Chapter 13

Metric, Normed, and Topological Spaces

A metric space is a set X that has a notion of the distance $d(x, y)$ between every pair of points $x, y \in X$. A fundamental example is $\mathbb{R}$ with the absolute-value metric $d(x, y) = |x - y|$, and nearly all of the concepts we discuss below for metric spaces are natural generalizations of the corresponding concepts for $\mathbb{R}$.

A special type of metric space that is particularly important in analysis is a normed space, which is a vector space whose metric is derived from a norm. On the other hand, every metric space is a special type of topological space, which is a set with the notion of an open set but not necessarily a distance.

The concepts of metric, normed, and topological spaces clarify our previous discussion of the analysis of real functions, and they provide the foundation for wide-ranging developments in analysis. The aim of this chapter is to introduce these spaces and give some examples, but their theory is too extensive to describe here in any detail.

13.1. Metric spaces

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

Definition 13.1. A metric d on a set X is a function $d : X \times X \to \mathbb{R}$ such that for all $x, y, z \in X$:

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

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

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

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

In general, many different metrics can be defined on the same set X, but if the metric on X is clear from the context, we refer to X as a metric space.

Subspaces of a metric space are subsets whose metric is obtained by restricting the metric on the whole space.

Definition 13.2. Let (X, d) be a metric space. A metric 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 intrinsic properties of a subset $A \subset X$ of a metric space X in terms of properties of the corresponding metric subspace (A, d_A).

When it is clear that we are discussing metric spaces, we refer to a metric subspace as a subspace, but metric subspaces should not be confused with other types of subspaces (for example, vector subspaces of a vector space).

13.1.1. Examples.

In the following examples of metric spaces, the verification of the properties of a metric is mostly straightforward and is left as an exercise.

Example 13.3. A rather trivial example of a metric on any set X is the discrete metric

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

This metric is nevertheless useful in illustrating the definitions and providing counter-examples.

Example 13.4. Define $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

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

Then d is a metric on $\mathbb{R}$. The natural numbers $\mathbb{N}$ and the rational numbers $\mathbb{Q}$ with the absolute-value metric are metric subspaces of $\mathbb{R}$, as is any other subset $A \subset \mathbb{R}$.

Example 13.5. 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. (Here, "ℓ^1" is pronounced "ell-one.") For example, writing $z = (z_1, z_2)$, we have

$$\begin{aligned} d(x, y) &= |x_1 - z_1 + z_1 - y_1| + |x_2 - z_2 + z_2 - y_2| \\ &\leq |x_1 - z_1| + |z_1 - y_1| + |x_2 - z_2| + |z_2 - y_2| \\ &\leq d(x, z) + d(z, y), \end{aligned}$$

so d satisfies the triangle inequality. This metric is sometimes referred to informally as the "taxicab" metric, since it's the distance one would travel by taxi on a rectangular grid of streets.

Example 13.6. 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 an analytical proof is non-trivial (see Theorem 13.26 below).

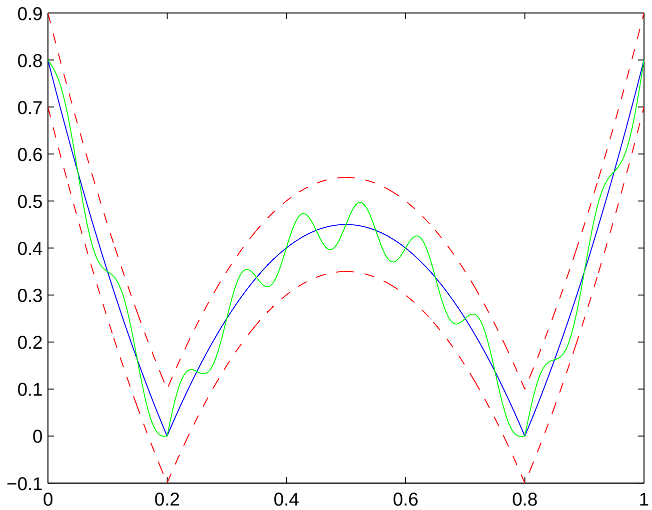

Figure 1. The graph of a function $f \in C([0,1])$ is in blue. A function whose distance from f with respect to the sup-norm is less than 0.1 has a graph that lies inside the dotted red lines $y = f(x) \pm 0.1$ e.g., the green graph.

Example 13.7. 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 13.8. 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 England, d is sometimes called the "British Rail" metric, because all the train lines radiate from London (located at 0). 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 13.9. Let $C(K)$ denote the set of continuous functions $f : K \to \mathbb{R}$, where $K \subset \mathbb{R}$ is compact; for example, $K = [a,b]$ is a closed, bounded interval. If $f, g \in C(K)$ define

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

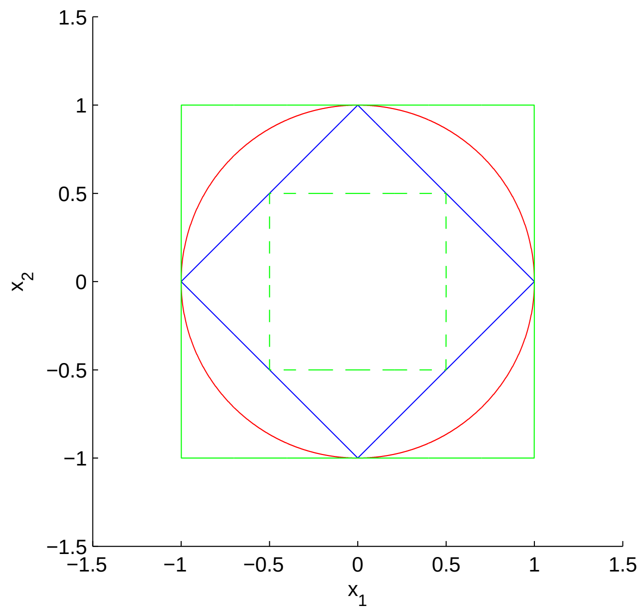

Figure 2. The unit balls $B_1(0)$ on $\mathbb{R}^2$ for different metrics: they are the interior of a diamond (ℓ^1-norm), a circle (ℓ^2-norm), or a square (ℓ^∞-norm). The ℓ^∞-ball of radius $1/2$ is also indicated by the dashed line.

The function $d : C(K) \times C(K) \to \mathbb{R}$ is well-defined, since a continuous function on a compact set is bounded, and d is a metric on $C(K)$. Two functions are close with respect to this metric if their values are close at every point $x \in K$. (See Figure 1.) We refer to $\|f\|_\infty$ as the sup-norm of f. Section 13.6 has further discussion.

13.1.2. Open and closed balls. A ball in a metric space is analogous to an interval in $\mathbb{R}$.

Definition 13.10. Let (X, d) be a metric space. The open ball $B_r(x)$ of radius $r > 0$ and center $x \in X$ is the set of points whose distance from x is less than r,

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

The closed ball $\bar{B}_r(x)$ of radius $r > 0$ and center $x \in X$ as the set of points whose distance from x is less than or equal to r,

$$\bar{B}_r(x) = \{y \in X : d(x, y) \leq r\}.$$

The term "ball" is used to denote a "solid ball," rather than the "sphere" of points whose distance from the center x is equal to r.

Example 13.11. Consider $\mathbb{R}$ with its standard absolute-value metric, defined in Example 13.4. Then the open ball $B_r(x) = \{y \in \mathbb{R} : |x - y| < r\}$ is the open interval of radius r centered at x, and the closed ball $\bar{B}_r(x) = \{y \in \mathbb{R} : |x - y| \leq r\}$ is the closed interval of radius r centered at x.

Example 13.12. For $\mathbb{R}^2$ with the Euclidean metric defined in Example 13.6, the ball $B_r(x)$ is an open disc of radius r centered at x. For the ℓ^1-metric in Example 13.5, the ball is a diamond of diameter $2r$, and for the ℓ^∞-metric in Example 13.7, it is a square of side $2r$. The unit ball $B_1(0)$ for each of these metrics is illustrated in Figure 2.

