Metric spaces, topology, and continuity

Written by Paul Schrimpf and modified by Hiro Kasahara
December 11, 2020
University of British Columbia
Economics 526

> Always consider a problem under the minimum structure in which it makes sense. ... one is naturally led to the study of problems with a kind of minimal and intrinsic structure. Besides the fact that it is much easier to find the crux of the matter in a simple structure than in a complicated one, there are not so many really basic structures, so one can hope that they will remain of interest for a very long time. — Talagrand (2005)

This lecture focuses on metric spaces, topology, and continuity. Similar material is covered in chapters 12 and 29, of Simon and Blume (1994), 1.3 of Carter (2001), chapter 2 of De la Fuente (2000), and any textbook on real analysis, such as Rudin (1976) or Tao (2006).

Many proofs in mathematics rely on showing some approximation can be made arbitrarily close. For example, showing that the first order conditions are necessary for an optimum relied on approximating the objective function with a first order expansion. To facilitate such arguments, we need some good ways to measure closeness. The most obvious way is to define some sort of distance. This is what metric spaces do. Given a measure of distance, we can think about convergence of sequences, continuity of functions, et cetera. We will see that a distance also gives rise to a classification of sets (as open and closed), and these open and close sets can also be used to describe convergence of sequences, continuity, and so on. This is what topology is about. Topology gives us another way of thinking about closeness without referring directly to distance. This can sometimes lead to easier proofs and new insights. There also exist sets that can be assigned a topology, but for which there does not exist an equivalent distance.[1]

1. Sequences and limits

A **sequence** is a list of elements, $\{x_1, x_2, ...\}$ or $\{x_n\}_{n=1}^{\infty}$ or sometimes just $\{x_n\}$. Although the notation for a sequence is similar to the notation for a set, they should not be confused. Sequences are different from sets in that the order of elements in a sequence matters, and the same element can appear many times in a sequence. Some examples of sequences with $x_i \in \mathbb{R}$ include

[1]In econometrics, you might eventually encounter weak convergence and hear references to the topology of weak convergence. This is an example of a non-metrizable topology.

(1) $\{1, 1, 2, 3, 5, 8, ...\}$
(2) $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...\}$
(3) $\{\frac{1}{2}, \frac{-2}{3}, \frac{3}{4}, \frac{-4}{5}, \frac{5}{6}, ...\}$

Some sequences, like 2, have elements that all get closer and closer to some fixed point. We say that these types of sequences converge. A sequence that does not converge diverges. Some divergent sequences like 1, increase without bound. Other divergent sequences, like 3, are bounded, but they do not converge to any single point.

To analyze sequences with elements that are not necessarily real numbers, we need to be able to say how far apart the entries in the sequence are.

Definition 1.1. A **metric space** is a set, X, and function $d : X \times X \to \mathbb{R}$ called a **metric** (or distance) such that $\forall x, y, z \in X$

(1) $d(x, y) > 0$ unless $x = y$ and then $d(x, x) = 0$
(2) (symmetry) $d(x, y) = d(y, x)$
(3) (triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 1.2. A **normed vector space** is a vector space X, together with a norm $|| \cdot || : X \to \mathbb{R}$, such that for all $x, y \in X$ and $\alpha \in \mathbb{R}$:

(1) $||x|| \geq 0$, with equality if and only if $x = 0$
(2) $||\alpha x|| = |\alpha| \cdot ||x||$
(3) (triangle inequality) $||x + y|| \leq ||x|| + ||y||$

Example 1.1. $\mathbb{R}$ is a metric space with $d(x, y) = |x - y|$.

Example 1.2. Any normed vector space is a metric space with $d(x, y) = \left\|x - y\right\|$.

Example 1.3. A set of bounded continuous functions on $[a, b] \subset \mathbb{R}$ with a metric $d(f, g) = \max_{x\in[a,b]} |f(x) - g(x)|$ is a metric space.

The most common metric space that we will encounter will be $\mathbb{R}^n$ with the Euclidean metric, $d(x, y) = \left\|x - y\right\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.

Definition 1.3. A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space **converges** to x if $\forall \epsilon > 0$ $\exists N$ such that

$$d(x_n, x) < \epsilon$$

for all $n \geq N$. We call x the **limit** of $\{x_n\}_{n=1}^{\infty}$ and write $\lim_{n\to\infty} x_n = x$ or $x_n \to x$.

Example 1.4. The sequence $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...\} = \{1/n\}_{n=1}^{\infty}$ converges. To see this, take any $\epsilon > 0$. Then $\exists$ N such that $1/N < \epsilon$. For all $n \geq N$, $d(1/n, 0) = 1/n < \epsilon$.

Example 1.5. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average of $\{X_i\}_{i=1}^{n}$ randomly sampled from identically distributed distribution with finite mean μ and finite variance. Then, the law of large numbers states that, for all $\delta > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| > \delta) = 0.$$

Or, equivalently, for all $\delta > 0$ and for all $\epsilon > 0$, there exists N such that

$$|\mathrm{P}(|\bar{X}_n - \mu| > \delta) - 0| = \mathrm{P}(|\bar{X}_n - \mu| > \delta) < \epsilon$$

for all $n \geq N$. Here, we consider a convergence of the sequence of positive numbers $\{p_n\}_{n=1}^{\infty}$ with $p_n := \mathrm{P}(|\bar{X}_n - \mu| > \delta)$ to a limit point 0, i.e., $p_n := \mathrm{P}(|\bar{X}_n - \mu| > \delta) \to 0$.

If a sequence does not converge, it diverges.

Example 1.6. The Fibonacci sequence, $\{1, 1, 2, 3, 5, 8, ...\}$ diverges.

Definition 1.4. a is an **accumulation point** of $\{x_n\}_{n=1}^{\infty}$ if $\forall \epsilon > 0$ $\exists$ infinitely many x_i such that

$$d(a, x_i) < \epsilon.$$

Example 1.7. The sequence $\{\frac{1}{2}, \frac{-2}{3}, \frac{3}{4}, \frac{-4}{5}, \frac{5}{6}, ...\}$ has two accumulation points, 1 and -1.

The third example of a sequence at the start of this section, 3, shows that the converse of this lemma is false. Not every accumulation point is a limit.

Definition 1.5. Given $\{x_n\}_{n=1}^{\infty}$ and any sequence of positive integers, $\{n_k\}$ such that $n_1 < n_2 < ...$ we call $\{x_{n_k}\}$ a **subsequence** of $\{x_n\}_{n=1}^{\infty}$.

In example 3, there are two accumulation points, -1 and 1, and you can find subsequences that converge to these points.

Lemma 1.1. *Let a be an accumulation point of $\{x_n\}$. Then $\exists$ a subsequence that converges to a.*

Proof. We can construct a subsequence as follows. Let $\{\epsilon_k\}$ be a sequence that converges to zero with $\epsilon_k > 0 \forall k$, (for example, $\epsilon_k = 1/k$). By the definition of accumulation point, for each ϵ_k $\exists$ infinitely many x_n such that

$$d(x_n, a) < \epsilon_k \tag{1}$$

Pick any x_{n_1} such that (1) holds for ϵ_1. For $k > 1$, pick $n_k \neq n_j$ for all $j < k$ and such that (1) holds for ϵ_k. Such an n_k always exists because there are infinite x_n that satisfy (1). By construction, $\lim_{k\to\infty} x_{n_k} = a$ (you should verify this using the definition of limit). □

The limit of any convergent sequence is an accumulation point of the sequence. In fact, it is the only accumulation point.

Lemma 1.2. *If $x_n \to x$, then x is the only accumulation point of $\{x_n\}_{n=1}^{\infty}$.*

Proof. Let $\epsilon > 0$ be given. By the definition of convergence, $\exists N$ such that

$$d(x_n, x) < \epsilon/2$$

for all $n \geq N$. $\{n \in \mathbb{N} : n \geq N\}$ is infinite, so x is an accumulation point.

Suppose x' is another accumulation point. Then, from Lemma 1.1, there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} x_{n_k} = x'$. Then, there exists N' such that $d(x_{n_k}, x') < \epsilon/2$

for all x_{n_k} with $n_k \geq N'$ in the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$. By the triangle inequality, $d(x, x') \leq d(x_{n_k}, x') + d(x_{n_k}, x) < \epsilon$ for all x_{n_k} with $n_k \geq \max\{N, N'\}$ in the subsequence. Since this inequality holds for any ϵ, it must be that $d(x, x') = 0$. d is a metric, so then $x = x'$, and the limit of sequence is the sequence's unique accumulation point. □

Convergence of sequences is often preserved by arithmetic operations, as in the following two theorems.

