Chapter 7

Metric Spaces

A metric space is a set X that has a notion of the distance $d(x, y)$ between every pair of points $x, y \in X$. The purpose of this chapter is to introduce metric spaces and give some definitions and examples. We do not develop their theory in detail, and we leave the verifications and proofs as an exercise. In most cases, the proofs are essentially the same as the ones for real functions or they simply involve chasing definitions.

7.1. Metrics

A metric on a set is a function that satisfies the minimal properties we might expect of a distance.

Definition 7.1. A metric d on a set X is a function $d : X \times X \to \mathbb{R}$ such that for all $x, y \in X$:

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

(2) $d(x, y) = d(y, x)$ (symmetry);

(3) $d(x, y) \leq d(x, z) + d(z, x)$ (triangle inequality).

A metric space (X, d) is a set X with a metric d defined on X.

We can define many different metrics on the same set, but if the metric on X is clear from the context, we refer to X as a metric space and omit explicit mention of the metric d.

Example 7.2. A rather trivial example of a metric on any set X is the discrete metric

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

Example 7.3. Define $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$d(x, y) = |x - y|.$$

Then d is a metric on $\mathbb{R}$. Nearly all the concepts we discuss for metric spaces are natural generalizations of the corresponding concepts for $\mathbb{R}$ with this absolute-value metric.

Example 7.4. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

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

Then d is a metric on $\mathbb{R}^2$, called the Euclidean, or ℓ^2, metric. It corresponds to the usual notion of distance between points in the plane. The triangle inequality is geometrically obvious, but requires an analytical proof (see Section 7.6).

Example 7.5. The Euclidean metric $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$ is defined by

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

where

$$x = (x_1, x_2, \dots, x_n), \qquad y = (y_1, y_2, \dots, y_n).$$

For $n = 1$ this metric reduces to the absolute-value metric on $\mathbb{R}$, and for $n = 2$ it is the previous example. We will mostly consider the case $n = 2$ for simplicity. The triangle inequality for this metric follows from the Minkowski inequality, which is proved in Section 7.6.

Example 7.6. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d(x,y) = |x_1 - y_1| + |x_2 - y_2| \qquad x = (x_1, x_2), \quad y = (y_1, y_2).$$

Then d is a metric on $\mathbb{R}^2$, called the ℓ^1 metric. It is also referred to informally as the "taxicab" metric, since it's the distance one would travel by taxi on a rectangular grid of streets.

Example 7.7. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d(x,y) = \max\left(|x_1 - y_1|, |x_2 - y_2|\right) \qquad x = (x_1, x_2), \quad y = (y_1, y_2).$$

Then d is a metric on $\mathbb{R}^2$, called the ℓ^∞, or maximum, metric.

Example 7.8. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ for $x = (x_1, x_2)$, $y = (y_1, y_2)$ as follows: if $(x_1, x_2) \neq k(y_1, y_2)$ for $k \in \mathbb{R}$, then

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

and if $(x_1, x_2) = k(y_1, y_2)$ for some $k \in \mathbb{R}$, then

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

That is, $d(x,y)$ is the sum of the Euclidean distances of x and y from the origin, unless x and y lie on the same line through the origin, in which case it is the Euclidean distance from x to y. Then d defines a metric on $\mathbb{R}^2$.

In Britain, d is sometimes called the "British Rail" metric, because all the train lines radiate from London (located at the origin). To take a train from town x to town y, one has to take a train from x to 0 and then take a train from 0 to y, unless x and y are on the same line, when one can take a direct train.

Example 7.9. Let $C(K)$ denote the set of continuous functions $f : K \to \mathbb{R}$, where $K \subset \mathbb{R}$ is compact; for example, we could take $K = [a, b]$ to be a closed, bounded interval. For $f, g \in C(K)$ define

$$d(f, g) = \sup_{x \in K} |f(x) - g(x)|.$$

The function $d : C(K) \times C(K) \to \mathbb{R}$ is well-defined, since a continuous function on a compact set is bounded; in fact, such a function attains it maximum value, so we could also write

$$d(f, g) = \max_{x \in K} |f(x) - g(x)|.$$

Then d is a metric on $C(K)$. Two functions are close with respect to this metric if their values are close at every point of K.

Subspaces of a metric space (X, d) are subsets $A \subset X$ with the metric d_A obtained by restricting the metric d on X to A.