Example 13.13. Consider the space $C(K)$ of continuous functions $f : K \to \mathbb{R}$ on a compact set $K \subset \mathbb{R}$ with the sup-norm metric defined in Example 13.9. The ball $B_r(f)$ consists of all continuous functions $g : K \to \mathbb{R}$ whose values are within r of the values of f at every $x \in K$. For example, for the function f shown in Figure 1 with $r = 0.1$, the open ball $B_r(f)$ consists of all continuous functions g whose graphs lie between the red lines.

One has to be a little careful with the notion of balls in a general metric space, because they don't always behave the way their name suggests.

Example 13.14. Let X be a set with the discrete metric given in Example 13.3. 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 $r \ge 1$. (See also Example 13.44.)

An another example, what are the open balls for the metric in Example 13.8?

A set in a metric space is bounded if it is contained in a ball of finite radius.

Definition 13.15. 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 $d(x, y) \le R$ for all $y \in A$, meaning that $A \subset B_R(x)$.

Unlike $\mathbb{R}$, or a vector space, a general metric space has no distinguished origin, but the center point of the ball is not important in this definition of a bounded set. The triangle inequality implies that $d(y, z) < R + d(x, y)$ if $d(x, z) < R$, so

$$B_R(x) \subset B_{R'}(y) \qquad \text{for } R' = R + d(x, y).$$

Thus, if Definition 13.15 holds for some $x \in X$, then it holds for every $x \in X$.

We can say equivalently that $A \subset X$ is bounded if the metric subspace (A, d_A) is bounded.

Example 13.16. Let X be a set with the discrete metric given in Example 13.3. Then X is bounded since $X = B_r(x)$ if $r > 1$ and $x \in X$.

Example 13.17. A subset $A \subset \mathbb{R}$ is bounded with respect to the absolute-value metric if $A \subset (-R, R)$ for some $0 < R < \infty$.

Example 13.18. Let $C(K)$ be the space of continuous functions $f : K \to \mathbb{R}$ on a compact set defined in Example 13.9. The set $\mathcal{F} \subset C(K)$ of all continuous functions $f : K \to \mathbb{R}$ such that $|f(x)| \le 1$ for every $x \in K$ is bounded, since $d(f, 0) = \|f\|_\infty \le 1$ for all $f \in \mathcal{F}$. The set of constant functions $\{f : f(x) = c \text{ for all } x \in K\}$ isn't bounded, since $\|f\|_\infty = |c|$ may be arbitrarily large.

We define the diameter of a set in an analogous way to Definition 3.5 for subsets of $\mathbb{R}$.

Definition 13.19. Let (X, d) be a metric space and $A \subset X$. The diameter $0 \le \operatorname{diam} A \le \infty$ of A is

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

It follows from the definitions that A is bounded if and only if $\operatorname{diam} A < \infty$.

The notions of an upper bound, lower bound, supremum, and infimum in $\mathbb{R}$ depend on its order properties. Unlike properties of $\mathbb{R}$ based on the absolute value, they do not generalize to an arbitrary metric space, which isn't equipped with an order relation.

13.2. Normed spaces

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.

We assume that the reader is familiar with the basic theory of vector spaces, and we consider only real vector spaces.

Definition 13.20. A normed vector space $(X, \|\cdot\|)$ is a vector space X 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) $0 \le \|x\| < \infty$ and $\|x\| = 0$ if and only if $x = 0$;
(2) $\|kx\| = |k|\|x\|$;
(3) $\|x + y\| \le \|x\| + \|y\|$.

The properties in Definition 13.20 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 a scalar 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 called the triangle inequality. We also refer to a normed vector space as a normed space for short.

Proposition 13.21. *If $(X, \|\cdot\|)$ is a normed vector space, 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 directly from the properties of a norm in Definition 13.20. The positivity is immediate. Also, we have

$$d(x, y) = \|x - y\| = \|-(x - y)\| = \|y - x\| = d(y, x),$$
$$d(x, y) = \|x - z + z - y\| \le \|x - z\| + \|z - y\| = d(x, z) + d(y, z),$$

which proves the symmetry of d and the triangle inequality. □

If X is a normed vector space, then we always use the metric associated with its norm, unless stated specifically otherwise.

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 imply that the open balls $B_r(x)$ in a normed space are rescaled, translated versions of the unit ball $B_1(0)$.

Example 13.22. The set of real numbers $\mathbb{R}$ with the absolute-value norm $|\cdot|$ is a one-dimensional normed vector space.

Example 13.23. The discrete metric in Example 13.3 on $\mathbb{R}$, and the metric in Example 13.8 on $\mathbb{R}^2$ are not derived from a norm. (Why?)

Example 13.24. The space $\mathbb{R}^2$ with any of the norms defined for $x = (x_1, x_2)$ by

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

is a normed vector space. The corresponding metrics are the "taxicab" metric in Example 13.5, the Euclidean metric in Example 13.6, and the maximum metric in Example 13.7, respectively.

The norms in Example 13.24 are special cases of a fundamental family of ℓ^p-norms on $\mathbb{R}^n$. All of the ℓ^p-norms reduce to the absolute-value norm if $n = 1$, but they are different if $n \geq 2$.

Definition 13.25. For $1 \leq p < \infty$, the ℓ^p-norm $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$ is defined for $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ by

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

The ℓ^2-norm is called the Euclidean norm. For $p = \infty$, the ℓ^∞-norm $\|\cdot\|_\infty : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\|x\|_\infty = \max\left(|x_1|, |x_2|, \ldots, |x_n|\right).$$

The notation for the ℓ^∞-norm is explained by the fact that

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p.$$

Moreover, consistent with its name, the ℓ^p-norm is a norm.

Theorem 13.26. Let $1 \leq p \leq \infty$. The space $\mathbb{R}^n$ with the ℓ^p-norm is a normed vector space.

Proof. The space $\mathbb{R}^n$ is an n-dimensional vector space, so we just need to verify the properties of the norm.

The positivity and homogeneity of the ℓ^p-norm follow immediately from its definition. We verify the triangle inequality here only for the cases $p = 1, \infty$.

Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be points in $\mathbb{R}^n$. For $p = 1$, we have

$$\begin{aligned}
\|x + y\|_1 &= |x_1 + y_1| + |x_2 + y_2| + \cdots + |x_n + y_n| \\
&\leq |x_1| + |y_1| + |x_2| + |y_2| + \cdots + |x_n| + |y_n| \\
&\leq \|x\|_1 + \|y\|_1.
\end{aligned}$$

For $p = \infty$, we have

$$\begin{aligned}\|x+y\|_\infty &= \max\left(|x_1+y_1|, |x_2+y_2|, \dots, |x_n+y_n|\right)\\ &\le \max\left(|x_1|+|y_1|, |x_2|+|y_2|, \dots, |x_n|+|y_n|\right)\\ &\le \max\left(|x_1|, |x_2|, \dots, |x_n|\right) + \max\left(|y_1|, |y_2|, \dots, |y_n|\right)\\ &\le \|x\|_\infty + \|y\|_\infty.\end{aligned}$$

The proof of the triangle inequality for $1 < p < \infty$ is more difficult and is given in Section 13.7. □

We can use Definition 13.25 to define $\|x\|_p$ for any $0 < p \le \infty$. However, if $0 < p < 1$, then $\|\cdot\|_p$ doesn't satisfy the triangle inequality, so it is not a norm. This explains the restriction $1 \le p \le \infty$.

Although the ℓ^p-norms are numerically different for different values of p, they are equivalent in the following sense (see Corollary 13.29).