Theorem 1.1. *Let $\{x_n\}$ and $\{y_n\}$ be sequences in $\mathbb{R}^n$ (or any normed vector space) V. If $x_n \to x$ and $y_n \to y$, then*

$$x_n + y_n \to x + y.$$

Proof. Let $\epsilon > 0$ be given. Then $\exists\ N_x$ such that for all $n \geq N_x$,

$$d(x_n, x) < \epsilon/2,$$

and $\exists N_y$ such that for all $n \geq N_y$,

$$d(y_n, y) < \epsilon/2.$$

Let $N = \max\{N_x, N_y\}$. Then for all $n \geq N$,

$$\begin{aligned} d(x_n + y_n, x + y) = \left\|(x_n + y_n) - (x + y)\right\| &\leq \|x_n - x\| + \left\|y_n - y\right\| \\ &< \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

□

Theorem 1.2. *Let $\{x_n\}$ be a sequence in a normed vector space with scalar field $\mathbb{R}$ and let $\{c_n\}$ be a sequence in $\mathbb{R}$. If $x_n \to x$ and $c_n \to c$ then*

$$x_n c_n \to x c.$$

Proof. Left as an exercise. □

In fact, later we will see that if $f(\cdot, \cdot)$ is continuous, then $\lim f(x_n, y_n) = f(x, y)$. The previous two theorems are examples of this with $f(x, y) = x + y$ and $f(c, x) = cx$, respectively.

1.1. **Series.** Infinite sums or series are formally defined as the limit of the sequence of partial sums.

Definition 1.6. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in a normed vector space. Let $s_n = \sum_{i=1}^{n} x_i$ denote the sum of the first n elements of the sequence. We call s_n the nth partial sum. We define the sum of all the x_is as

$$\sum_{i=1}^{\infty} x_i \equiv \lim_{n \to \infty} s_n$$

This is called a(n infinite) **series**.

Example 1.8. Let $\beta \in \mathbb{R}$. $\sum_{i=0}^{\infty} \beta^i$ is called a geometric series. Geometric series appear often in economics, where β will be the subjective discount factor or perhaps $1/(1 + r)$.

Notice that

$$\begin{aligned} s_n =& 1 + \beta + \beta^2 + \cdots + \beta^n \\ =& 1 + \beta(1 + \beta + \cdots + \beta^{n-1}) \\ =& 1 + \beta(1 + \beta + \cdots + \beta^{n-1} + \beta^n) - \beta^{n+1} \\ s_n(1-\beta) =& 1 - \beta^{n+1} \\ s_n =& \frac{1-\beta^{n+1}}{1-\beta}, \end{aligned}$$

so,

$$\begin{aligned} \sum_{i=0}^{\infty} \beta^i =& \lim s_n \\ =& \lim \frac{1-\beta^{n+1}}{1-\beta} \\ =& \frac{1}{1-\beta} \text{ if } |\beta| < 1. \end{aligned}$$

1.2. **Cauchy sequences.** We have defined convergent sequences as ones whose entries all get close to a fixed limit point. This means that all the entries of the sequence are also getting closer together. You might imagine a sequence where the entries get close together without necessarily reaching a fixed limit.

Definition 1.7. A sequence $\{x_n\}_{n=1}^{\infty}$ is a **Cauchy** sequence if for any $\epsilon > 0$ $\exists N$ such that for all $i, j \geq N$, $d(x_i, x_j) < \epsilon$.

It turns that in $\mathbb{R}^n$ Cauchy sequences and convergent sequences are the same. This is a consequence of the way the real numbers are defined.

Theorem 1.3. *Every Cauchy sequence in $\mathbb{R}$ converges.*

There is a more detailed discussion of this in Chapter 29.1 of Simon and Blume. If you want more practice with the sort of proofs in this lecture, it would be good to read that section. The convergence of Cauchy sequences in the real numbers is equivalent to the least upper bound property that is discussed in the appendix to the lecture notes on sets.

Example 1.9. Consider a mapping $T : \mathbb{R} \to \mathbb{R}$ given by

$$T(x) := \alpha + \beta x,$$

where $\mathbb{R}$ is the set of real numbers. We assume that $0 < \beta < 1$. $T(x)$ is a contraction mapping, i.e., for any $x \in \mathbb{R}$ and $y \in \mathbb{R}$, $|T(x) - T(y)| < |x - y|$ because $|T(x) - T(y)| = \beta|x - y| < |x - y|$. Given $x_0 \in \mathbb{R}$, construct a sequence of the numbers $\{x_i\}_{i=0}^{\infty}$ by recursively computing $x_i = T(x_{i-1})$ starting from x_0. Then, $\{x_i\}$ is a Cauchy sequence.

Cauchy sequences do not converge in all metric spaces. For example, the rational numbers are a metric space, and any sequence of rationals that converges to an irrational number in $\mathbb{R}$ is a Cauchy sequence in $\mathbb{Q}$ but has no limit in $\mathbb{Q}$. Having Cauchy sequences converge is necessary for proving many theorems, so we have a special name for metric spaces where Cauchy sequences converge.

Definition 1.8. A metric space, X, is **complete** if every Cauchy sequence of points in X converges in X.

Example 1.10. $\mathbb{R}^n$ is a complete metric space. A brief argument follows. You may want to state the details as an exercise. Let $\{\mathbf{x}_i\}$ be a Cauchy sequence in $\mathbb{R}^n$. Each coordinate of $\mathbf{x}_i = (x_{1i}, ..., x_{ni})$ is a Cauchy sequence in $\mathbb{R}$. $\mathbb{R}$ is complete, so each coordinate has a limit, $x_{ji} \to x_j$ for $j = 1, ..., n$. Finally, show that $\mathbf{x} = (x_1, ..., x_n)$ is the limit of the original sequence in $\mathbb{R}^n$.

Example 1.11. $\ell^p = \{(x_1, x_2, ...)\, s.t.\, x_i \in \mathbb{R}, \sum_{i=1}^\infty |x_i|^p < \infty\}$ with metric

$$d_p(x, y) = \left(\sum_{i=1}^\infty |x_i - y_i|^p \right)^{1/p}$$

is a complete metric space.

Showing that ℓ^p is complete is slightly tricky because you have deal with a sequence of $\mathbf{x}_i \in \ell^p$, each element of which is itself an infinite sequence.

To show that ℓ^p is complete, let $\{\mathbf{x}_n\}_{n=1}^\infty$ be a Cauchy sequence. Let $\left\|x - y\right\|_p$ denote $d_p(x, y)$.

Denote the elements of $\mathbf{x}_i$ by $x_{i1}, x_{i2}, ...$. First, let's show that for any n, $x_{1n}, x_{2n}, ...$ is a Cauchy sequence in $\mathbb{R}$. Let $\epsilon > 0$. Since $\{\mathbf{x}_n\}_{n=1}^\infty$ is Cauchy, $\exists N_\epsilon$ such that for all $i, j \geq N_\epsilon$,

$$\left\|\mathbf{x}_i - \mathbf{x}_j\right\|_p < \epsilon.$$

Since

$$\left\|\mathbf{x}_i - \mathbf{x}_j\right\|_p^p = \left(\sum_{m=1}^\infty \left|x_{im} - x_{jm}\right|^p \right)$$

All terms in the sum on the right are non-negative and the sum includes $\left|x_{in} - x_{jn}\right|$, so

$$\left|x_{in} - x_{jn}\right|^p \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|_p^p$$
$$\left|x_{in} - x_{jn}\right| \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|_p$$

Therefore, $\left|x_{in} - x_{jn}\right| < \epsilon$ for all $i, j \geq N_\epsilon$, i.e. $x_{1n}, x_{2n}, ...$ is a Cauchy sequence in $\mathbb{R}$. $\mathbb{R}$ is complete, so it has some limit. Denote the limit by x_n^*.