Definition 7.10. Let (X, d) be a metric space. A subspace (A, d_A) of (X, d) consists of a subset $A \subset X$ whose metric $d_A : A \times A \to \mathbb{R}$ is is the restriction of d to A; that is, $d_A(x, y) = d(x, y)$ for all $x, y \in A$.

We can often formulate properties of subsets $A \subset X$ of a metric space (X, d) in terms of properties of the corresponding metric subspace (A, d_A).

7.2. Norms

In general, there are no algebraic operations defined on a metric space, only a distance function. Most of the spaces that arise in analysis are vector, or linear, spaces, and the metrics on them are usually derived from a norm, which gives the "length" of a vector

Definition 7.11. A normed vector space $(X, \|\cdot\|)$ is a vector space X (which we assume to be real) together with a function $\|\cdot\| : X \to \mathbb{R}$, called a norm on X, such that for all $x, y \in X$ and $k \in \mathbb{R}$:

(1) $0 \leq \|x\| < \infty$ and $\|x\| = 0$ if and only if $x = 0$;

(2) $\|kx\| = |k|\|x\|$;

(3) $\|x + y\| \leq \|x\| + \|y\|$.

The properties in Definition 7.11 are natural ones to require of a length: The length of x is 0 if and only if x is the 0-vector; multiplying a vector by k multiplies its length by $|k|$; and the length of the "hypoteneuse" $x + y$ is less than or equal to the sum of the lengths of the "sides" x, y. Because of this last interpretation, property (3) is referred to as the triangle inequality.

Proposition 7.12. If $(X, \|\cdot\|)$ is a normed vector space X, then $d : X \times X \to \mathbb{R}$ defined by $d(x, y) = \|x - y\|$ is a metric on X.

Proof. The metric-properties of d follow immediately from properties (1) and (3) of a norm in Definition 7.11. □

A metric associated with a norm has the additional properties that for all $x, y, z \in X$ and $k \in \mathbb{R}$

$$d(x + z, y + z) = d(x, y), \qquad d(kx, ky) = |k|d(x, y),$$

which are called translation invariance and homogeneity, respectively. These properties do not even make sense in a general metric space since we cannot add points or multiply them by scalars. If X is a normed vector space, we always use the metric associated with its norm, unless stated specifically otherwise.

Example 7.13. The set of real numbers $\mathbb{R}$ with the absolute-value norm $|\cdot|$ is a one-dimensional normed vector space.

Example 7.14. The set $\mathbb{R}^2$ with any of the norms defined for $x = (x_1, x_2)$ by

$$\|x\|_1 = |x_1| + |x_2|, \qquad \|x\|_2 = \sqrt{x_1^2 + x_2^2}, \qquad \|x\|_\infty = \max(|x_1|, |x_2|)$$

is a two-dimensional normed vector space. The corresponding metrics are the "taxicab" metric, the Euclidean metric, and the maximum metric, respectively.

These norms are special cases of the following example.

Example 7.15. The set $\mathbb{R}^n$ with the ℓ^p-norm defined for $x = (x_1, x_2, \dots, x_n)$ and $1 \le p < \infty$ by

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$$

and for $p = \infty$ by

$$\|x\|_\infty = \max(|x_1|, |x_2|^p, \dots, |x_n|^p)$$

is an n-dimensional normed vector space for every $1 \le p \le \infty$. The Euclidean case $p = 2$ is distinguished by the fact that the norm $\|\cdot\|_2$ is derived from an inner product on $\mathbb{R}^n$:

$$\|x\|_2 = \sqrt{\langle x, x\rangle}, \qquad \langle x, y\rangle = \sum_{i=1}^{n} x_i y_i.$$

The triangle inequality for the ℓ^p-norm is called Minkowski's inequality. It is straightforward to verify if $p = 1$ or $p = \infty$, but it is not obvious if $1 < p < \infty$. We give a proof of the simplest case $p = 2$ in Section 7.6.

Example 7.16. Let $K \subset \mathbb{R}$ be compact. Then the space $C(K)$ of continuous functions $f : K \to \mathbb{R}$ with the sup-norm $\|\cdot\| : C(K) \to \mathbb{R}$, defined by

$$\|f\| = \sup_{x \in K} |f(x)|,$$

is a normed vector space. The corresponding metric is the one described in Example 7.9.

Example 7.17. The discrete metric in Example 7.2 and the metric in Example 7.8 are not derived from a norm.