Definition 13.27. Let X be a vector space. Two norms $\|\cdot\|_a$, $\|\cdot\|_b$ on X are *equivalent* if there exist strictly positive constants $M \ge m > 0$ such that

$$m\|x\|_a \le \|x\|_b \le M\|x\|_a \qquad \text{for all } x \in X.$$

Geometrically, two norms are equivalent if and only if an open ball with respect to either one of the norms contains an open ball with respect to the other. Equivalent norms define the same open sets, convergent sequences, and continuous functions, so there are no topological differences between them.

The next theorem shows that every ℓ^p-norm is equivalent to the ℓ^∞-norm. (See Figure 2.)

Theorem 13.28. *Suppose that $1 \le p < \infty$. Then, for every $x \in \mathbb{R}^n$,*

$$\|x\|_\infty \le \|x\|_p \le n^{1/p}\|x\|_\infty.$$

Proof. Let $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. Then for each $1 \le i \le n$, we have

$$|x_i| \le \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p} = \|x\|_p,$$

which implies that

$$\|x\|_\infty = \max\left\{|x_i| : 1 \le i \le n\right\} \le \|x\|_p.$$

On the other hand, since $|x_i| \le \|x\|_\infty$ for every $1 \le i \le n$, we have

$$\|x\|_p \le \left(n\|x\|_\infty^p\right)^{1/p} = n^{1/p}\|x\|_\infty,$$

which proves the result □

As an immediate consequence, we get the equivalence of the ℓ^p-norms.

Corollary 13.29. *The ℓ^p and ℓ^q norms on $\mathbb{R}^n$ are equivalent for every $1 \le p, q \le \infty$.*

Proof. We have $n^{-1/q}\|x\|_q \le \|x\|_\infty \le \|x\|_p \le n^{1/p}\|x\|_\infty \le n^{1/p}\|x\|_q$. □

With more work, one can prove that that $\|x\|_q \le \|x\|_p$ for $1 \le p \le q \le \infty$, meaning that the unit ball with respect to the ℓ^q-norm contains the unit ball with respect to the ℓ^p-norm.

13.3. Open and closed sets

There are natural definitions of open and closed sets in a metric space, analogous to the definitions in $\mathbb{R}$. Many of the properties of such sets in $\mathbb{R}$ carry over immediately to general metric spaces.

Definition 13.30. 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 $B_r(x) \subset G$. A subset $F \subset X$ is closed if $F^c = X \setminus F$ is open.

We can rephrase this definition more geometrically in terms of neighborhoods.

Definition 13.31. 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 $r > 0$.

Definition 13.30 then states that a subset of a metric space is open if and only if every point in the set has a neighborhood that is contained in the set. In particular, a set is open if and only if it is a neighborhood of every point in the set.

Example 13.32. If d is the discrete metric on a set X in Example 13.3, then every subset $A \subset X$ is open, since for every $x \in A$ we have $B_{1/2}(x) = \{x\} \subset A$. Every set is also closed, since its complement is open.

Example 13.33. Consider $\mathbb{R}^2$ with the Euclidean norm (or any other ℓ^p-norm). If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, then

$$E = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_2 < f(x_1)\right\}$$

is an open subset of $\mathbb{R}^2$. If f is discontinuous, then E needn't be open. We leave the proofs as an exercise.

Example 13.34. If (X, d) is a metric space and $A \subset X$, then $B \subset A$ is open in the metric subspace (A, d_A) if and only if $B = A \cap G$ where G is an open subset of X. This is consistent with our previous definition of relatively open sets in $A \subset \mathbb{R}$.

Open sets with respect to one metric on a set need not be open with respect to another metric. For example, every subset of $\mathbb{R}$ with the discrete metric is open, but this is not true of $\mathbb{R}$ with the absolute-value metric.

Consistent with our terminology, open balls are open and closed balls are closed.

Proposition 13.35. Let X be a metric space. If $x \in X$ and $r > 0$, then the open ball $B_r(x)$ is open and the closed ball $\bar{B}_r(x)$ is closed.

Proof. Suppose that $y \in B_r(x)$ where $r > 0$, and let $\epsilon = r - d(x, y) > 0$. The triangle inequality implies that $B_\epsilon(y) \subset B_r(x)$, which proves that $B_r(x)$ is open. Similarly, if $y \in \bar{B}_r(x)^c$ and $\epsilon = d(x, y) - r > 0$, then the triangle inequality implies that $B_\epsilon(y) \subset \bar{B}_r(x)^c$, which proves that $\bar{B}_r(x)^c$ is open and $\bar{B}_r(x)$ is closed. □

The next theorem summarizes three basic properties of open sets.

Theorem 13.36. Let X be a metric space.

(1) The empty set $\varnothing$ and the whole set X are open.

(2) An arbitrary union of open sets is open.

(3) A finite intersection of open sets is open.

Proof. Property (1) follows immediately from Definition 13.30. (The empty set satisfies the definition vacuously: since it has no points, every point has a neighborhood that is contained in the set.)

To prove (2), let $\{G_i \subset X : i \in I\}$ be an arbitrary collection of open sets. If

$$x \in \bigcup_{i\in I} G_i,$$

then $x \in G_i$ for some $i \in I$. Since G_i is open, there exists $r > 0$ such that $B_r(x) \subset G_i$, and then

$$B_r(x) \subset \bigcup_{i\in I} G_i.$$

Thus, the union $\bigcup G_i$ is open.

The prove (3), let $\{G_1, G_2, \ldots, G_n\}$ be a finite collection of open sets. If

$$x \in \bigcap_{i=1}^{n} G_i,$$

then $x \in G_i$ for every $1 \le i \le n$. Since G_i is open, there exists $r_i > 0$ such that $B_{r_i}(x) \subset G_i$. Let $r = \min(r_1, r_2, \ldots, r_n) > 0$. Then $B_r(x) \subset B_{r_i}(x) \subset G_i$ for every $1 \le i \le n$, which implies that

$$B_r(x) \subset \bigcap_{i=1}^{n} G_i.$$

Thus, the finite intersection $\bigcap G_i$ is open. □

The previous proof fails if we consider the intersection of infinitely many open sets $\{G_i : i \in I\}$ because we may have $\inf\{r_i : i \in I\} = 0$ even though $r_i > 0$ for every $i \in I$.

The properties of closed sets follow by taking complements of the corresponding properties of open sets and using De Morgan's laws, exactly as in the proof of Proposition 5.20.

Theorem 13.37. Let X be a metric space.

(1) The empty set $\varnothing$ and the whole set X are closed.
(2) An arbitrary intersection of closed sets is closed.
(3) A finite union of closed sets is closed.

The following relationships of points to sets are entirely analogous to the ones in Definition 5.22 for $\mathbb{R}$.

Definition 13.38. 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 a point in A and a point 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 is an interior point and closed if and only if every accumulation point belongs to the set.

We define the interior, boundary, and closure of a set as follows.

Definition 13.39. Let A be a subset of a metric space. The interior A° of A is the set of interior points of A. The boundary ∂A of A is the set of boundary points. The closure of A is $\bar{A} = A \cup \partial A$.

It follows that $x \in \bar{A}$ if and only if the ball $B_r(x)$ contains some point in A for every $r > 0$. The next proposition gives equivalent topological definitions.

Proposition 13.40. Let X be a metric space and $A \subset X$. The interior of A is the largest open set contained in A,

$$A^\circ = \bigcup \{G \subset A : G \text{ is open in } X\},$$

the closure of A is the smallest closed set that contains A,

$$\bar{A} = \bigcap \{F \supset A : F \text{ is closed in } X\},$$

and the boundary of A is their set-theoretic difference,

$$\partial A = \bar{A} \setminus A^\circ.$$

Proof. Let A_1 denote the set of interior points of A, as in Definition 13.38, and $A_2 = \bigcup \{G \subset A : G \text{ is open}\}$. If $x \in A_1$, then there is an open neighborhood $G \subset A$ of x, so $G \subset A_2$ and $x \in A_2$. It follows that $A_1 \subset A_2$. To get the opposite inclusion, note that A_2 is open by Theorem 13.36. Thus, if $x \in A_2$, then $A_2 \subset A$ is a neighborhood of x, so $x \in A_1$ and $A_2 \subset A_1$. Therefore $A_1 = A_2$, which proves the result for the interior.