Now we will show that $\mathbf{x}^* = (x_1^*, x_2^*, ...)$ is the limit of $\{\mathbf{x}_n\}_{n=1}^\infty$. First, we should show that $\mathbf{x}^* \in \ell^p$. Let

$$s_m^* = \sum_{n=1}^m |x_n^*|^p.$$

We need to show that $\lim s_m^*$ exists. Since $\{\mathbf{x}_n\}_{n=1}^{\infty}$ is Cauchy, $\exists j$ such that if $i \geq j$, $\left\|\mathbf{x}_i - \mathbf{x}_j\right\| < 1$. Using the triangle inequality,

$$\|\mathbf{x}_i\| \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\| + \left\|\mathbf{x}_j\right\| < 1 + \left\|\mathbf{x}_j\right\| \equiv M$$

for all $i \geq j$ and some fixed j. Thus, $\|\mathbf{x}_i\| \leq M$ for some constant M and all $i \geq j$. Then,

$$s_m^* = \sum_{n=1}^{m} |x_n^*|^p = \lim_{i\to\infty} \sum_{n=1}^{m} |x_{in}|^p \leq M^p.$$

The final inequality comes from the fact that $\sum_{n=1}^{m} |x_{in}|^p \leq \sum_{n=1}^{\infty} |x_{in}|^p = \|\mathbf{x}_i\|^p \leq M^p$. Importantly, the upper bound M^p does not depend on i or m. Thus, we have shown that

$$s_m^* \leq M^p$$

for all m. s_m^* is a bounded weakly increasing sequence in $\mathbb{R}$, so it must converge.[a]

Finally, we should show that $\{\mathbf{x}_n\}_{n=1}^{\infty}$ converges to $\mathbf{x}^*$. Let $\epsilon > 0$. Since the original sequence is Cauchy, there is an N such that if $i, j > N$, then

$$\sum_{m=1}^{M} \left|x_{im} - x_{jm}\right|^p \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^p < \epsilon$$

for all M. Therefore,

$$\lim_{j\to\infty} \sum_{m=1}^{M} \left|x_{im} - x_{jm}\right|^p = \sum_{m=1}^{M} |x_{im} - x_m^*| < \epsilon$$

for all $i \geq N$ and all M. Thus,

$$\|\mathbf{x}_i - \mathbf{x}^*\| = \lim_{M\to\infty} \sum_{m=1}^{M} |x_{im} - x_m^*| < \epsilon$$

for all $i \geq N$, so the sequence converges.

[a] Let $\{x_n\}_{n=1}^{\infty} \in \mathbb{R}$ and suppose $x_1 \leq x_2 \leq x_3 \leq \ldots$ and $\{x_n\}_{n=1}^{\infty}$ is bounded, then we will show $\{x_n\}_{n=1}^{\infty}$ converges. Suppose not. Then the sequence has no accumulation points. In particular, x_i is not an accumulation point of the sequence for any i i.e. there is an $\epsilon > 0$ such that for all i there are finitely many j with $d(x_i, x_j) < \epsilon$. Then we can construct a subsequence by choosing j_k such that $j_k > j_{k-1}$ and $|x_{j_k} - x_{j_{k-1}}| > \epsilon$. But then

$$\begin{aligned} x_{j_k} =& x_{j_1} + (x_{j_2} - x_{j_1}) + (x_{j_3} - x_{j_2}) + \ldots + (x_{j_k} - x_{j_{k-1}}) \\ \geq& x_{j_1} + (k-1)\epsilon \end{aligned}$$

which is not bounded.

Example 1.12. The space of bounded real-valued functions, denoted by $B(X)$, with the sup norm $||f||_\infty = \sup_{x\in X} |f(x)|$ is complete. Consider a mapping $T : B(X) \to B(X)$ and suppose that T is a contraction mapping. Then, starting from $v_0 \in B(X)$, construct a sequence of bounded functions $\{v_i\}_{i=0}^{\infty}$ recursively computing $v_i = T(v_{i-1})$. Then, $\{v_i\}_{i=1}^{\infty}$ is a Cauchy sequence. Therefore, this sequence has the limit in $B(X)$ and thus, $v(x) = \lim_{i\to\infty} v_i(x)$ is a bounded real-valued function. Furthermore, the limit of this sequence is the unique fixed point $v = T(v)$ by contraction mapping theorem. This argument is useful for proving that the fixed point of the Bellman equation is a real-valued bounded function. Analogous argument can be applied to the space of continuous (bounded) functions on a closed and bounded interval with the sup norm, which is also complete.

2. Open sets

Definition 2.1. Let X be a metric space and $x \in X$. A **neighborhood** of x is the set

$$N_\epsilon(x) = \{y \in X : d(x,y) < \epsilon\}.$$

A neighborhood is sometimes also called an open ϵ**-ball** of x and written $B_\epsilon(x)$. I will try to stick with the $N_\epsilon(x)$ notation, but I might sometimes use ball and neighborhood interchangeably.

Definition 2.2. A set, $S \subseteq X$ is **open** if $\forall x \in S$, $\exists\ \epsilon > 0$ such that

$$N_\epsilon(x) \subset S.$$

For every point in an open set, you can find a small neighborhood around that point such that the neighborhood lies entirely within the set.

Example 2.1. Any open interval, $(a,b) = \{x \in \mathbb{R} : a < x < b\}$, is an open set.

Theorem 2.1.
(1) *Any union of open sets is open. (finite or infinite)*
(2) *The* finite *intersection of open sets is open.*

Proof. Let S_j, $j \in J$ be a collection of open sets. If $x \in \cup_{j\in J} S_j$, then there exists j_0 such that $x \in S_{j_0}$. Because S_{j_0} is an open set, for every $\epsilon > 0$, $N_{\epsilon_{j_0}}(x) \subset S_{j_0}$ and the stated result of (1) follows form $S_{j_0} \subset \cup_{j\in J} S_j$.

Let $S_1, .., S_k$ be a finite collection of open sets. For each i $\exists \epsilon_i > 0$ such that $N_{\epsilon_i}(x) \subset S_i$. Let $\underline{\epsilon} = \min_{i\in\{1,\dots,k\}} \epsilon_i$. Then $\underline{\epsilon} > 0$ since it is the minimum of a finite set of positive numbers. Also, $N_{\underline{\epsilon}}(x) \subset S_i$ for each i, so $N_{\underline{\epsilon}}(x) \subset \cap_{i=1}^k S_i$. □

Definition 2.3. The **interior** of a set A is the union of all open sets contained in A. It is denoted as int(A).

From the previous, theorem, we know that the interior of any set is open.

Example 2.2. Here some examples of the interior of sets in $\mathbb{R}$.
(1) $A = (a,b)$, int$(A) = (a,b)$.
(2) $A = [a,b]$, int$(A) = (a,b)$.
(3) $A = \{1,2,3,4,\dots\}$, A $= \emptyset$

Exercise 2.1. Let X be a metric space and $\{x_n\}_{n=1}^\infty$ a sequence in X. Show that $x_n \to x$ if and only if for every open set U containing x $\exists N$ such that $x_n \in U$ for all $n \geq N$.

3. Closed sets

A closed set is almost like the opposite of an open set. However, a set can be both open and closed, and a set can be neither, so they are not exactly opposites.

Definition 3.1. A set $S \subseteq X$ is closed if its complement, S^c, is open.

For any metric space, X, the whole space, X is open. Therefore, the $X^c = \emptyset$ is closed. The empty set is considered open in any metric space. For now, you could just consider this to be by convention, but below we will that closed sets can also be defined by convergence of sequences. By that definition, the whole space is closed, so its complement, the empty set must be open. Anyway, the point is that the empty set and entire metric space are always both open and closed.

Example 3.1 ($\mathbb{R}$ is open and closed in the set $\mathbb{R}$)**.** The set of real numbers $\mathbb{R}$ is open and closed in the set $\mathbb{R}$. $\mathbb{R}$ is open because every point in the set has an open neighbourhood in $\mathbb{R}$. $\mathbb{R}$ is also closed because $\mathbb{R}$ contains the limits of any convergent sequences in $\mathbb{R}$ (See Theorem 3.2 below).

Closed sets behavior under union and intersections is the mirror image of that of open sets.

Theorem 3.1.

(1) *The intersection of any collection of closed sets is closed.*
(2) *The union of any finite collection of closed sets is closed.*

Proof. Let C_j, $j \in J$ be a collection of closed sets. Then $\left(\cap_{j\in J} C_j\right)^c = \cup_{j \in J} C_j^c$. C_j^c are open, so by theorem 2.1, $\cup_{j\in J} C_j^c = \left(\cap_{j\in J} C_j\right)^c =$ is open.