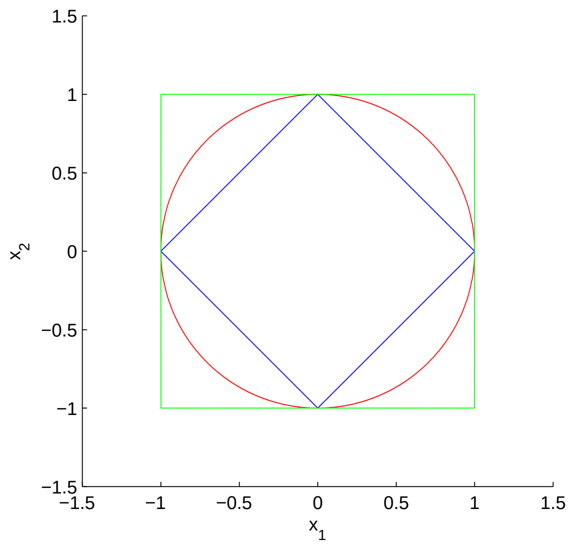

Figure 1. Boundaries of the unit balls $B_1(0)$ in $\mathbb{R}^2$ for the ℓ^1-norm (diamond), the ℓ^2-norm (circle), and the ℓ^∞-norm (square).

7.3. Sets

We first define an open ball in a metric space, which is analogous to a bounded open interval in $\mathbb{R}$.

Definition 7.18. Let (X, d) be a metric space. The open ball of radius $r > 0$ and center $x \in X$ is the set

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

Example 7.19. Consider $\mathbb{R}$ with its standard absolute-value metric, defined in Example 7.3. Then the open ball

$$B_r(x) = \{y \in \mathbb{R} : |x - y| < r\}$$

is the open interval of radius r centered at x.

Next, we describe the unit balls in $\mathbb{R}^2$ with respect to some different metrics.

Example 7.20. Consider $\mathbb{R}^2$ with the Euclidean metric defined in Example 7.4. Then $B_r(x)$ is a disc of diameter $2r$ centered at x. For the ℓ^1-metric in Example 7.6, the ball $B_r(x)$ is a diamond of diameter $2r$, and for the ℓ^∞-metric in Example 7.7, it is a square of side $2r$ (see Figure 1).

The norms $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ on $\mathbb{R}^n$ satisfy

$$\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le n\|x\|_\infty.$$

These inequalities correspond to the nesting of one ball inside another in Figure 1. Furthermore, the ℓ^∞-ball of radius 1 is included in the ℓ^1-ball of radius 2. As a result, every open ball with respect to one norm contains an open ball with respect to the other norms, and we say that the norms are equivalent. It follows from the definitions below that, despite the different geometries of their unit balls, these norms define the same collection of open sets and neighborhoods (i.e. the same topologies) and the same convergent sequences, limits, and continuous functions.

Example 7.21. Consider the space $C(K)$ of continuous functions $f : K \to \mathbb{R}$ with the sup-metric defined in Example 7.9. The ball $B_r(f)$ consists of all continuous functions $g : K \to \mathbb{R}$ whose values are strictly within r of the values of f at every point $x \in K$.

One has to be a little careful with the notion of open balls in a general metric space because they do not always behave the way their name suggests.

Example 7.22. Let X be a set with the discrete metric given in Example 7.2. Then $B_r(x) = \{x\}$ consists of a single point if $0 \le r < 1$ and $B_r(x) = X$ is the whole space if $r \ge 1$.

An another example, what are the open balls for the metric in Example 7.8?

We define open sets in a metric space analogously to open sets in $\mathbb{R}$.

Definition 7.23. Let X be a metric space. A set $G \subset X$ is open if for every $x \in G$ there exists $r > 0$ such that $B_r(x) \subset G$.

We can give a more geometrical definition of an open set in terms of neighborhoods.

Definition 7.24. Let X be a metric space. A set $U \subset X$ is a neighborhood of $x \in X$ if $B_r(x) \subset U$ for some $r > 0$.

Thus, a set is open if and only if every point in the set has a neighborhood that is contained in the set. In particular, an open set is itself a neighborhood of every point in the set.

The following is the topological definition of a closed set.

Definition 7.25. Let X be a metric space. A set $F \subset X$ is closed if

$$F^c = \{x \in X : x \notin F\}$$

is open.

Bounded sets in a metric space are defined in the obvious way.