Next, Definition 13.38 and the previous result imply that

$$(\bar{A})^c = (A^c)^\circ = \bigcup \{G \subset A^c : G \text{ is open}\}.$$

Using De Morgan's laws, and writing $G^c = F$, we get that

$$\bar{A} = \bigcup \{G \subset A^c : G \text{ is open}\}^c = \bigcap \{F \supset A : F \text{ is closed}\},$$

which proves the result for the closure.

Finally, if $x \in \partial A$, then $x \in \bar{A} = A \cup \partial A$, and no neighborhood of x is contained in A, so $x \notin A^\circ$. It follows that $x \in \bar{A} \setminus A^\circ$ and $\partial A \subset \bar{A} \setminus A^\circ$. Conversely, if $x \in \bar{A} \setminus A^\circ$, then every neighborhood of x contains points in A, since $x \in \bar{A}$, and every neighborhood contains points in A^c, since $x \notin A^\circ$. It follows that $x \in \partial A$ and $\bar{A} \setminus A^\circ \subset \partial A$, which completes the proof. □

It follows from Theorem 13.36, and Theorem 13.37 that the interior A° is open, the closure $\bar{A}$ is closed, and the boundary ∂A is closed. Furthermore, A is open if and only if $A = A^\circ$, and A is closed if and only if $A = \bar{A}$.

Let us illustrate these definitions with some examples, whose verification we leave as an exercise.

Example 13.41. Consider $\mathbb{R}$ with the absolute-value metric. If $I = (a, b)$ and $J = [a, b]$, then $I^\circ = J^\circ = (a, b)$, $\bar{I} = \bar{J} = [a, b]$, and $\partial I = \partial J = \{a, b\}$. Note that $I = I^\circ$, meaning that I is open, and $J = \bar{J}$, meaning that J is closed. If $A = \{1/n : n \in \mathbb{N}\}$, then $A^\circ = \varnothing$ and $\bar{A} = \partial A = A \cup \{0\}$. Thus, A is neither open ($A \neq A^\circ$) nor closed ($A \neq \bar{A}$). If $\mathbb{Q}$ is the set of rational numbers, then $\mathbb{Q}^\circ = \varnothing$ and $\bar{\mathbb{Q}} = \partial\mathbb{Q} = \mathbb{R}$. Thus, $\mathbb{Q}$ is neither open nor closed. Since $\bar{\mathbb{Q}} = \mathbb{R}$, we say that $\mathbb{Q}$ is dense in $\mathbb{R}$.

Example 13.42. Let A be the unit open ball in $\mathbb{R}^2$ with the Euclidean metric,

$$A = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\right\}.$$

Then $A^\circ = A$, the closure of A is the closed unit ball

$$\bar{A} = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\},$$

and the boundary of A is the unit circle

$$\partial A = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\right\},$$

Example 13.43. Let A be the unit open ball with the x-axis deleted in $\mathbb{R}^2$ with the Euclidean metric,

$$A = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1,\, y \neq 0\right\}.$$

Then $A^\circ = A$, the closure of A is the closed unit ball

$$\bar{A} = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\},$$

and the boundary of A consists of the unit circle and the x-axis,

$$\partial A = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\right\} \cup \left\{(x, 0) \in \mathbb{R}^2 : |x| \leq 1\right\}.$$

Example 13.44. Suppose that X is a set containing at least two elements with the discrete metric defined in Example 13.3. If $x \in X$, then the unit open ball is $B_1(x) = \{x\}$, and it is equal to its closure $\overline{B_r(x)} = \{x\}$. On the other hand, the closed unit ball is $\bar{B}_1(x) = X$. Thus, in a general metric space, the closure of an open ball of radius $r > 0$ need not be the closed ball of radius r.

13.4. Completeness, compactness, and continuity

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

Definition 13.45. 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, $x_n \to x$ in X if $d(x_n, x) \to 0$ in $\mathbb{R}$. 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 13.46. If d is the discrete metric on a set X, then a sequence (x_n) converges in (X, d) if and only if it is eventually constant. That is, there exists $x \in X$ and $N \in \mathbb{N}$ such that $x_n = x$ for all $n > N$; and, in that case, the sequence converges to x.

Example 13.47. For $\mathbb{R}$ with its standard absolute-value metric, Definition 13.45 is the definition of the convergence of a real sequence.

As for subsets of $\mathbb{R}$, we can give a sequential characterization of closed sets in a metric space.

Theorem 13.48. A subset $F \subset X$ of a metric space X is closed if and only if the limit of every convergent sequence in F belongs to F.

Proof. First suppose that F is closed, meaning that F^c is open. If (x_n) be a sequence in F and $x \in F^c$, then there is a neighborhood $U \subset F^c$ of x which contains no terms in the sequence, so (x_n) cannot converge to x. Thus, the limit of every convergent sequence in F belongs to F.

Conversely, suppose that F is not closed. Then F^c is not open, and there exists a point $x \in F^c$ such that every neighborhood of x contains points in F. Choose $x_n \in F$ such that $x_n \in B_{1/n}(x)$. Then (x_n) is a sequence in F whose limit x does not belong to F, which proves the result. $\square$

We define the completeness of metric spaces in terms of Cauchy sequences.

Definition 13.49. 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

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

Every convergent sequence is Cauchy: if $x_n \to x$ then given $\epsilon > 0$ there exists N such that $d(x_n, x) < \epsilon/2$ for all $n > N$, and then for all $m, n > N$ we have

$$d(x_m, x_n) \leq d(x_m, x) + d(x_n, x) < \epsilon.$$

Complete spaces are ones in which the converse is also true.

Definition 13.50. A metric space is complete if every Cauchy sequence converges.

Example 13.51. If d is the discrete metric on a set X, then (X, d) is a complete metric space since every Cauchy sequence is eventually constant.

Example 13.52. The space $(\mathbb{R}, |\cdot|)$ is complete, but the metric subspace $(\mathbb{Q}, |\cdot|)$ is not complete.

In a complete space, we have the following simple criterion for the completeness of a subspace.

Proposition 13.53. A subspace (A, d_A) of a complete metric space (X, d) is complete if and only if A is closed in X.

Proof. If A is a closed subspace of a complete space X and (x_n) is a Cauchy sequence in A, then (x_n) is a Cauchy sequence in X, so it converges to $x \in X$. Since A is closed, $x \in A$, which shows that A is complete.

Conversely, if A is not closed, then by Proposition 13.48 there is a convergent sequence in A whose limit does not belong to A. Since it converges, the sequence is Cauchy, but it doesn't have a limit in A, so A is not complete. □

The most important complete metric spaces in analysis are the complete normed spaces, or Banach spaces.

Definition 13.54. A Banach space is a complete normed vector space.

For example, $\mathbb{R}$ with the absolute-value norm is a one-dimensional Banach space. Furthermore, it follows from the completeness of $\mathbb{R}$ that every finite-dimensional normed vector space over $\mathbb{R}$ is complete. We prove this for the ℓ^p-norms given in Definition 13.25.

Theorem 13.55. Let $1 \le p \le \infty$. The vector space $\mathbb{R}^n$ with the ℓ^p-norm is a Banach space.