The proof of part 2 is similar. □

Example 3.2 (Closed sets)**.** Some examples of closed sets include

(1) $[a, b] \subseteq \mathbb{R}$
(2) $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$
(3) $\{x \in \mathbb{R}^n : Ax = b\}$ where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$

Example 3.3 (Discrete metric and topology)**.** Let X be any set. Define a metric on X by

$$\Delta(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

This is a metric because it is positive definite, symmetric, and satisfies the triangle inequality. This metric is called the discrete metric. In this space, the set of any single point $\{x\}$ is open since $N_\epsilon(x) = \{x\} \subseteq \{x\}$ for any $\epsilon < 1$. Any union of open sets is open, so all subsets of X are open. Since every set is open, the complement of every set is open, so every set is closed as well.

Closed sets can also be defined as sets that contain the limit of any convergent sequence in the set. Simon and Blume use this definition. The next theorem shows that their definition is equivalent to ours.

Theorem 3.2. *Let $\{x_n\}$ be any convergent sequence with each element contained in a set C. Then $\lim x_n = x \in C$ for all such $\{x_n\}$ if and only if C is closed.*

Proof. ($\Rightarrow$) Suppose that C contains all of its limit points but C is not closed. Because C is not close, C^c is not open. Therefore, there exists $x \in C^c$ such that, for every $\epsilon > 0$, $N_\epsilon(x)$ contains at least one point in $(C^c)^c = C$. Setting $\epsilon = 1/n$, let this point be x_n so that $x_n \in N_{\frac{1}{n}}(x) \cap C$ for all n. Then, $\{x_n\}$ is a sequence in C but $\lim_{n\to\infty} x_n = x \in C^c$. This contradicts the assumption that C contains all of its limit points.
($\Leftarrow$) Suppose that C is closed. Let $\{x_n\}$ be a sequence in C such that $x_n \in C$ for all n but its limit point is not in C, i.e., $\lim_{n\to\infty} x_n = x \notin C$. Then $x \in C^c$. Because C^c is open, for every $\epsilon > 0$, $N_\epsilon(x) \subset C^c$. On the other hand, because $x_n \to x$, there exists N such that $d(x_n, x) < \epsilon$ for all $n \geq N$. This means that $x_n \in N_\epsilon(x) \subset C^c$ for all $n \geq N$. But this contradicts the assumption that $x_n \in C$ for n. □

Definition 3.2. The **closure** of a set S, denoted by $\overline{S}$ (or cl(S)), is the intersection of all closed sets containing S.

Example 3.4. If S is closed, $\overline{S} = S$.

Example 3.5. $\overline{(0,1]} = [0,1]$

Lemma 3.1. *$\overline{S}$ is the set of limits of convergent sequences in S.*

Proof. Let $\{x_n\}$ be a convergent sequence in S with limit x. If C is any closed set containing S, then $\{x_n\}$ is in C and by theorem 3.2, $x \in C$. Therefore, $x \in \overline{S}$.

Let $x \in \overline{S}$. For any $\epsilon > 0$, $N_\epsilon(x) \cap S \neq \emptyset$ because otherwise $N_\epsilon(x)^c$ is a closed set containing S, but not x. Therefore, we can construct a sequence $x_n \in S \cap N_{1/n}(x)$ that converges to x and is in S. □

Example 3.6. $\overline{\{1/n\}_{n\in\mathbb{N}}} = \{0, 1, 1/2, 1/3, ...\}$

Definition 3.3. The **boundary** of a set S is $\overline{S} \cap \overline{S^c}$.

Example 3.7. The boundary of $[0,1]$ is $\{0,1\}$.

Example 3.8. The boundary of the unit ball, $\{x \in \mathbb{R}^2 : \|x\| < 1\}$ is the unit circle, $\{x \in \mathbb{R}^2 : \|x\| = 1\}$.

Lemma 3.2. *If x is in the boundary of S then $\forall \epsilon > 0$, $N_\epsilon(x) \cap S \neq \emptyset$ and $N_\epsilon(x) \cap S^c \neq \emptyset$.*

Proof. As in the proof of lemma 3.1, all ϵ-neighborhoods of $x \in \overline{S}$ must intersect with S. The same applies to S^c. □

Exercise 2.1 and theorem 3.2 show that there is an important relationship between convergence of sequences and open and closed sets. Given a definition of what it means for a sequence to converge, we could use theorem 3.2 to define closed sets. Open sets could then be defined as the complement of closed sets. Conversely, if we specify which sets are open and closed, we can then define convergence of sequences as in exercise 2.1. Metrics, convergence of sequences, and open and closed sets are three different ways of describing notions of proximity and continuity.

4. Compact sets

Compact sets are a generalization of finite sets. Compact sets are essential for proving many important theorems. Compact sets have a somewhat difficult to understand definition, but they are incredibly useful.

Definition 4.1. An **open cover** of a set S is a collection of open sets, $\{G_\alpha\}$ $\alpha \in \mathcal{A}$ such that $S \subseteq \cup_{\alpha \in \mathcal{A}} G_\alpha$.

Example 4.1. Some open covers of $\mathbb{R}$ are:

- $\{\mathbb{R}\}$
- $\{(-\infty, 1), (-1, \infty)\}$
- $\{..., (-3, -1), (-2, 0), (-1, 1), (0, 2), (1, 3), ...\}$
- $\{(x, y) : x < y\}$

The first two are finite open covers since they consist of finitely many open sets. The third is a countably infinite open cover. The fourth is an uncountably infinite open cover.

Example 4.2. Let X be a metric space and $A \subseteq X$. The set of open balls of radius ϵ centered at all points in A is an open cover of A. If A is finite / countable / uncountable, then this open cover will also be finite / countable / uncountable.

Open covers of the form in the previous example are often used to prove some property applies to all of A by verifying the property in each small $N_\epsilon(x)$. Unfortunately, this often involves taking a maximum or sum of something for each set in the open cover. When the open cover is infinite, it can be hard to ensure that the infinite sum or maximum stays finite. When we have a finite open cover, we know that things will remain finite.

Definition 4.2. A set K is **compact** if every open cover of K has a finite subcover.

By a finite subcover, we mean that there is finite subset of the sets in the open cover, $G_{\alpha_1}, ...G_{\alpha_k}$ such that $S \subset \cup_{j=1}^{k} G_{\alpha_j}$. Compact sets are a generalization of finite sets. Many facts that are obviously true of finite sets are also true for compact sets, but not true for infinite sets that are not compact. Suppose we want to show a set has some property. If the set is compact, we can cover it with a finite number of small ϵ balls and then we just need to show that each small ball has the property we want. We will see many concrete examples of this technique in the next few weeks.

Example 4.3. $\mathbb{R}$ is not compact. $\{..., (-3,-1), (-2,0), (-1,1), (0,2), (1,3), ...\}$ is an infinite cover, but if we leave out any single interval (the one beginning with n) we will fail to cover some number $(n+1)$.

Example 4.4. Let $K = \{x\}$, a set of a single point. Then K is compact. Let $\{G_\alpha\}_{\alpha\in\mathcal{A}}$ be an open cover of K. Then $\exists\ \alpha$ such that $x \in G_\alpha$. This single set is a finite subcover.

Example 4.5. Let $K = \{x_1, ..., x_n\}$ be a finite set. Then K is compact. Let $\{G_\alpha\}_{\alpha\in\mathcal{A}}$ be an open cover of K. Then for each i, $\exists\ \alpha_i$ such that $x_i \in G_{\alpha_i}$. The collection $\{G_{\alpha_1}, ...G_{\alpha_n}\}$ is a finite subcover.

Example 4.6. $(0,1) \subseteq \mathbb{R}$ is not compact. $\{(1/n, 1)\}_{n=2}^\infty$ is an open cover, but there can be no finite subcover. Any finite subcover would have a largest n and could not contain, e.g. $1/(n+1)$.

Example 4.7. Pick any $x \in \mathbb{R}^n$. Let $K = \{x\frac{1}{2}, x\frac{2}{3}, x\frac{3}{4}, ...\}$. Then K is not compact. Consider the open cover $N_{\|x\|\frac{1}{3(n+2)^2}}(x\frac{n}{n+1})$ for $n = 1, 2, ...$ Assuming $x \neq 0$, each of these neighborhoods contains exactly one point of K, so there is no finite subcover.

Before using compactness, let's investigate how being compact relates to other properties of sets, such as closed/open.