Definition 7.26. Let (X, d) be a metric space A set $A \subset X$ is bounded if there exist $x \in X$ and $0 \le R < \infty$ such that

$$d(x, y) \le R \qquad \text{for all } y \in A.$$

Equivalently, this definition says that $A \subset B_R(x)$. The center point $x \in X$ is not important here. The triangle inequality implies that

$$B_R(x) \subset B_S(y), \qquad S = R + d(x, y),$$

so if the definition holds for some $x \in X$, then it holds for every $x \in X$. Alternatively, we define the diameter $0 \le \operatorname{diam} A \le \infty$ of a set $A \subset X$ by

$$\operatorname{diam} A = \sup \{d(x, y) : x, y \in A\}.$$

Then A is bounded if and only if $\operatorname{diam} A < \infty$.

Example 7.27. Let X be a set with the discrete metric given in Example 7.2. Then X is bounded since $X = B_1(x)$ for any $x \in X$.

Example 7.28. Let $C(K)$ be the space of continuous functions $f : K \to \mathbb{R}$ on a compact set $K \subset \mathbb{R}$ equipped with the sup-norm. The set $F \subset C(K)$ of all functions f such that $|f(x)| \leq 1$ for every $x \in K$ is a bounded set since $\|f\| = d(f, 0) \leq 1$ for all $f \in F$.

Compact sets are sets that have the Heine-Borel property

Definition 7.29. A subset $K \subset X$ of a metric space X is compact if every open cover of K has a finite subcover.

A significant property of $\mathbb{R}$ (or $\mathbb{R}^n$) that does *not* generalize to arbitrary metric spaces is that a set is compact if and only if it is closed and bounded. In general, a compact subset of a metric space is closed and bounded; however, a closed and bounded set need not be compact.

Finally, we define some relationships of points to a set that are analogous to the ones for $\mathbb{R}$.

Definition 7.30. Let X be a metric space and $A \subset X$.

(1) A point $x \in A$ is an interior point of A if $B_r(x) \subset A$ for some $r > 0$.
(2) A point $x \in A$ is an isolated point of A if $B_r(x) \cap A = \{x\}$ for some $r > 0$, meaning that x is the only point of A that belongs to $B_r(x)$.
(3) A point $x \in X$ is a boundary point of A if, for every $r > 0$, the ball $B_r(x)$ contains points in A and points not in A.
(4) A point $x \in X$ is an accumulation point of A if, for every $r > 0$, the ball $B_r(x)$ contains a point $y \in A$ such that $y \neq x$.

A set is open if and only if every point in the set is an interior point, and a set is closed if and only if every accumulation point of the set belongs to the set.

7.4. Sequences

A sequence (x_n) in a set X is a function $f : \mathbb{N} \to X$, where we write $x_n = f(n)$ for the nth term in the sequence.

Definition 7.31. Let (X, d) be a metric space. A sequence (x_n) in X converges to $x \in X$, written $x_n \to x$ as $n \to \infty$ or

$$\lim_{n\to\infty} x_n = x,$$

if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n > N \text{ implies that } d(x_n, x) < \epsilon.$$

That is, $x_n \to x$ if $d(x_n, x) \to 0$ as $n \to \infty$. Equivalently, $x_n \to x$ as $n \to \infty$ if for every neighborhood U of x there exists $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$.

Example 7.32. For $\mathbb{R}$ with its standard absolute value metric, Definition 7.31 is just the definition of the convergence of a real sequence.

Example 7.33. Let $K \subset \mathbb{R}$ be compact. A sequence of continuous functions (f_n) in $C(K)$ converges to $f \in C(K)$ with respect to the sup-norm if and only if $f_n \to f$ as $n \to \infty$ uniformly on K.

We define closed sets in terms of sequences in the same way as for $\mathbb{R}$.

Definition 7.34. A subset $F \subset X$ of a metric space X is sequentially closed if the limit every convergent sequence (x_n) in F belongs to F.

Explicitly, this means that if (x_n) is a sequence of points $x_n \in F$ and $x_n \to x$ as $n \to \infty$ in X, then $x \in F$. A subset of a metric space is sequentially closed if and only if it is closed.

Example 7.35. Let $F \subset C(K)$ be the set of continuous functions $f : K \to \mathbb{R}$ such that $|f(x)| \le 1$ for all $x \in K$. Then F is a closed subset of $C(K)$.

We can also give a sequential definition of compactness, which generalizes the Bolzano-Weierstrass property.

Definition 7.36. A subset $K \subset X$ of a metric space X is sequentially compact if every sequence in K has a convergent subsequence whose limit belongs to K.

Explicitly, this means that if (x_n) is a sequence of points $x_n \in K$ then there is a subsequence (x_{n_k}) such that $x_{n_k} \to x$ as $k \to \infty$, and $x \in K$.

