cl}(A)$, the set $A \cap B(a, r)$ is infinite for each $r > 0$. We can construct a subsequence of \mathbf{x} that converges to $a \in C$.

Conversely, suppose that C is sequentially compact. If $\mathbf{x} = (x_n : n \in \mathbb{N}_+)$ is a Cauchy sequence in C , then by assumption, \mathbf{x} has a subsequence that converges to some $x \in C$. But then by (17) the sequence \mathbf{x} itself converges to x , so it follows that C is complete. We next show that C is totally bounded. Our goal is to show that C can be covered by a finite number of balls of an arbitrary radius $r > 0$. Pick $x_1 \in C$. If $C \subseteq B(x_1, r)$ then we are done. Otherwise, pick $x_2 \in C \setminus B(x_1, r)$. If $C \subseteq B(x_1, r) \cup B(x_2, r)$ then again we are done. Otherwise there exists $x_3 \in C \setminus [B(x_1, r) \cup B(x_2, r)]$. This process must terminate in a finite number of steps or otherwise we would have a sequence of points $(x_n : n \in \mathbb{N}_+)$ in C with the property that $d(x_n, x_m) \geq r$ for every $n, m \in \mathbb{N}_+$. Such a sequence does not have a convergent subsequence. Suppose now that \mathcal{U} is an open cover of C and let $c = \text{diam}(C)$. Then C can be covered by a finite number of closed balls of with centers in C and with radius $c/4$. It follows that at least one of these balls cannot be covered by a finite subcover from \mathcal{U} . Let C_1 denote the intersection of this ball with C . Then C_1 is closed and is sequentially compact with $\text{diam}(C_1) \leq c/4$. Repeating the argument, we generate a nested sequence of close sets (C_1, C_2, \dots) such that $\text{diam}(C_n) \leq c/2^n$, and with the property that C_n cannot be finitely covered by \mathcal{U} for each $n \in \mathbb{N}_+$. Pick $x_n \in C_n$ for each $n \in \mathbb{N}_+$. Then $\mathbf{x} = (x_n : n \in \mathbb{N}_+)$ is a Cauchy sequence in C and hence has a subsequence that converges to some $x \in C$. Then $x \in \bigcap_{n=1}^{\infty} C_n$ and since $\text{diam}(C_n) \rightarrow 0$ as $n \rightarrow \infty$ it follows that in fact, $\bigcap_{n=1}^{\infty} C_n = \{x\}$. Now, since \mathcal{U} covers C , there exists $U \in \mathcal{U}$ such that $x \in U$. Since U is open, there exists $r > 0$

such that $B(x, r) \subseteq U$. Now let $n \in \mathbb{N}_+$ be sufficiently large that $d(x, x_n) \leq r/2$ and $\text{diam}(C_n) < r/2$. Then $C_n \subseteq B(x, r) \subseteq U$, which contradicts the fact that C_n cannot be finitely covered by \mathcal{U} .

Hausdorff Measure and Dimension

Our last discussion is somewhat advanced, but is important for the study of certain random processes, particularly Brownian motion. The idea is to measure the “size” of a set in a metric space in a topological way, and then use this measure to define a type of “dimension”. We need a preliminary definition, using our convenient *cover* terminology. If (S, d) is a metric space, $A \subseteq S$, and $\delta \in (0, \infty)$, then a *countable δ cover* of A is a countable cover \mathcal{B} of A with the property that $\text{diam}(B) < \delta$ for each $B \in \mathcal{B}$.

Suppose again that (S, d) is a metric space and that $A \subseteq S$. For $\delta \in (0, \infty)$ and $k \in [0, \infty)$, define

$$H_\delta^k(A) = \inf \left\{ \sum_{B \in \mathcal{B}} [\text{diam}(B)]^k : \mathcal{B} \text{ is a countable } \delta \text{ cover of } A \right\} \quad (1.10.9)$$

The k -dimensional *Hausdorff measure* of A is

$$H^k(A) = \sup \{ H_\delta^k(A) : \delta > 0 \} = \lim_{\delta \downarrow 0} H_\delta^k(A) \quad (1.10.10)$$

Additional details

Note that if \mathcal{B} is a countable δ cover of A then it is also a countable ϵ cover of A for every $\epsilon > \delta$. This means that $H_\delta^k(A)$ is decreasing in $\delta \in (0, \infty)$ for fixed $k \in [0, \infty)$. Hence

$$\sup \{ H_\delta^k(A) : \delta > 0 \} = \lim_{\delta \downarrow 0} H_\delta^k(A) \quad (1.10.11)$$

Note that the k -dimensional Hausdorff measure is defined for every $k \in [0, \infty)$, not just nonnegative integers. Nonetheless, the integer dimensions are interesting. The 0-dimensional measure of A is the number of points in A . In Euclidean space, which we consider in (36), the measures of dimension 1, 2, and 3 are related to length, area, and volume, respectively.