Lemma 4.1. *Let X be a metric space and $K \subseteq X$. If K is compact, then K is closed.*

Proof. Let $x \in K^c$. The collection $\{N_{d(x.y)/3}(y)\}$, $y \in K$ is an open cover of K. K is compact, so there is a finite subcover, $N_{d(x,y_1)/3}(y_1), ..., N_{d(x,y_n)/3}(y_n)$. For each i, $N_{d(x,y_i)/3}(y_i) \cap N_{d(x,y_i)/3}(x) = \emptyset$, so

$$\cap_{i=1}^n N_{d(x,y_i)/3}(x)$$

is an open neighborhood of x that is contained in K^c. K^c is open, so K is closed. □

Lemma 4.2. *Let X be a metric space, $C \subseteq K \subseteq X$. If K is compact and C is closed. Then C is also compact.*

Proof. Let $\{G_\alpha\}_{\alpha\in\mathcal{A}}$ be an open cover for C. Then $\{G_\alpha\}_{\alpha\in\mathcal{A}}$ plus C^c is an open cover for K. Since K is compact there is a finite subcover. Since $C \subseteq K$, the finite subcover also covers C. Therefore, C is compact. □

An equivalent definition of compactness is in terms of collection of sets with "the finite intersection property."

Definition 4.3. A collection of sets $\{C_\alpha\}_{\alpha\in\mathcal{B}}$ has the **finite intersection property** if for all finite subsets, $F \subseteq \mathcal{A}$, the intersection, $\cap_{\alpha\in F} C_\alpha$ is not empty.

Lemma 4.3. *Let X be a metric space and $K \subseteq X$. K is compact if and only if for every collection of closed subsets, $\{C_\alpha\}_{\alpha\in\mathcal{B}}$ with $C_\alpha \subseteq K$, with the finite intersection property, the intersection of all the subsets is not empty i.e. $\cap_{\alpha\in\mathcal{B}} C_\alpha$ is not empty.*

Proof. Left as an exercise. The statement in the lemma is basically the contrapositive of the definition of compact, combined with the observation that the complement of closed sets are open and vice versa, and the fact that $(A \cup B)^c = A^c \cap B^c$. □

The following corollary is useful.

Corollary 4.1. *Choose, $a, b \in \mathbb{R}$ with $a \leq b$. Let C_1, C_2, ..., be a sequence of non-empty closed sets in $[a, b]$ with*

$$C_1 \supseteq C_2 \supseteq \ldots.$$

Then, $\cap_{n=1}^{\infty} C_n \neq \emptyset$.

Proof. Any finite collection of $\{C_n\}_{n=1}^{\infty}$ is written as $\{C_{n_1}, C_{n_2}, ..., C_{n_k}\}$. Let $n^* = \max\{n_1, n_2, ..., n_k\}$ and then $\cap_{i=1}^{k} C_{n_i} = C_{n^*}$ is not empty. Because $[a, b]$ is compact, $\{C_n\}_{n=1}^{\infty}$ satisfies the finite intersection property and the state result follows from Lemma 4.3. □

The definition of compactness is somewhat abstract. We just saw that compact sets are always closed. Another property of compact sets is that they are bounded.

Definition 4.4. Let X be a metric space and $S \subseteq X$. S is **bounded** if $\exists x_0 \in S$ and $r \in \mathbb{R}$ such that

$$d(x, x_0) < r$$

for all $x \in S$.

A bounded set is one that fits inside an open ball of finite radius. For subsets of $\mathbb{R}$ this definition is equivalent to there being a lower and upper bound for the set. For subsets of a normed vector space, if S is bounded then there exists some M such that $\|x\| < M$ for all $x \in S$.

Lemma 4.4. *Let $K \subseteq X$ be compact. Then K is bounded.*

Proof. Pick $x_0 \in K$. $\{N_r(x_0)\}_{r\in\mathbb{R}}$ is an open cover of K, so there must be a finite subcover. The finite subcover has some maximum r^*. Then $K \subseteq N_{r^*}(x_0)$, so K is bounded. □

This lemma along with lemma 4.1 show that if a set is compact then it is also closed and bounded. In $\mathbb{R}^n$, the converse is also true.

Theorem 4.1 (Heine-Borel)**.** *A set $S \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Proof. We already showed that if S is compact, then it is closed and bounded.

Now suppose S is closed and bounded. Since S is bounded, it is a subset of some n-dimensional cube, say $[-a, a]^n$ (i.e. the set of all vectors $x = (x_1, .., x_n)$ with $-a \leq x_i \leq a$). We will show $[-a, a]^n$ is compact, and then use the fact that a closed subset of a compact set is compact.

Let's just show $[-a, a]^n$ is compact for $n = 1$. The argument for larger n is similar, but the notation is more cumbersome. If $[-a, a]$ is not compact, then there is an infinite open cover with no finite subcover, say $\{G_\alpha\}_{\alpha\in\mathcal{A}}$. If we cut the interval into two halves, $[-a, 0]$ and $[0, a]$, at least one of them must have no finite subcover. We can repeat this argument many times to get nested closed intervals of length $a/(2^k)$ for any k. Call the kth interval I_k. We claim that $\cap_{k=1}^\infty I_k \neq \emptyset$. To show this take the sequence of lower endpoints of the intervals, call it $\{x_n\}_{n=1}^\infty$. This is a Cauchy sequence, so it converges to some limit, x_0. Also, for any k, $\{x_n\}_{n=k}^\infty$ is a sequence in I_k. I_k is closed so $x_0 \in I_k$. Thus $x_0 \in \cap_{k=1}^\infty I_k$. On the other hand, $I_k \subset \cup_{\alpha\in\mathcal{A}} G_\alpha$ for all k. Therefore, x_0 must be in some open G_α as part of this cover. Then $\exists\epsilon > 0$ such that $N_\epsilon(x_0) \subset G_\alpha$. However, for k sufficiently large but finite so that $a/(2^k) < \epsilon$ or $k > -(\log\epsilon - \log(a))/\log(2)$, $I_k \subset N_\epsilon(x_0) \subset G_\alpha$, and then I_k has a finite subcover. Therefore, $[-a, a]$ must be compact.

The argument for $n > 1$ is very similar. For $n = 2$, we would divide the square $[-a, a]^2$ into four smaller squares. For $n = 3$, we would divide the cube into eight smaller cubes. In general we would divide the hypercube $[-a, a]^n$ into 2^n hypercubes with half the side length. □

You may wonder whether closed and bounded sets are always compact. We will see that all finite dimensional real vector spaces are isomorphic to $\mathbb{R}^n$. In any such space, sets are compact iff they are closed and bounded. However, in infinite dimensional spaces, there are closed and bounded sets that are not compact. The argument in the previous proof does not apply to infinite-dimensional spaces because an infinite dimensional hypercube can only be divided into infinitely many hypercubes with half the side length.

Example 4.8. $\ell^\infty = \{(x_1, x_2, ...) : \sup_i |x_i| < \infty$ with norm $\|x\| = \sup_i |x_i|$ is a normed vector space. Let e_i be the element of all 0s except for the ith position, which is 1. Then $E = \{e_i\}_{i=1}^\infty$ is closed and bounded. However, E is not compact because $\{N_{1/2}(e_i)\}_{i=1}^\infty$ is an open cover with no finite subcover.

It is always true that a closed subset of a compact set is compact.

Lemma 4.5. *Let $C \subseteq K \subseteq X$, where X is a metric space. If K is compact and C is closed, then C is compact.*

Proof. Let $\{G_\alpha\}_{\alpha\in\mathcal{A}}$ be an open cover for C. C is closed, so C^c is open. Also, C^c along with $\{G_\alpha\}_{\alpha\in\mathcal{A}}$ is an open cover of K. K is compact so there exists a finite subcover. This finite subcover (with C^c removed) is also a finite subcover of C. Therefore, C is compact. □

We saw that closed sets contain the limit points of all their convergent sequences. There is also a relationship between compactness and sequences.

Definition 4.5. Let X be a metric space and $K \subseteq X$. K is **sequentially compact** if every sequence in K has an accumulation point in K.

Sometimes this definition is written as: K is sequentially compact if every sequence in K has a subsequence that converges in K. Compactness implies sequential compactness.