Proof. Suppose that $(x_k)_{k=1}^{\infty}$ is a sequence of points

$$x_k = (x_{1,k}, x_{2,k}, \dots, x_{n,k})$$

in $\mathbb{R}^n$ that is Cauchy with respect to the ℓ^p-norm. From Theorem 13.28,

$$|x_{i,j} - x_{i,k}| \le \|x_j - x_k\|_p,$$

so each coordinate sequence $(x_{i,k})_{k=1}^{\infty}$ is Cauchy in $\mathbb{R}$. The completeness of $\mathbb{R}$ implies that $x_{i,k} \to x_i$ as $k \to \infty$ for some $x_i \in \mathbb{R}$. Let $x = (x_1, x_2, \dots, x_n)$. Then, from Theorem 13.28 again,

$$\|x_k - x\|_p \le C \max\{|x_{i,k} - x_i| : i = 1, 2, \dots, n\},$$

where $C = n^{1/p}$ if $1 \le p < \infty$ or $C = 1$ if $p = \infty$. Given $\epsilon > 0$, choose $N_i \in \mathbb{N}$ such that $|x_{i,k} - x_i| < \epsilon/C$ for all $k > N_i$, and let $N = \max\{N_1, N_2, \dots, N_n\}$. Then $k > N$ implies that $\|x_k - x\|_p < \epsilon$, which proves that $x_k \to x$ with respect to the ℓ^p-norm. Thus, $(\mathbb{R}^n, \|\cdot\|_p)$ is complete. □

The Bolzano-Weierstrass property provides a sequential definition of compactness in a general metric space.

Definition 13.56. A subset $K \subset X$ of a metric space X is sequentially compact, or compact for short, if every sequence in K has a convergent subsequence whose limit belongs to K.

Explicitly, this condition 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 $k \to \infty$ where $x \in K$. Compactness is an intrinsic property of a subset: $K \subset X$ is compact if and only if the metric subspace (K, d_K) is compact.

Although this definition is similar to the one for compact sets in $\mathbb{R}$, there is a significant difference between compact sets in a general metric space and in $\mathbb{R}$. Every compact subset of a metric space is closed and bounded, as in $\mathbb{R}$, but it is not always true that a closed, bounded set is compact.

First, as the following example illustrates, a set must be complete, not just closed, to be compact. (A closed subset of $\mathbb{R}$ is complete because $\mathbb{R}$ is complete.)

Example 13.57. Consider the metric space $\mathbb{Q}$ with the absolute value norm. The set $[0, 2] \cap \mathbb{Q}$ is a closed, bounded subspace, but it is not compact since a sequence of rational numbers that converges in $\mathbb{R}$ to an irrational number such as $\sqrt{2}$ has no convergent subsequence in $\mathbb{Q}$.

Second, completeness and boundedness is not enough, in general, to imply compactness.

Example 13.58. Consider $\mathbb{N}$, or any other infinite set, with the discrete metric,

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

Then $\mathbb{N}$ is complete and bounded with respect to this metric. However, it is not compact since $x_n = n$ is a sequence with no convergent subsequence, as is clear from Example 13.46.

The correct generalization to an arbitrary metric space of the characterization of compact sets in $\mathbb{R}$ as closed and bounded replaces "closed" with "complete" and "bounded" with "totally bounded," which is defined as follows.

Definition 13.59. Let (X, d) be a metric space. A subset $A \subset X$ is totally bounded if for every $\epsilon > 0$ there exists a finite set $\{x_1, x_2, \dots, x_n\}$ of points in X such that

$$A \subset \bigcup_{i=1}^{n} B_\epsilon(x_i).$$

The proof of the following result is then completely analogous to the proof of the Bolzano-Weierstrass theorem in Theorem 3.57 for $\mathbb{R}$.

Theorem 13.60. A subset $K \subset X$ of a metric space X is sequentially compact if and only if it is is complete and totally bounded.

The definition of the continuity of functions between metric spaces parallels the definitions for real functions.

Definition 13.61. 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.

Example 13.62. 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

$$|f(x_1, x_2) - f_1(c_1, c_2)| < \epsilon.$$

Example 13.63. 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 $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, this condition reads: $|x - c| < \delta$ implies that

$$\sqrt{[f_1(x) - f_1(c)]^2 + [f_2(x) - f_2(c)]^2} < \epsilon.$$

The previous examples generalize in a natural way to define the continuity of an m-component vector-valued function of n variables, $f : \mathbb{R}^n \to \mathbb{R}^m$. The definition looks complicated if it is written out explicitly, but it is much clearer if it is expressed in terms or metrics or norms.

Example 13.64. Define $F : C([0, 1]) \to \mathbb{R}$ by

$$F(f) = f(0),$$

where $C([0, 1])$ is the space of continuous functions $f : [0, 1] \to \mathbb{R}$ equipped with the sup-norm described in Example 13.9, and $\mathbb{R}$ has the absolute value norm. That is, F evaluates a function $f(x)$ at $x = 0$. Thus F is a function acting on functions, and its values are scalars; such a function, which maps functions to scalars, is called a functional. Then F is continuous, since $\|f - g\|_\infty < \epsilon$ implies that $|f(0) - g(0)| < \epsilon$. (That is, we take $\delta = \epsilon$).

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

Theorem 13.65. Let X and Y be metric spaces. A function $f : X \to Y$ is continuous at $c \in X$ if and only if $f(x_n) \to f(c)$ as $n \to \infty$ for every sequence (x_n) in X such that $x_n \to c$ as $n \to \infty$,

We define uniform continuity similarly to the real case.

Definition 13.66. 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 identically to the proofs we gave for functions $f : A \subset \mathbb{R} \to \mathbb{R}$. First, a function on a metric space is continuous if and only if the inverse images of open sets are open.

Theorem 13.67. 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 13.68. Let $f : K \to Y$ be a continuous function from a compact metric space K to a metric space Y. Then $f(K)$ is a compact subspace of Y.

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

Theorem 13.69. If $f : K \to Y$ is a continuous function on a compact set K, then f is uniformly continuous.

13.5. Topological spaces

A collection of subsets of a set X with the properties of the open sets in a metric space given in Theorem 13.36 is called a topology on X, and a set with such a collection of open sets is called a topological space.

Definition 13.70. Let X be a set. A collection $\mathcal{T} \subset \mathcal{P}(X)$ of subsets of X is a topology on X if it satisfies the following conditions.

(1) The empty set $\varnothing$ and the whole set X belong to $\mathcal{T}$.

(2) The union of an arbitrary collection of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

(3) The intersection of a finite collection of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

A set $G \subset X$ is open with respect to $\mathcal{T}$ if $G \in \mathcal{T}$, and a set $F \subset X$ is closed with respect to $\mathcal{T}$ if $F^c \in \mathcal{T}$. A topological space $(X, \mathcal{T})$ is a set X together with a topology $\mathcal{T}$ on X.

We can put different topologies on a set with two or more elements. If the topology on X is clear from the context, then we simply refer to X as a topological space and we don't specify the topology when we refer to open or closed sets.

Every metric space with the open sets in Definition 13.30 is a topological space; the resulting collection of open sets is called the metric topology of the metric space. There are, however, topological spaces whose topology is not derived from any metric on the space.

Example 13.71. Let X be any set. Then $\mathcal{T} = \mathcal{P}(X)$ is a topology on X, called the discrete topology. In this topology, every set is both open and closed. This topology is the metric topology associated with the discrete metric on X in Example 13.3.

Example 13.72. Let X be any set. Then $\mathcal{T} = \{\varnothing, X\}$ is a topology on X, called the trivial topology. The empty set and the whole set are both open and closed, and no other subsets of X are either open or closed.

If X has at least two elements, then this topology is different from the discrete topology in the previous example, and it is not derived from a metric. To see this, suppose that $x, y \in X$ and $x \neq y$. If $d : X \times X \to \mathbb{R}$ is a metric on X, then $d(x, y) = r > 0$ and $B_r(x)$ is a nonempty open set in the metric topology that doesn't contain y, so $B_r(x) \notin \mathcal{T}$.

The previous example illustrates a separation property of metric topologies that need not be satisfied by non-metric topologies.

Definition 13.73. A topological space $(X, \mathcal{T})$ is Hausdorff if for every $x, y \in X$ with $x \neq y$ there exist open sets $U, V \in \mathcal{T}$ such that $x \in U$, $y \in V$ and $U \cap V = \varnothing$.