Suppose again that (S, d) is a metric space and that $A \subseteq S$. The *Hausdorff dimension* of A is

$$\dim_H(A) = \inf \{ k \in [0, \infty) : H^k(A) = 0 \} \quad (1.10.12)$$

Of special interest, as before, is the case when $S = \mathbb{R}^n$ for some $n \in \mathbb{N}_+$ and d is the standard Euclidean distance, reviewed in (36). As you might guess, the Hausdorff dimension of a point is 0, the Hausdorff dimension of a “simple curve” is 1, the Hausdorff dimension of a “simple surface” is 2, and so on. But there are also sets with fractional Hausdorff dimension, and the stochastic process Brownian motion provides some fascinating examples. The graph of standard Brownian motion has Hausdorff dimension $3/2$ while the set of zeros has Hausdorff dimension $1/2$.

Examples and Special Cases

Normed Vector Spaces

A norm on a vector space generates a metric on the space in a very simple, natural way.

Suppose that $(S, +, \cdot)$ is a vector space, and that $\| \cdot \|$ is a norm on the space. Then d defined by $d(x, y) = \|y - x\|$ for $x, y \in S$ is a metric on S .

Proof

The metric axioms follow easily from the norm axioms.

1. The positive property for d follows since $\|x\| = 0$ if and only if $x = 0$.
2. The symmetric property for d follows since $\| -x \| = \|x\|$.
3. The triangle inequality for d follows from the triangle inequality for the norm: $\|x + y\| \leq \|x\| + \|y\|$.

On \mathbb{R}^n , we have a variety of norms, and hence a variety of metrics.

For $n \in \mathbb{N}_+$ and $k \in [1, \infty)$, the function d_k given below is a metric on \mathbb{R}^n :

$$d_k(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n |x_i - y_i|^k \right)^{1/k}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n), \mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n \quad (1.10.13)$$

Proof

This follows from the general result above, since $\|\cdot\|_k$ defined below is a norm on \mathbb{R}^n :

$$\|\mathbf{x}\|_k = \left(\sum_{i=1}^n |x_i|^k \right)^{1/k}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad (1.10.14)$$

Of course the metric d_2 is *Euclidean distance*, named for Euclid of course. This is the most important one, in a practical sense because it's the usual one that we use in the real world, and in a mathematical sense because the associated norm corresponds to the standard inner product on \mathbb{R}^n given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i, \quad \mathbf{x} = (x_1, x_2, \dots, x_n), \mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n \quad (1.10.15)$$

For $n \in \mathbb{N}_+$, the function d_∞ defined below is a metric on \mathbb{R}^n :

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max\{|x_i - y_i| : i \in \{1, 2, \dots, n\}\}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad (1.10.16)$$

Proof

This follows from the general result above, since $\|\cdot\|_\infty$ defined below is a norm on \mathbb{R}^n :

$$\|\mathbf{x}\|_\infty = \max\{|x_i| : i \in \{1, 2, \dots, n\}\}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad (1.10.17)$$

To justify the notation, recall that $\|\mathbf{x}\|_k \rightarrow \|\mathbf{x}\|_\infty$ as $k \rightarrow \infty$ for $\mathbf{x} \in \mathbb{R}^n$, and hence $d_k(\mathbf{x}, \mathbf{y}) \rightarrow d_\infty(\mathbf{x}, \mathbf{y})$ as $k \rightarrow \infty$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

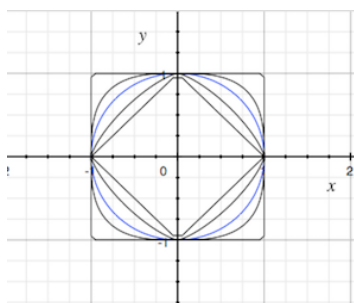


Figure 1.10.1: From inside out, the boundaries of the unit balls centered at the origin in \mathbb{R}^2 for the metrics d_k with $k \in \{1, 3/4, 2, 3, \infty\}$.

Suppose now that S is a nonempty set. Recall that the collection \mathcal{V} of all functions $f : S \rightarrow \mathbb{R}$ is a vector space under the usual pointwise definition of addition and scalar multiplication. That is, if $f, g \in \mathcal{V}$ and $c \in \mathbb{R}$, then $f + g \in \mathcal{V}$ and $cf \in \mathcal{V}$ are defined by $(f + g)(x) = f(x) + g(x)$ and $(cf)(x) = cf(x)$ for $x \in S$. Recall further that the collection \mathcal{U} of *bounded* functions $f : S \rightarrow \mathbb{R}$ is a vector subspace of \mathcal{V} , and moreover, $\|\cdot\|$ defined by $\|f\| = \sup\{|f(x)| : x \in S\}$ is a norm on \mathcal{U} , known as the *supremum norm*. It follows that \mathcal{U} is a metric space with the metric d defined by

$$d(f, g) = \|f - g\| = \sup\{|f(x) - g(x)| : x \in S\} \quad (1.10.18)$$

Vector spaces of bounded, real-valued functions, with the supremum norm are very important in the study of probability and stochastic processes. Note that the supremum norm on \mathcal{U} generalizes the [maximum norm](#) on \mathbb{R}^n , since we can think of a point in \mathbb{R}^n as a function from $\{1, 2, \dots, n\}$ into \mathbb{R} . Later, as part of our discussion on integration with respect to a positive measure, we will see how to generalize the k norms on \mathbb{R}^n to spaces of functions.