Lemma 4.6. *Let X be a metric space and $K \subseteq X$ be compact. Then K is sequentially compact.*

Proof. Let $\{x_n\}_{n=1}^\infty$ be a sequence in K. Pick any $\epsilon > 0$, $N_\epsilon(x)$, $x \in K$ is an open cover of K, so there is a finite subcover. Therefore, one of the ϵ neighborhoods must contain an infinite number of the elements from the sequence. Call this neighborhood $N_\epsilon(x_1^*)$. Pick the smallest n such that $x_n \in N_\epsilon(x_1^*)$ and call it n_1. $K_1 = \overline{N_\epsilon(x_1^*)} \cap K$ is a closed subset of the compact set K, so is itself compact. Repeat the above argument with $\epsilon/2$ in place of ϵ and $K_1 = \overline{N_\epsilon(x_1^*)} \cap K$ in place of K to find K_2, K_3, etc. The sets $K_1, K_2, ...$ are all compact, (and also closed) subsets of K. Moreover, for any finite collection $K_{\ell_1}, ...K_{\ell_m}$, $\cap_{j=1}^m K_{\ell_j} = K_{\max_j \ell_j}$ is not empty. Since K is compact, $\cap_{j=1}^\infty K_j$ is also not empty. Finally, let $x^* \in \cap_{j=1}^\infty K_j$. Then x^* is an accumulation point of the original sequence $\{x_n\}$. □

In $\mathbb{R}^n$, a set is sequentially compact iff it is compact iff it is closed and bounded.

Theorem 4.2 (Bolzano-Weierstrass). *A set $S \subseteq \mathbb{R}^n$ is closed and bounded if and only if it is sequentially compact.*

Proof. Let S be closed and bounded. By the Heine-Borel theorem (4.1), S is compact. By lemma 4.6, S is sequentially compact.

Let S be sequentially compact. Let $\{x_n\}$ be a convergent sequence in S. Its limit is an accumulation point, so it must be in S. Therefore, S is closed. To show S is bounded, pick $x_0 \in S$. Suppose $\exists x_1 \in S$ such that $d(x_1, x_0) \geq 1$, and $x_2 \in S$ such that $d(x_2, x_0) \geq 2$ etc. This sequence is not Cauchy because of the reverse triangle inequality,

$$d(x_i, x_j) \geq \left|d(x_i, x_0) - d(x_j, x_0)\right| = |i - j|$$

This would be a sequence in S with no accumulation points. Therefore, it must not always be possible to find such x_n. In other words, S must be bounded. □

Comment 4.1. This theorem is sometimes stated as "each bounded sequence in $\mathbb{R}^n$ has a convergent subsequence." As an exercise, you may want to verify that this statement is equivalent to the one above.

Simon and Blume also prove this theorem in chapter 29.2. They do not prove the Heine-Borel theorem first though, so their proof is of the Bolzano-Weierstrass theorem is longer. Perhaps unsurprisingly, the details of their proof are somewhat similar to our proof of the Heine-Borel theorem.

In $\mathbb{R}^n$, compactness, sequential compactness, and closed and bounded are all the same. In general metric spaces, this need not be true. We saw above that in ℓ^∞, there are closed and bounded sets which are not compact. However it is always true that sequential compactness and compactness are the same for metric spaces. We already showed that compactness implies sequential compactness. The proof that sequential compactness implies compactness is a somewhat long and difficult, and you may want to skip it unless you are especially interested.

Theorem 4.3. *Let X be a metric space and $K \subseteq X$. K is compact if and only if K is sequentially compact.*

Proof. Lemma 4.6 shows that if K is compact, then K is sequentially compact.

Suppose every sequence in K has a convergent subsequence with a limit point in K. Let G_α, $\alpha \in \mathcal{A}$ be an open cover of K. $\mathcal{A}$ could be uncountable, so we will begin by showing that there must be a countable subcover. Let $n = 1$. Pick $x_1 \in K$. If possible choose $x_2 \in K$ such that $d(x_1, x_2) \geq 1/n$. Repeat this process, choosing x_j in K such that $d(x_j, x_i) \geq 1/n$ for each $i < j$. Eventually this will no longer be possible because otherwise we could construct a sequence with no convergent subsequence. When it is no longer possible, set $n = n + 1$. This gives a countable collection of open neighborhoods $N_{1/n}(x_i)$ that cover K for each n and get arbitrarily small as n increases. Call these neighborhoods η_j for $j = 1, 2, ...$ Let J be set of all η_j such that $\eta_j \subseteq G_\alpha$ for some α. J is a subset of a countable set, so J is countable. Note that $\cup_{j\in J}\eta_j \supset K$ because if $x \in K$, then $x \in G_\alpha$ for some α, and then $\exists\epsilon$ such that $N_\epsilon(x) \in G_\alpha$ and $\exists j$ s.t. $\eta_j \subset N_\epsilon(x)$. Finally, for each $j \in J$ choose G_{α_j} such that $\eta_j \subseteq G_{\alpha_j}$. Such α_j exist by construction. Also $\cup_{j\in J}G_{\alpha_j} \supset \cup_{j\in J}\eta_j \supset K$. So G_{α_j} is a countable subcover.

If G_{α_j} has no finite subcover, then for each n,

$$F_n = \left(\cup_{i=1}^n G_{\alpha_i}\right)^c \cap K$$

is not empty (if it were empty, then $\cup_{i=1}^n G_{\alpha_i}$ would be a finite subcover). Choose $x_n \in F_n$. Then $\{x_n\}$ is a sequence in K, and it must have a convergent subsequence with a limit, x_0, in K. However, each $F_{i+1} \subset F_i$ and F_i are all closed. Therefore, the sequence $\{x_j\}_{j=i}^\infty$ is also in F_i and so is its limit. Then $x_0 \in \cap_{i=1}^\infty F_i$. However,

$$\cap_{i=1}^\infty F_i = \left(\cup_{i=1}^\infty G_{\alpha_i}\right)^c \cap K,$$

but G_{α_i} is a countable cover of K, which implies

$$\cap_{i=1}^\infty F_i = \left(\cup_{i=1}^\infty G_{\alpha_i}\right)^c \cap K = \emptyset$$

and we have a contradiction. Therefore, G_{α_j} must have a finite subcover, and K is compact. □

Comment 4.2. There are non-metric spaces where sequential compactness and compactness are not equivalent. One can define open sets on a space without a metric by simply specifying which sets are open and making it such that theorem 2.1 holds. Such a space is called a topological space. You can then define closed sets, compact sets, and sequential compactness in terms of open sets. Exercise 2.1 showed that it is possible to define the convergence of sequences using only open sets, without referring to a metric at all. Similarly, you can define continuity of functions in terms of open and closed sets. Topology is the branch of mathematics that studies topological spaces. One interesting observation is that on $\mathbb{R}^n$, if a set is open with respect to some p-norm, then it is also open with respect to any other p-norm. Thus, we say that $\mathbb{R}^n$ with the p-norms are topologically equivalent or homeomorphic. Properties like continuity and compactness are the same regardless of what p-norm we use.

Topological spaces that are not metric spaces sound exotic, but they do sometimes appear. An example of a non-metrizable topological space is the following. Let $X = \{f : \mathbb{R} \to \mathbb{R}\}$. Define convergence of sequences in this space as $f_n \to f$ if $f_n(x) \to f(x)$ for all $x \in \mathbb{R}$. Let $S \subset X$ be closed if S contains the accumulation points of all sequences in S. This definition of closed and open sets is called the topology of pointwise convergence. This space and definition of convergence sounds reasonable (and is reasonable), but there is no metric on this space that leads to the same definition of convergence and closed sets.

In an area of econometrics and statistics called empirical process theory, you often have to work with something called the topology of weak convergence, which has a somewhat similar definition and is also non-metrizable. We saw in example 4.8 that in an infinite dimensional space, the unit sphere (or a closed neighborhood of any radius) is not compact. It is also not sequentially compact. Econometricians care about sequential compactness to show that limits and asymptotic distributions exist. Working with weak convergence instead of a metrizable topology makes many nice sets sequentially compact.

To review, in $\mathbb{R}^n$ a set is compact if any of the following four things hold:

(1) For every open cover there exists a finite subcover,
(2) Every collection of closed sets with the finite intersection property has a non-empty intersection
(3) Every sequence in the set has a convergent subsequence, or
(4) The set is closed and bounded.

The first two are always equivalent. In infinite dimensional spaces, closed and bounded sets need not be compact, but compact sets are always closed and bounded. In any metric space, a set is compact if and only if it is sequentially compact.

5. Functions and continuity

We have already used functions in this course, so perhaps we should have defined them earlier. Anyway, a **function** from a set A to a set B is a rule that assigns to each $a \in A$ one and only one $b \in B$. If we want to call this function f, we denote this by $f : A \to B$, which is read as "f is a function from A to B" or simply "f from A to B." The set A is called the **domain** of f. B is called the **target** of f. The set

$$\{y \in B : f(x) = y \text{ for some } x \in A\}$$

is called the **image** or **range** of f.

Example 5.1.