That is, a topological space is Hausdorff if distinct points have disjoint neighborhoods. In that case, we also say that the topology is Hausdorff. Nearly all topological spaces that arise in analysis are Hausdorff, including, in particular, metric spaces.

Proposition 13.74. Every metric topology is Hausdorff.

Proof. Let (X, d) be a metric space. If $x, y \in X$ and $x \neq y$, then $d(x, y) = r > 0$, and $B_{r/2}(x)$, $B_{r/2}(y)$ are disjoint open neighborhoods of x, y. □

Compact sets are defined topologically as sets with the Heine-Borel property.

Definition 13.75. Let X be a topological space. A set $K \subset X$ is compact if every open cover of K has a finite subcover. That is, if $\{G_i : i \in I\}$ is a collection of open sets such that

$$K \subset \bigcup_{i \in I} G_i,$$

then there is a finite subcollection $\{G_{i_1}, G_{i_2}, \dots, G_{i_n}\}$ such that

$$K \subset \bigcup_{k=1}^{n} G_{i_k}.$$

The Heine-Borel and Bolzano-Weierstrass properties are equivalent in every metric space.

Theorem 13.76. A metric space is compact if and only if it sequentially compact.

We won't prove this result here, but we remark that it does not hold for general topological spaces.

Finally, we give the topological definitions of convergence, continuity, and connectedness which are essentially the same as the corresponding statements for $\mathbb{R}$. We also show that continuous maps preserve compactness and connectedness.

The definition of the convergence of a sequence is identical to the statement in Proposition 5.9 for $\mathbb{R}$.

Definition 13.77. Let X be a topological space. A sequence (x_n) in X converges to $x \in X$ if for every neighborhood U of x there exists $N \in \mathbb{N}$ such that $x_n \in U$ for every $n > N$.

The following definition of continuity in a topological space corresponds to Definition 7.2 for $\mathbb{R}$ (with the relative absolute-value topology on the domain A of f) and Theorem 7.31.

Definition 13.78. Let $f : X \to Y$ be a function between topological spaces X, Y. Then f is continuous at $x \in X$ if for every neighborhood $V \subset Y$ of $f(x)$, there exists a neighborhood $U \subset X$ of x such that $f(U) \subset V$. The function f is continuous on X if $f^{-1}(V)$ is open in X for every open set $V \subset Y$.

These definitions are equivalent to the corresponding "ϵ-δ" definitions in a metric space, but they make sense in a general topological space because they refer only to neighborhoods and open sets. We illustrate them with two simple examples.

Example 13.79. If X is a set with the discrete topology in Example 13.71, then a sequence converges to $x \in X$ if an only if its terms are eventually equal to x, since $\{x\}$ is a neighborhood of x. Every function $f : X \to Y$ is continuous with respect to the discrete topology on X, since every subset of X is open. On the other hand, if Y has the discrete topology, then $f : X \to Y$ is continuous if and only if $f^{-1}(\{y\})$ is open in X for every $y \in Y$.

Example 13.80. Let X be a set with the trivial topology in Example 13.72. Then every sequence converges to every point $x \in X$, since the only neighborhood of x is X itself. As this example illustrates, non-Hausdorff topologies have the unpleasant feature that limits need not be unique, which is one reason why they rarely arise in analysis. If Y has the trivial topology, then every function $X \to Y$ is continuous, since $f^{-1}(\varnothing) = \varnothing$ and $f^{-1}(Y) = X$ are open in X. On the other hand, if X has the trivial topology and Y is Hausdorff, then the only continuous functions $f : X \to Y$ are the constant functions.

Our last definition of a connected topological space corresponds to Definition 5.58 for connected sets of real numbers (with the relative topology).

Definition 13.81. A topological space X is disconnected if there exist nonempty, disjoint open sets U, V such that $X = U \cup V$. A topological space is connected if it is not disconnected.

The following proof that continuous functions map compact sets to compact sets and connected sets is the same as the proofs given in Theorem 7.35 and Theorem 7.32 for sets of real numbers. Note that a continuous function maps compact or connected sets in the opposite direction to open or closed sets, whose inverse image is open or closed.

Theorem 13.82. Suppose that $f : X \to Y$ is a continuous map between topological spaces X and Y. Then $f(X)$ is compact if X is compact, and $f(X)$ is connected if X is connected.

Proof. For the first part, suppose that X is compact. If $\{V_i : i \in I\}$ is an open cover of $f(X)$, then since f is continuous $\{f^{-1}(V_i) : i \in I\}$ is an open cover of X, and since X is compact there is a finite subcover

$$\left\{f^{-1}(V_{i_1}), f^{-1}(V_{i_2}), \dots, f^{-1}(V_{i_n})\right\}.$$

It follows that

$$\{V_{i_1}, V_{i_2}, \dots, V_{i_n}\}$$

is a finite subcover of the original open cover of $f(X)$, which proves that $f(X)$ is compact.

For the second part, suppose that $f(X)$ is disconnected. Then there exist nonempty, disjoint open sets U, V in Y such that $U \cup V \supset f(X)$. Since f is continuous, $f^{-1}(U)$, $f^{-1}(V)$ are open, nonempty, disjoint sets such that

$$X = f^{-1}(U) \cup f^{-1}(V),$$

so X is disconnected. It follows that $f(X)$ is connected if X is connected. □

13.6. * Function spaces

There are many function spaces, and their study is a central topic in analysis. We discuss only one important example here: the space of continuous functions on a compact set equipped with the sup norm. We repeat its definition from Example 13.9.

Definition 13.83. Let $K \subset \mathbb{R}$ be a compact set. The space $C(K)$ consists of the continuous functions $f : K \to \mathbb{R}$. Addition and scalar multiplication of functions is defined pointwise in the usual way: if $f, g \in C(K)$ and $k \in \mathbb{R}$, then

$$(f+g)(x) = f(x) + g(x), \qquad (kf)(x) = k\,(f(x))\,.$$

The sup-norm of a function $f \in C(K)$ is defined by

$$\|f\|_\infty = \sup_{x \in K} |f(x)|.$$

Since a continuous function on a compact set attains its maximum and minimum value, for $f \in C(K)$ we can also write

$$\|f\|_\infty = \max_{x \in K} |f(x)|.$$

Thus, the sup-norm on $C(K)$ is analogous to the ℓ^∞-norm on $\mathbb{R}^n$. In fact, if $K = \{1, 2, \dots, n\}$ is a finite set with the discrete topology, then it is identical to the ℓ^∞-norm.

Our previous results on continuous functions on a compact set can be formulated concisely in terms of this space. The following characterization of uniform convergence in terms of the sup-norm is easily seen to be equivalent to Definition 9.8.

Definition 13.84. A sequence (f_n) of functions $f_n : K \to \mathbb{R}$ converges uniformly on K to a function $f : K \to \mathbb{R}$ if

$$\lim_{n \to \infty} \|f_n - f\|_\infty = 0.$$

Similarly, we can rephrase Definition 9.12 for a uniformly Cauchy sequence in terms of the sup-norm.

Definition 13.85. A sequence (f_n) of functions $f_n : K \to \mathbb{R}$ is uniformly Cauchy on K if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n > N \text{ implies that } \|f_m - f_n\|_\infty < \epsilon.$$

Thus, the uniform convergence of a sequence of functions is defined in exactly the same way as the convergence of a sequence of real numbers with the absolute value $|\cdot|$ replaced by the sup-norm $\|\cdot\|$. Moreover, like $\mathbb{R}$, the space $C(K)$ is complete.

Theorem 13.86. The space $C(K)$ with the sup-norm $\|\cdot\|_\infty$ is a Banach space.

Proof. From Theorem 7.15, the sum of continuous functions and the scalar multiple of a continuous function are continuous, so $C(K)$ is closed under addition and scalar multiplication. The algebraic vector-space properties for $C(K)$ follow immediately from those of $\mathbb{R}$.