Products of Metric Spaces

If we have a finite number of metric spaces, then we can combine the individual metrics, together with an norm on the vector space \mathbb{R}^n , to create a norm on the Cartesian product space.

Suppose $n \in \{2, 3, \dots\}$, and that (S_i, d_i) is a metric space for each $i \in \{1, 2, \dots, n\}$. Suppose also that $\|\cdot\|$ is a norm on \mathbb{R}^n . Then the function d given as follows is a metric on $S = S_1 \times S_2 \times \dots \times S_n$:

$$d(\mathbf{x}, \mathbf{y}) = \|(d_1(x_1, y_1), d_2(x_2, y_2), \dots, d_n(x_n, y_n))\|, \quad \mathbf{x} = (x_1, x_2, \dots, x_n), \mathbf{y} = (y_1, y_2, \dots, y_n) \in S \quad (1.10.19)$$

Proof

1. Note that $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $d_i(x_i, y_i) = 0$ for $i \in \{1, 2, \dots, n\}$ if and only if $x_i = y_i$ for $i \in \{1, 2, \dots, n\}$ if and only if $\mathbf{x} = \mathbf{y}$.
2. Since $d_i(x_i, y_i) = d_i(y_i, x_i)$ for $i \in \{1, 2, \dots, n\}$ we have $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.
3. The triangle inequality follows from the triangle inequality for each metric, and the triangle inequality for the norm.

Graphs

Recall that a *graph* (in the combinatorial sense) consists of a countable set S of *vertices* and a set $E \subseteq S \times S$ of *edges*. In this discussion, we assume that the graph is *undirected* in the sense that $(x, y) \in E$ if and only if $(y, x) \in E$, and has no loops so that $(x, x) \notin E$ for $x \in S$. Finally, recall that a *path* of length $n \in \mathbb{N}_+$ from $x \in S$ to $y \in S$ is a sequence $(x_0, x_1, \dots, x_n) \in S^{n+1}$ such that $x_0 = x$, $x_n = y$, and $(x_{i-1}, x_i) \in E$ for $i \in \{1, 2, \dots, n\}$. The graph is *connected* if there exists a path of finite length between any two distinct vertices in S . Such a graph has a natural metric:

Suppose that $G = (S, E)$ is a connected graph. Then d defined as follows is a metric on S : $d(x, x) = 0$ for $x \in S$, and $d(x, y)$ is the length of the shortest path from x to y for distinct $x, y \in S$.

Proof

1. The positive property follows from the definition: $d(x, y) = 0$ if and only if $x = y$.
2. The symmetric property follows since the graph is undirected: $d(x, y) = d(y, x)$ for all $x, y \in S$.
3. For the triangle inequality, suppose that $x, y, z \in S$, and that $m = d(x, y)$ and $n = d(y, z)$. Then there is a path of length m from x to y and a path of length n from y to z . Concatenating the paths produces a path of length $m + n$ from x to z . But $d(x, z)$ is the length of the shortest such path, so it follows that $d(x, z) \leq m + n$.

The Discrete Topology

Suppose that S is a nonempty set. Recall that the discrete topology on S is $\mathcal{P}(S)$, the power set of S , so that every subset of S is open (and closed). The discrete topology is [metrizable](#), and there are lots of metrics that generate this topology.

Suppose again that S is a nonempty set. A metric d on S with the property that there exists $c \in (0, \infty)$ such that $d(x, y) \geq c$ for distinct $x, y \in S$ generates the discrete topology.

Proof

Note that $B(x, c) = \{x\}$ for $x \in S$. Hence $\{x\}$ is open for each $x \in S$.

So any metric that is bounded from below (for distinct points) generates the discrete topology. It's easy to see that there are such metrics.

Suppose again that S is a nonempty set. The function d on $S \times S$ defined by $d(x, x) = 0$ for $x \in S$ and $d(x, y) = 1$ for distinct $x, y \in S$ is a metric on S , known as the *discrete metric*. This metric generates the discrete topology.

Proof

Clearly $d(x, y) = 0$ if and only if $x = y$, and $d(x, y) = d(y, x)$ for $x, y \in S$, so the positive and symmetric properties hold. For the triangle inequality, suppose $x, y, z \in S$. The inequality trivially holds if the points are not distinct. If the points are distinct, then $d(x, z) = 1$ and $d(x, y) + d(y, z) = 2$.

In probability applications, the discrete topology is often appropriate when S is countable. Note also that the discrete metric is the graph distance if S is made into the *complete graph*, so that (x, y) is an edge for every pair of distinct vertices $x, y \in S$.

This page titled [1.10: Metric Spaces](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.