Theorem 7.37. A subset of a metric space is sequentially compact if and only if it is compact.

We can also define Cauchy sequences in a metric space.

Definition 7.38. Let (X, d) be a metric space. A sequence (x_n) in X is a Cauchy sequence for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n > N \text{ implies that } d(x_m, x_n) < \epsilon.$$

Completeness of a metric space is defined using the Cauchy condition.

Definition 7.39. A metric space is complete if every Cauchy sequence converges.

For $\mathbb{R}$, completeness is equivalent to the existence of suprema, but general metric spaces are not ordered so this property does not apply to them.

Definition 7.40. A Banach space is a complete normed vector space.

Nearly all metric and normed spaces that arise in analysis are complete.

Example 7.41. The space $(\mathbb{R}, |\cdot|)$ is a Banach space. More generally, $\mathbb{R}^n$ with the ℓ^p-norm defined in Example 7.15 is a Banach space.

Example 7.42. If $K \subset \mathbb{R}$ is compact, the space $C(K)$ with the sup-norm is a Banach space. A sequence of functions (f_n) is Cauchy in $C(K)$ if and only if it is uniformly Cauchy. Thus, Theorem 5.21 states that $C(K)$ is complete.

7.5. Continuous functions

The definitions of limits and continuity of functions between metric spaces parallel the definitions for real functions.

Definition 7.43. Let (X, d_X) and (Y, d_Y) be metric spaces, and suppose that $c \in X$ is an accumulation point of X. If $f : X \setminus \{c\} \to Y$, then $y \in Y$ is the limit of $f(x)$ as $x \to c$, or

$$\lim_{x \to c} f(x) = y,$$

if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$0 < d_X(x, c) < \delta \text{ implies that } d_Y(f(x), y) < \epsilon.$$

In terms of neighborhoods, the definition says that for every neighborhood V of y in Y there exists a neighborhood U of c in X such that f maps $U \setminus \{c\}$ into V.

Definition 7.44. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous at $c \in X$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$d_X(x, c) < \delta \text{ implies that } d_Y(f(x), f(c)) < \epsilon.$$

The function is continuous on X if it is continuous at every point of X.

In terms of neighborhoods, the definition says that for every neighborhood V of $f(c)$ in Y there exists a neighborhood U of c in X such that f maps U into V.

Example 7.45. A function $f : \mathbb{R}^2 \to \mathbb{R}$, where $\mathbb{R}^2$ is equipped with the Euclidean norm $\|\cdot\|$ and $\mathbb{R}$ with the absolute value norm $|\cdot|$, is continuous at $c \in \mathbb{R}^2$ if

$$\|x - c\| < \delta \text{ implies that } |f(x) - f(c)| < \epsilon$$

Explicitly, if $x = (x_1, x_2)$, $c = (c_1, c_2)$ and

$$f(x) = (f_1(x_1, x_2), f_2(x_1, x_2)),$$

this condition reads:

$$\sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2} < \delta$$

implies that

$$|f(x_1, x_2) - f(c_1, c_2)| < \epsilon.$$

Example 7.46. A function $f : \mathbb{R} \to \mathbb{R}^2$, where $\mathbb{R}^2$ is equipped with the Euclidean norm $\|\cdot\|$ and $\mathbb{R}$ with the absolute value norm $|\cdot|$, is continuous at $c \in \mathbb{R}^2$ if

$$|x - c| < \delta \text{ implies that } \|f(x) - f(c)\| < \epsilon$$

Explicitly, if $f(x) = (f_1(x), f_2(x))$, where where $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, this condition reads: $|x - c| < \delta$ implies that

$$\sqrt{[f_1(x) - f_1(c)]^2 + [f_1(x) - f_1(c)]^2} < \epsilon.$$

The previous examples generalize in a natural way to define the continuity of an m-component vector-valued function of n variables.

Example 7.47. A function $f : \mathbb{R}^n \to \mathbb{R}^m$, where both $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with the Euclidean norm, is continuous at c if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$\|x - c\| < \delta \text{ implies that } \|f(x) - f(c)\| < \epsilon.$$

This definition would look complicated if it was written out explicitly, but it is much clearer when it is expressed in terms or metrics or norms.

We also have a sequential definition of continuity in a metric space.

Definition 7.48. Let X and Y be metric spaces. A function $f : X \to Y$ is sequentially continuous at $c \in X$ if

$$f(x_n) \to f(c) \qquad \text{as } n \to \infty$$

for every sequence (x_n) in X such that

$$x_n \to c \qquad \text{as } n \to \infty$$

As for real functions, this is equivalent to continuity.