From Theorem 7.37, a continuous function on a compact set is bounded, so $\|\cdot\|_\infty$ is well-defined on $C(K)$. The sup-norm is clearly non-negative, and $\|f\|_\infty = 0$ implies that $f(x) = 0$ for every $x \in K$, meaning that $f = 0$ is the zero function.

We also have for all $f, g \in C(K)$ and $k \in \mathbb{R}$ that

$$\begin{aligned}\|kf\|_\infty &= \sup_{x\in K} |kf(x)| = |k| \sup_{x\in K} |f(x)| = |k|\,\|f\|_\infty, \\ \|f+g\|_\infty &= \sup_{x\in K} |f(x)+g(x)| \\ &\le \sup_{x\in K} \{|f(x)| + |g(x)|\} \\ &\le \sup_{x\in K} |f(x)| + \sup_{x\in K} |g(x)| \\ &\le \|f\|_\infty + \|g\|_\infty,\end{aligned}$$

which verifies the properties of a norm.

Finally, Theorem 9.13 implies that a uniformly Cauchy sequence converges uniformly so $C(K)$ is complete. □

For comparison with the sup-norm, we consider a different norm on $C([a,b])$ called the one-norm, which is analogous to the ℓ^1-norm on $\mathbb{R}^n$.

Definition 13.87. If $f : [a,b] \to \mathbb{R}$ is a Riemann integrable function, then the one-norm of f is

$$\|f\|_1 = \int_a^b |f(x)|\,dx.$$

Theorem 13.88. The space $C([a,b])$ of continuous functions $f : [a,b] \to \mathbb{R}$ with the one-norm $\|\cdot\|_1$ is a normed space.

Proof. As shown in Theorem 13.86, $C([a,b])$ is a vector space. Every continuous function is Riemann integrable on a compact interval, so $\|\cdot\|_1 : C([a,b]) \to \mathbb{R}$ is well-defined, and we just have to verify that it satisfies the properties of a norm.

Since $|f| \ge 0$, we have $\|f\|_1 = \int_a^b |f| \ge 0$. Furthermore, since f is continuous, Proposition 11.42 shows that $\|f\|_1 = 0$ implies that $f = 0$, which verifies the positivity. If $k \in \mathbb{R}$, then

$$\|kf\|_1 = \int_a^b |kf| = |k| \int_a^b |f| = |k|\,\|f\|_1,$$

which verifies the homogeneity. Finally, the triangle inequality is satisfied since

$$\|f+g\|_1 = \int_a^b |f+g| \le \int_a^b |f| + |g| = \int_a^b |f| + \int_a^b |g| = \|f\|_1 + \|g\|_1.$$

□

Although $C([a,b])$ equipped with the one-norm $\|\cdot\|_1$ is a normed space, it is not complete, and therefore it is not a Banach space. The following example gives a non-convergent Cauchy sequence in this space.

Example 13.89. Define the continuous functions $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 0 & \text{if } 0 \le x \le 1/2, \\ n(x - 1/2) & \text{if } 1/2 < x < 1/2 + 1/n, \\ 1 & \text{if } 1/2 + 1/n \le x \le 1. \end{cases}$$

If $n > m$, we have

$$\|f_n - f_m\|_1 = \int_{1/2}^{1/2+1/m} |f_n - f_m| \leq \frac{1}{m},$$

since $|f_n - f_n| \leq 1$. Thus, $\|f_n - f_m\|_1 < \epsilon$ for all $m, n > 1/\epsilon$, so (f_n) is a Cauchy sequence with respect to the one-norm.

We claim that if $\|f - f_n\|_1 \to 0$ as $n \to \infty$ where $f \in C([0,1])$, then f would have to be

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x \leq 1/2, \\ 1 & \text{if } 1/2 < x \leq 1, \end{cases}$$

which is discontinuous at $1/2$, so (f_n) does not have a limit in $(C([0,1]), \|\cdot\|_1)$.

To prove the claim, note that if $\|f - f_n\|_1 \to 0$, then $\int_0^{1/2} |f| = 0$ since

$$\int_0^{1/2} |f| = \int_0^{1/2} |f - f_n| \leq \int_0^1 |f - f_n| \to 0,$$

and Proposition 11.42 implies that $f(x) = 0$ for $0 \leq x \leq 1/2$. Similarly, for every $0 < \epsilon < 1/2$, we get that $\int_{1/2+\epsilon}^1 |f - 1| = 0$, so $f(x) = 1$ for $1/2 < x \leq 1$.

The sequence (f_n) is not uniformly Cauchy since $\|f_n - f_m\|_\infty \to 1$ as $n \to \infty$ for every $m \in \mathbb{N}$, so this example does not contradict the completeness of $(C([0,1]), \|\cdot\|_\infty)$.

The ℓ^∞-norm and the ℓ^1-norm on the finite-dimensional space $\mathbb{R}^n$ are equivalent, but the sup-norm and the one-norm on $C([a,b])$ are not. In one direction, we have

$$\int_a^b |f| \leq (b-a) \cdot \sup_{[a,b]} |f|,$$

so $\|f\|_1 \leq (b-a)\|f\|_\infty$, and $\|f - f_n\|_\infty \to 0$ implies that $\|f - f_n\|_1 \to 0$. As the following example shows, the converse is not true. There is no constant M such that $\|f\|_\infty \leq M\|f\|_1$ for all $f \in C([a,b])$, and $\|f - f_n\|_1 \to 0$ does not imply that $\|f - f_n\|_\infty \to 0$.

Example 13.90. For $n \in \mathbb{N}$, define the continuous function $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 - nx & \text{if } 0 \leq x \leq 1/n, \\ 0 & \text{if } 1/n < x \leq 1. \end{cases}$$

Then $\|f_n\|_\infty = 1$ for every $n \in \mathbb{N}$, but

$$\|f_n\|_1 = \int_0^{1/n} (1 - nx)\, dx = \left[x - \frac{1}{2}nx^2\right]_0^{1/n} = \frac{1}{2n},$$

so $\|f_n\|_1 \to 0$ as $n \to \infty$.

Thus, unlike the finite-dimensional vector space $\mathbb{R}^n$, an infinite-dimensional vector space such as $C([a,b])$ has many inequivalent norms and many inequivalent notions of convergence.

The incompleteness of $C([a,b])$ with respect to the one-norm suggests that we use the larger space $R([a,b])$ of Riemann integrable functions on $[a,b]$, which includes some discontinuous functions. A slight complication arises from the fact that if f is Riemann integrable and $\int_a^b |f| = 0$, then it does not follows that $f = 0$, so $\|f\|_1 = 0$ does not imply that $f = 0$. Thus, $\|\cdot\|_1$ is not, strictly speaking, a norm on $R([a,b])$. We can, however, get a normed space of equivalence classes of Riemann integrable functions, by defining $f, g \in R([a,b])$ to be equivalent if $\int_a^b |f - g| = 0$. For instance, the function in Example 11.14 is equivalent to the zero-function.

A much more fundamental defect of the space of (equivalence classes of) Riemann integrable functions with the one-norm is that it is still not complete. To get a space that is complete with respect to the one-norm, we have to use the space $L^1([a,b])$ of (equivalence classes of) Lebesgue integrable functions on $[a,b]$. This is another reason for the superiority of the Lebesgue integral over the Riemann integral: it leads function spaces that are complete with respect to integral norms.

The inclusion of the smaller incomplete space $C([a,b])$ of continuous functions in the larger complete space $L^1([a,b])$ of Lebesgue integrable functions is analogous to the inclusion of the incomplete space $\mathbb{Q}$ of rational numbers in the complete space $\mathbb{R}$ of real numbers.

13.7. * The Minkowski inequality

Inequalities are essential to analysis. Their proofs, however, may require considerable ingenuity, and there are often many different ways to prove the same inequality. In this section, we complete the proof that the ℓ^p-spaces are normed spaces by proving the triangle inequality given in Definition 13.25. This inequality is called the Minkowski inequality, and it's one of the most important inequalities in mathematics.