(1) Production functions: $f : \mathbb{R}^2 \to \mathbb{R}$
- Linear $f(x_1, x_2) = a_1 x_1 + a_2 x_2$
- Cobb-Douglas: $f(x_1, x_2) = K x_1^{\alpha_1} x_2^{\alpha_2}$
- Constant elasticity of substitution: $f(x_1, x_2) = K(c_1 x_1^{-a} + c_2 x_2^{-a})^{-b/a}$

(2) Utility functions: $u : \mathbb{R}^T \to \mathbb{R}$
- Constant relative risk aversion: $u(c_1, ..., c_T) = \sum_{t=1}^T \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}$
- Constant absolute risk aversion: $u(c_1, ..., c_T) = \sum_{t=1}^T \beta^t (-e^{-\alpha c_t})$

(3) Demand function with constant elasticity, $D : \mathbb{R}^3 \rightarrow \mathbb{R}^2$

$$D(p_1, p_2, y) = \begin{pmatrix} M p_1^{\alpha_{11}} p_2^{\alpha_{12}} y^{\beta_1} \\ M p_1^{\alpha_{21}} p_2^{\alpha_{22}} y^{\beta_2} \end{pmatrix}$$

where p_1 and p_2 are the prices of two goods and y is income.

I do not expect you to remember the names of these functions, but it is very likely that you will repeatedly encounter them this year.

A continuous function is a function without any jumps or holes. Formally,

Definition 5.1. A function $f : X \rightarrow Y$ where X and Y are metric spaces is **continuous** at x if whenever $\{x_n\}_{n=1}^{\infty}$ converges to x in X, then $f(x_n) \rightarrow f(x)$ in Y.

We simply say that f is continuous if it is continuous at every $x \in X$. There are some equivalent definitions of continuity that are also useful. You may have seen continuity defined as in the following lemma.

Lemma 5.1. *$f : X \rightarrow Y$ is continuous at x if and only if for every $\epsilon > 0$ $\exists$ $\delta > 0$ such that $d(x, x') < \delta$ implies $d(f(x), f(x')) < \epsilon$.*

Proof. Suppose f is continuous and there is an $\epsilon > 0$ such that such that for any $\delta > 0$, $d(x, x') < \delta$ does not imply $d(f(x), f(x')) < \epsilon$. Then by letting $\delta = 1/n$ we can construct a convergent sequence by choosing x_n such that $d(x, x_n) < 1/n$ and $d(f(x), f(x_n)) \geq \epsilon$. $x_n \rightarrow x$, but $f(x_n) \not\rightarrow f(x)$. Therefore, if f is continuous it must be impossible to construct such a sequence. This means that there must be some $\delta > 0$ such that $d(x, x') < \delta$ implies $d(f(x), f(x')) < \epsilon$.

Now suppose $\forall \epsilon > 0$ $\exists \delta > 0$ such that $d(x, x') < \delta$ implies $d(f(x), f(x')) < \epsilon$. Let $x_n \rightarrow x$. Then $\exists N$ s.t. $n \geq N$ implies $d(x, x_n) < \delta$. But this implies $d(f(x), f(x_n)) < \epsilon$, so f is continuous. ☐

Both the definition of continuity and (5.1) are about f being continuous at a point x. We say that f is continuous on $U \subseteq X$ if f is continuous at all $x \in U$.

A third way of defining continuity is in terms of open sets. First, another definition.

Definition 5.2. Let $f : X \rightarrow Y$. f is continuous The **preimage** of $V \subseteq Y$ is the set in X, $f^{-1}(V)$ defined by

$$f^{-1}(V) = \{x \in X : f(x) \in V\}$$

A function is continuous if and only if the preimage of any open set is open.

Lemma 5.2. *$f : X \rightarrow Y$ is continuous at x if and only if for all open V with $f(x) \in V$, $\exists$ open $U \subseteq X$ such that $x \in U \subseteq f^{-1}(V)$.*

Proof. Suppose for all open $V \subseteq Y$ that $\exists$ open U such that $x \in U \subseteq f^{-1}(V)$. We want to show that then f is continuous on U. To do that, let $x_n \rightarrow x$ and let $\epsilon > 0$. $N_\epsilon(f(x))$ is open, so by assumption, $\exists$ open U such that $x \in U \subseteq f^{-1}(N_\epsilon(f(x)))$. By the definition of open sets, $\exists$ $\delta > 0$ such that $N_\delta(x) \subseteq U \subseteq f^{-1}(N_\epsilon(f(x)))$. By the definition of $x_n \rightarrow x$, $\exists N$ such that if $n \geq N$, $x_n \in N_\delta(x)$. Then $x_n \in f^{-1}(N_\epsilon(f(x)))$, so $f(x_n) \in N_\epsilon(f(x))$, i.e.

$$d\left(f(x_n), f(x)\right) < \epsilon.$$

Therefore, $f(x_n) \to f(x)$.

Conversely, suppose f is continuous. Let $V \subseteq Y$ be open. We want to show that $\exists$ open U such that $x \in U \subseteq f^{-1}(V)$. Suppose there is no such U. Then for any $\epsilon > 0$, $\exists \tilde{x}_\epsilon \notin f^{-1}(V)$ with

$$d(x, \tilde{x}_\epsilon) < \epsilon.$$

Pick a sequence of ϵ_n that converges to zero, such as $\epsilon_n = 1/n$. Then the associated $\tilde{x}_n \to x$. However, since each $\tilde{x}_n \notin f^{-1}(V)$, $f(\tilde{x}_n) \in V^c$. But then having $f(\tilde{x}_n) \to f(x)$ would mean that V^c is not closed, which contradict V being open. Thus, there must exist an open U such that $x \in U \subseteq f^{-1}(V)$ when f is continuous. □

Since the a set is open if and only its complement is closed, we can also define continuity using closed sets.

Corollary 5.1. *$f : X \to Y$ is continuous if and only if $f^{-1}(V)$ is closed for all closed $V \subseteq Y$.*

Proof. Let $V \subseteq Y$ be closed. Then V^c is open. Also, note that the complement of the preimage of V is the preimage of V^c. In symbols,

$$f^{-1}(V)^c = \{x \in X : f(x) \notin V\} = \{x \in X : f(x) \in V^c\} = f^{-1}(V^c).$$

From lemma 5.2, f is continuous iff $f^{-1}(V^c) = f^{-1}(V)^c$ is open for all open sets V^c, which is true iff $f^{-1}(V)$ is closed for all closed sets V. □

Earlier we saw that convergence of sequences is preserved by arithmetic. Since continuity can be defined using sequences, it should be no surprise that continuity is also preserved by arithmetic.

Theorem 5.1. *Let $f : X \to Y$ and $g : X \to Y$ be continuous and X and Y be vector spaces. Then $(f + g)(x) = f(x) + g(x)$ is continuous.*

Proof. If f and g are continuous, then by definition $f(x_n) \to f(x)$ and $g(x_n) \to g(x)$ whenever $x_n \to x$. From the previous lecture the limit of a (finite) sum is the sum of limits, so $f(x_n) + g(x_n) \to f(x) + g(x)$, and $f + g$ is continuous. □

Similar results can be shown for subtraction, multiplication, etc, whenever they are well defined.

Continuity is also preserved by composition.

Theorem 5.2. *Let $f : X \to Y$ and $g : Y \to Z$ be continuous where X, Y, and Z are metric spaces. Then $f \circ g$ is continuous, where*

$$(f \circ g)(x) = f(g(x)).$$

Proof. Let $x_n \to x$. g is continuous, so $g(x_n) \to g(x)$. f is also continuous, so $f(g(x_n)) \to f(g(x))$. □

$f \circ g$ is called the composition of f and g.

Exercise 5.1. Rewrite the above proofs of Theorems 5.1 and 5.2 more rigorously using the definition of convergence, i.e., $x_n \to x$ indicates that for all $\epsilon > 0$, there exists N

such that $n \geq N$ implies that $d(x_n, x) < \epsilon$. Can you also rewrite the proof using the alternative characterization of contnuity based on ϵ-δ in Lemma 5.1?

5.1. **Onto, one-to-one, and inverses.** We have already used the concepts of onto, one-to-one, and inverses. We restate the definitions here.

Definition 5.3. $f : X \to Y$ is **one-to-one** or **injective** if for all $x_1, x_2 \in X$,

$$f(x_1) = f(x_2)$$

if and only if $x_1 = x_2$.