Proposition 7.49. A function $f : X \to Y$ is sequentially continuous at $c \in X$ if and only if it is continuous at c.

We define uniform continuity similarly.

Definition 7.50. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is uniformly continuous on X if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$d_X(x, y) < \delta \text{ implies that } d_Y(f(x), f(y)) < \epsilon.$$

The proofs of the following theorems are identical to the proofs we gave for functions $f : \mathbb{R} \to \mathbb{R}$.

First, a function on a metric space is continuous if and only if the inverse images of open sets are open.

Theorem 7.51. A function $f : X \to Y$ between metric spaces X and Y is continuous on X if and only if $f^{-1}(V)$ is open in X for every open set V in Y.

Second, the continuous image of a compact set is compact.

Theorem 7.52. Let K be a compact metric space and Y a metric space. If $f : K \to Y$ is a continuous function, then $f(K)$ is a compact subset of Y.

Third, a continuous functions on a compact set is uniformly continuous.

Theorem 7.53. If $f : K \to Y$ is a continuous function on a compact set K, then f is uniformly continuous.

7.6. Appendix: The Minkowski inequality

Inequalities are essential to analysis. Their proofs, however, are often not obvious and may require considerable ingenuity. Moreover, there may be many different ways to prove the same inequality.

The triangle inequality for the ℓ^p-norm on $\mathbb{R}^n$ defined in Example 7.15 is called the Minkowski inequality, and it is one of the most important inequalities in analysis. In this section, we prove it in the Euclidean case $p = 2$. The general case, with arbitrary $1 < p < \infty$, is more involved, and we will not prove it here.

We first prove the Cauchy-Schwartz inequality, which is itself a fundamental inequality.

Theorem 7.54 (Cauchy-Schwartz)**.** If $(x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, then

$$\left| \sum_{i=1}^n x_i y_i \right| \le \left(\sum_{i=1}^n x_i^2 \right)^{1/2} \left(\sum_{i=1}^n y_i^2 \right)^{1/2} .$$

Proof. Since $|\sum x_i y_i| \le \sum |x_i|\,|y_i|$, it is sufficient to prove the inequality for $x_i, y_i \ge 0$. Furthermore, the inequality is obvious if either $x = 0$ or $y = 0$, so we assume at least one x_i and one y_i is nonzero.

For every $\alpha, \beta \in \mathbb{R}$, we have

$$0 \le \sum_{i=1}^n (\alpha x_i - \beta y_i)^2 .$$

Expanding the square on the right-hand side and rearranging the terms, we get that

$$2\alpha\beta \sum_{i=1}^n x_i y_i \le \alpha^2 \sum_{i=1}^n x_i^2 + \beta^2 \sum_{i=1}^n y_i^2 .$$

We choose $\alpha, \beta > 0$ to "balance" the terms on the right-hand side,

$$\alpha = \left(\sum_{i=1}^n y_i^2 \right)^{1/2}, \qquad \beta = \left(\sum_{i=1}^n x_i^2 \right)^{1/2} .$$

Then division of the resulting inequality by $2\alpha\beta$ proves the theorem. □

The Minkowski inequality for $p = 2$ is an immediate consequence of the Cauchy-Schwartz inequality.

Corollary 7.55. If $(x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, then

$$\left[\sum_{i=1}^n (x_i + y_i)^2 \right]^{1/2} \le \left(\sum_{i=1}^n x_i^2 \right)^{1/2} + \left(\sum_{i=1}^n y_i^2 \right)^{1/2} .$$

Proof. Expanding the square in the following equation and using the Cauchy-Schwartz inequality, we have

$$\begin{aligned}\sum_{i=1}^{n}(x_i+y_i)^2 &= \sum_{i=1}^{n}x_i^2+2\sum_{i=1}^{n}x_iy_i+\sum_{i=1}^{n}y_i^2\\ &\le \sum_{i=1}^{n}x_i^2+2\left(\sum_{i=1}^{n}x_i^2\right)^{1/2}\left(\sum_{i=1}^{n}y_i^2\right)^{1/2}+\sum_{i=1}^{n}y_i^2\\ &\le \left[\left(\sum_{i=1}^{n}x_i^2\right)^{1/2}+\left(\sum_{i=1}^{n}y_i^2\right)^{1/2}\right]^2.\end{aligned}$$

Taking the square root of this inequality, we get the result. □