The simplest case is for the Euclidean norm with $p = 2$. We begin by proving the following fundamental Cauchy-Schwartz inequality.

Theorem 13.91 (Cauchy-Schwartz inequality)**.** If $(x_1, x_2, \dots, x_n)$, $(y_1, y_2, \dots, y_n)$ are points 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 $x = 0$ or $y = 0$, so we assume that 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. $\square$

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

Corollary 13.92 (Minkowski inequality)**.** If $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are points in $\mathbb{R}^n$, then

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

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

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

Taking the square root of this inequality, we obtain the result. $\square$

To prove the Minkowski inequality for general $1 < p < \infty$, we first define the Hölder conjugate p' of p and prove Young's inequality.

Definition 13.93. If $1 < p < \infty$, then the Hölder conjugate $1 < p' < \infty$ of p is the number such that

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

If $p = 1$, then $p' = \infty$; and if $p = \infty$ then $p' = 1$.

The Hölder conjugate of $1 < p < \infty$ is given explicitly by

$$p' = \frac{p}{p-1}.$$

Note that if $1 < p < 2$, then $2 < p' < \infty$; and if $2 < p < \infty$, then $1 < p' < 2$. The number 2 is its own Hölder conjugate. Furthermore, if p' is the Hölder conjugate of p, then p is the Hölder conjugate of p'.

Theorem 13.94 (Young's inequality)**.** Suppose that $1 < p < \infty$ and $1 < p' < \infty$ is its Hölder conjugate. If $a, b \geq 0$ are nonnegative real numbers, then

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'}.$$

Moreover, there is equality if and only if $a^p = b^{p'}$.

Proof. There are several ways to prove this inequality. We give a proof based on calculus.

The result is trivial if $a = 0$ or $b = 0$, so suppose that $a, b > 0$. We write

$$\frac{a^p}{p} + \frac{b^{p'}}{p'} - ab = b^{p'}\left(\frac{1}{p}\frac{a^p}{b^{p'}} + \frac{1}{p'} - \frac{a}{b^{p'-1}}\right).$$

The definition of p' implies that $p'/p = p' - 1$, so that

$$\frac{a^p}{b^{p'}} = \left(\frac{a}{b^{p'/p}}\right)^p = \left(\frac{a}{b^{p'-1}}\right)^p$$

Therefore, we have

$$\frac{a^p}{p} + \frac{b^{p'}}{p'} - ab = b^{p'} f(t), \qquad f(t) = \frac{t^p}{p} + \frac{1}{p'} - t, \qquad t = \frac{a}{b^{p'-1}}.$$

The derivative of f is

$$f'(t) = t^{p-1} - 1.$$

Thus, for $p > 1$, we have $f'(t) < 0$ if $0 < t < 1$, and Theorem 8.36 implies that $f(t)$ is strictly decreasing; moreover, $f'(t) > 0$ if $1 < t < \infty$, so $f(t)$ is strictly increasing. It follows that f has a strict global minimum on $(0, \infty)$ at $t = 1$. Since

$$f(1) = \frac{1}{p} + \frac{1}{p'} - 1 = 0,$$

we conclude that $f(t) \geq 0$ for all $0 < t < \infty$, with equality if and only if $t = 1$. Furthermore, $t = 1$ if and only if $a = b^{p'-1}$ or $a^p = b^{p'}$. It follows that

$$\frac{a^p}{p} + \frac{b^{p'}}{p'} - ab \geq 0$$

for all $a, b \geq 0$, with equality if and only $a^p = b^{p'}$, which proves the result. □

For $p = 2$, Young's inequality reduces to the more easily proved inequality in Proposition 2.8.

Before continuing, we give a scaling argument which explains the appearance of the Hölder conjugate in Young's inequality. Suppose we look for an inequality of the form

$$ab \leq Ma^p + Na^q \qquad \text{for all } a, b \geq 0$$

for some exponents p, q and some constants M, N. Any inequality that holds for all positive real numbers must remain true under rescalings. Rescaling $a \mapsto \lambda a$, $b \mapsto \mu b$ in the inequality (where $\lambda, \mu > 0$) and dividing by $\lambda\mu$, we find that it becomes

$$ab \leq \frac{\lambda^{p-1}}{\mu} Ma^p + \frac{\mu^{q-1}}{\lambda} Nb^q.$$

We take $\mu = \lambda^{p-1}$ to make the first scaling factor equal to one, and then the inequality becomes

$$ab \leq Ma^p + \lambda^r Nb^q, \qquad r = (p-1)(q-1) - 1.$$

If the exponent r of λ is non-zero, then we can violate the inequality by taking λ sufficiently small (if $r > 0$) or sufficiently large (if $r < 0$), since it is clearly impossible to bound ab by a^p for all $b \in \mathbb{R}$. Thus, the inequality can only hold if $r = 0$, which implies that $q = p'$.

This argument does not, of course, prove the inequality, but it shows that the only possible exponents for which an inequality of this form can hold must satisfy $q = p'$. Theorem 13.94 proves that such an inequality does in fact hold in that case provided $1 < p < \infty$.

Next, we use Young's inequality to deduce Hólder's inequality, which is a generalization of the Cauchy-Schwartz inequality for $p \neq 2$.

Theorem 13.95 (Hölder's inequality)**.** Suppose that $1 < p < \infty$ and $1 < p' < \infty$ is its Hölder conjugate. If $(x_1, x_2, \dots, x_n)$ and $(y_1, y_2, \dots, y_n)$ are points in $\mathbb{R}^n$, then

$$\left|\sum_{i=1}^n x_i y_i\right| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^{p'}\right)^{1/p'}.$$

Proof. We assume without loss of generality that x_i, y_i are nonnegative and $x, y \neq 0$. Let $\alpha, \beta > 0$. Then applying Young's inequality in Theorem 13.94 with $a = \alpha x_i$, $b = \beta y_i$ and summing over i, we get

$$\alpha\beta \sum_{i=1}^n x_i y_i \leq \frac{\alpha^p}{p} \sum_{i=1}^n x_i^p + \frac{\beta^{p'}}{p'} \sum_{i=1}^n y_i^{p'}.$$

Then, choosing

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

to "balance" the terms on the right-hand side, dividing by $\alpha\beta$, and using the fact that $1/p + 1/p' = 1$, we get Hölder's inequality. □

Minkowski's inequality follows from Hölder's inequality.

Theorem 13.96 (Minkowski's inequality)**.** Suppose that $1 < p < \infty$ and $1 < p' < \infty$ is its Hölder conjugate. If $(x_1, x_2, \dots, x_n)$ and $(y_1, y_2, \dots, y_n)$ are points in $\mathbb{R}^n$, then

$$\left(\sum_{i=1}^n |x_i + y_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^n |y_i|^p\right)^{1/p}.$$

Proof. We assume without loss of generality that x_i, y_i are nonnegative and $x, y \neq 0$. We split the sum on the left-hand side as follows:

$$\begin{aligned}\sum_{i=1}^n |x_i + y_i|^p &= \sum_{i=1}^n |x_i + y_i|\ |x_i + y_i|^{p-1} \\ &\leq \sum_{i=1}^n |x_i|\ |x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i|\ |x_i + y_i|^{p-1}.\end{aligned}$$

By Hölder's inequality, we have

$$\sum_{i=1}^n |x_i|\ |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i + y_i|^{(p-1)p'}\right)^{1/p'},$$

and using the fact that $p' = p/(p-1)$, we get

$$\sum_{i=1}^{n} |x_i|\ |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-1/p}.$$

Similarly,

$$\sum_{i=1}^{n} |y_i|\ |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-1/p}.$$

Combining these inequalities, we obtain

$$\sum_{i=1}^{n} |x_i + y_i|^p \leq \left[\left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p}\right] \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-1/p}.$$

Fianlly, dividing this inequality by $(\sum |x_i + y_i|^p)^{1-1/p}$, we get the result. $\square$