Equivalently, f is injective if for each $y \in Y$, the set $\{x : f(x) = y\}$ is either a singleton or empty. In terms of a nonlinear equation, if f is one-to-one, then $f(x) = b$ has at most one solution.

Definition 5.4. $f : X \to Y$ is **onto** or **surjective** if $\forall y \in Y$, $\exists x \in X$ such that $f(x) = y$.

In terms of a nonlinear equation, if f is onto, then $f(x) = b$ has at least one solution. When f is one-to-one and onto, we say that f is **bijective**. A bijective function has an inverse.

Definition 5.5. If $f : X \to Y$ is bijective, then the **inverse** of f, written f^{-1} satisfies

$$f(f^{-1}(y)) = y$$

and

$$f^{-1}(f(x)) = x.$$

Comment 5.1. While writing these notes, I briefly tried to prove that if $f : X \to Y$ is bijective and continuous, then f^{-1} is continuous. I could not do this, which is good, because that statement is false. You have to be a little creative in defining X and Y to come up with a counterexample. Let $X = [0, 2\pi)$ and $Y = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Then $f(x) = (cos(x), sin(x))$ is bijective and continuous, but f^{-1} is not continuous at $(1, 0)$.

This counterexample is actually related to a fundamental fact in topology. You may remember from last lecture that topology is about studying spaces with open and closed sets that do not necessarily have a metric. One thing that people are interested when studying such spaces is finding a continuous (in both directions) bijections between them. Loosely speaking, two topological spaces will have a continuous bijection between them if one can be bent and stretched from one into the form of another. You cannot bend a circle into an interval without breaking the circle, so there is no continuous bijection between the circle and an interval. When there is a continuous bijection between two spaces, they have the same collection of open sets, so to a topologist, they are the same. We then call the spaces homeomorphic (or topologically isomorphic). Loosely speaking, spaces will be homeomorphic if they are the same dimension and their shapes have the same number of holes. The circle has one hole, an interval has none, so they are not topologically isomorphic. I'd be remiss not to make a joke now, so here goes: Why did the topologist eat her/his coffee mug and drink from his/her donut? Because they're topologically isomorphic. Hahaha.

We begin the course by studying how to solve optimization problems. However, we did not worry too much about conditions that ensure an optimum exists. One fundamental result that ensures the existence of optima is Weierstrass's theorem.

Theorem 5.3 (Weierstrass)**.** *Let X be a metric space and $f : X \to \mathbb{R}$ be continuous and $K \subset X$ be compact and nonempty. Then $\exists x^* \in K$ such that $f(x^*) \geq f(x) \forall x \in K$.*

Proof. First, we show that $\sup_{x \in K} f(x)$ exists. $\mathbb{R}$ has the least upper bound property, so it is enough to show that f is bounded. Let $\epsilon > 0$. For each x, let $x \in U_x$ be open and $U_x \subseteq f^{-1}(N_\epsilon(f(x)))$. Such U_x exist since f is continuous. Also, $\{U_x\}_{x \in K}$ is open cover of K. K is compact, so there exists a finite subcover. Let this subcover be $U_{x_1}, ..., U_{x_n}$. For any $x \in K$, $x \in U_{x_j}$ for some j and

$$f(x) \leq f(x_j) + \epsilon.$$

Hence,

$$f(x) \leq \max_{j \in \{1,...,n\}} f(x_j) + \epsilon$$

and f is bounded on K. $\mathbb{R}$ has the least upper bound property — any set bounded above has a least upper bound, so $\bar{f} = \sup_{x \in K} f(x)$ exists.

Let $\bar{f} = \sup_{x \in K} f(x)$. Let $y_n = \bar{f} - 1/n$. By definition of supremum, for each n, $\exists\, x_n \in K$ such that $y_n \leq f(x_n) \leq \bar{f}$ (otherwise, y_n would be the supremum). Since K is compact, the sequence $\{x_n\}$ must have a convergent subsequence $x_{n_k} \to x^* \in K$. f is continuous, so $f(x_{n_k}) \to f(x)$. Also, by construction, $d(f(x_{n_k}), \bar{f}) \to 0$, so it must be that $f(x^*) = \bar{f}$ for $x^* \in K$. □

6. Correspondences

[2]A function, $f : X \to Y$ associates exactly one element of Y, $f(x)$, with each $x \in X$. Often we encounter things that are like functions, but for each $x \in X$, there are multiple elements of Y. We call this generalization of a function as correspondence.

Definition 6.1. A **correspondence** from a set X to a set Y, is a rule that assigns to each $x \in X$ a subset of Y. We denote a correspondence by $\phi : X \rightrightarrows Y$.

An equivalent definition is that $\phi : X \rightrightarrows Y$ is a function from X to the power set of Y. Correspondences appear often in economics, especially as constraint sets in optimization problems.

Example 6.1 (Budget correspondence)**.** Suppose there are n goods with prices $p \in \mathbb{R}^n$. Then given income of m, a consumer can afford $\chi(p, m) = \{x \in X \subseteq \mathbb{R}^n : p'x \leq m\}$, which defines a correspondence $\chi : \mathbb{R}^{n+1} \rightrightarrows X$. We can write the consumer's problem

[2]This section is largely based on section 2.1.5 of Carter (2001).

of maximizing utility subject to the budget constraint as

$$\max_{x \in \chi(p,m)} u(x)$$

If this problem has a solution, then the indirect utility function is the maximized utility,

$$v(p,m) = \max_{x \in \chi(p,m)} u(x).$$

The demand correspondence (usually function) is

$$x^*(p,m) = \arg\max_{x \in \chi(p,m)} u(x).$$

Such maximization problems are central to economics. To derive properties of the indirect utility and demand functions it is often useful to treat the budge set as a correspondence.

Correspondences also appear in economics in models with multiple equilibria, such as many games.

Defining continuity is a bit more complicated for correspondences than for functions. A function can either be continuous or it can jump. A correspondence can also expand or contract. For example, consider $\xi : \mathbb{R} \rightrightarrows \mathbb{R}$ defined by

$$\xi(x) = \begin{cases} [0,1] & \text{if } x > 0 \\ [1/4, 3/4] & \text{if } x \leq 0 \end{cases}$$

and $\psi : \mathbb{R} \rightrightarrows \mathbb{R}$ defined by

$$\psi(x) = \begin{cases} [0,1] & \text{if } x \geq 0 \\ [1/4, 3/4] & \text{if } x < 0 \end{cases}$$

Both these correspondences are somewhat continuous because they contain a continuous function, e.g. $f(x) = 1/2$, for all x. However, they are also somewhat discontinuous because the corresponding set changes suddenly at 0. Motivated by this observation we define the following:

Definition 6.2. A correspondence, $\phi : X \rightrightarrows Y$ is **upper hemicontinuous** at x if for all sequences $x_n \rightarrow x$ and $y_n \in \phi(x_n)$ with $y_n \rightarrow y$, then $y \in \phi(x)$.

In the previous example, ψ is upper hemicontinuous at 0, but ξ is not. To see this consider $x_n = 1/n$ and $y_n = 1$.

Definition 6.3. A correspondence, $\phi : X \rightrightarrows Y$ is **lower hemicontinuous** at x if for all sequences $x_n \rightarrow x$ and $y \in \phi(x)$, there exists a subsequence, x_{nk} and $y_k \in \phi(x_{nk})$ with $y_k \rightarrow y$.

In the previous example, ξ is lower hemicontinuous at 0, but ψ is not. To see this consider $x_n = -1/n$ and $y = 1$.

Definition 6.4. We say that a correspondence is **continuous** if it is both upper and lower hemicontinuous.

At all $x \neq 0$, ξ and ψ are continuous.

Exercise 6.1. A correspondence $\phi : X \rightrightarrows Y$ can also be viewed as a function $\phi : X \to \mathcal{P}(Y)$. There is a metric on the set of compact subsets of Y (which is a subset of $\mathcal{P}(Y)$) that leads to the same definition of continuity as in definition 6.4. What is that metric?

References

Carter, Michael. 2001. *Foundations of mathematical economics*. MIT Press.

De la Fuente, Angel. 2000. *Mathematical methods and models for economists*. Cambridge University Press.

Rudin, Walter. 1976. *Principles of Mathematical Analysis (International Series in Pure & Applied Mathematics)*. McGraw-Hill Publishing Co.

Simon, Carl P and Lawrence Blume. 1994. *Mathematics for economists*. Norton New York.

Talagrand, Michel. 2005. *The generic chaining*. Springer.

Tao, Terence. 2006. *Analysis*, vol. 1. Springer.