

2

Metric Spaces

As mentioned in Chapter 0, our objectives in this book include thoroughly understanding and generalizing the ideas of convergence and continuity that we encounter in calculus. Undergirding both of these notions is the concept of distance: convergence means that a sequence of objects gets closer and closer to some limiting object; continuity means that inputs of a function that are close to one another have corresponding outputs that are also close to one another. In this chapter we develop the basic mathematical theory of sets for which the “distance” between two elements is defined. Such sets are called *metric spaces*.

2.1 The definition of a metric space

When we measure the distance between two points x and y , what we are really doing is taking x and y as inputs to a “distance function” whose corresponding output is a real number — namely, the distance between x and y . This is the idea underlying the following definition.

Definition 2.1.1 (Metric Space). Let X be any nonempty set. A function $d : X \times X \rightarrow \mathbb{R}$ is called a *metric* (on X) if it satisfies the following properties:

- Positivity: $d(x, y) \geq 0$ for all $(x, y) \in X \times X$, and $d(x, y) = 0$ if and only if $x = y$;
- Symmetry: $d(x, y) = d(y, x)$ for all $(x, y) \in X \times X$;
- Triangle Inequality: Given any three elements $x, y, z \in X$, we have

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

A set X for which a metric $d : X \times X \rightarrow \mathbb{R}$ is defined is called a *metric space*.

Elements of a metric space are usually called *points*.

We think of $d(x, y)$ as the “distance between x and y .” With this interpretation, the first property in Definition 2.1.1 says that there is a positive distance between any

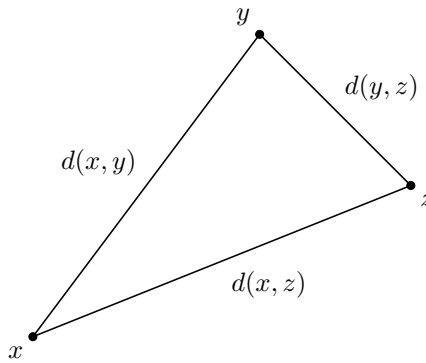


Figure 2.1. The idea of the triangle inequality.

two distinct points (that makes sense, right?). The second property says that the distance between x and y doesn't depend on which point is listed first (that also makes sense!). The third property says that the distance between x and z is never greater than the distance between x and y plus the distance between y and z . To see why this is a reasonable property for any function measuring distance to have, imagine marking three points x , y , and z on a piece of paper and measuring the distances among them with a ruler; see Figure 2.1. Since a straight line is the shortest path between any two points, the distance shown on the ruler between x and z will be less than or equal to the sum of the distances from x to y and from y to z (that is, any two sides of a triangle sum to be greater than or equal to the third side — hence the name “triangle inequality”). This familiar “ruler distance” in the x, y -plane is given by the so-called *Euclidean metric*, whose formula is probably familiar to you and which we define in Section 2.2.

As we shall see, there are many different metrics that one can define on a nonempty set. Therefore, strictly speaking, a metric space is not merely a set but a *set together with a particular metric* — in other words, a metric space is really a *pair* (X, d) , where X is a set and d is a metric on X . Accordingly, we sometimes write (X, d) to mean the set X equipped with the particular metric d .

Example 2.1.2 (The Standard Metric on \mathbb{R}). The real line \mathbb{R} becomes a metric space if we choose the metric $d(x, y) = |x - y|$. This is the *standard metric* on \mathbb{R} . It is easy to see that the metric properties are satisfied: for suppose that x , y , and z are real numbers.

- Is it true that $|x - y| \geq 0$, and that $|x - y| = 0$ if and only if $x = y$? Yes.
- Is it true that $|x - y| = |y - x|$? Yes.
- Is it true that $|x - z| \leq |x - y| + |y - z|$? Yes. We proved this in Proposition 1.6.1!

There are lots of other metrics that can be defined on \mathbb{R} , though most people don't use them very often. (In Exercise 5.28 we shall will learn an infinite source of examples of such metrics.)

Analytical Advice 2.1.3. Whenever you encounter a definition or theorem in mathematics that feels abstract or general, you should, if at all possible, go over

what that definition or theorem says in some familiar contexts. Very often, abstract definitions or theorems are formulated not to capture some strange idea that you've never thought about before, but rather to *extend* familiar ideas to broader settings.

Here is one particularly simple metric that can be put on *any* nonempty set.

Example 2.1.4 (The discrete metric). Let X be any nonempty set. The *discrete metric* is the map $\sigma : X \times X \rightarrow \mathbb{R}$ given by

$$\sigma(x, y) = \begin{cases} 0, & \text{if } x = y; \\ 1, & \text{otherwise.} \end{cases}$$

It is good practice to show that σ is in fact a metric; see Exercise 2.2.

The triangle inequality allows us to deduce another fact with a familiar geometric interpretation: the difference in the lengths of any two legs of a triangle is no more than the length of the third side. This fact, suitably generalized, is called the *second triangle inequality*.

Proposition 2.1.5 (The second triangle inequality). *Let (X, d) be a metric space. For any three points x, y , and z in X , we have*

$$d(x, z) \geq |d(x, y) - d(y, z)|.$$

Proof. By the triangle inequality we have that

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

Subtracting $d(x, y)$ from both sides yields

$$d(x, z) \geq d(y, z) - d(x, y).$$

By the triangle inequality we also have that

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

Subtracting $d(y, z)$ from both sides yields

$$d(x, z) \geq d(x, y) - d(y, z).$$

Since $d(x, z)$ is greater than or equal to both $d(y, z) - d(x, y)$ and its negative $d(x, y) - d(y, z)$, we see that $d(x, z)$ is greater than or equal to $|d(x, y) - d(y, z)|$, as desired. \square

Example 2.1.6. Let $S = \{1, 2, 3\}$ and let $\mathcal{M}(S)$ be the set of functions from S to itself. We define a metric d on $\mathcal{M}(S)$ as follows: if $f : S \rightarrow S$ and $g : S \rightarrow S$ are two elements of $\mathcal{M}(S)$, then

$$d(f, g) = \frac{|f(1) - g(1)| + |f(2) - g(2)| + |f(3) - g(3)|}{3}.$$

(Said otherwise, $d(f, g)$ is the average of $|f(x) - g(x)|$, for $x \in S$.)

We see that $d(f, g)$ is an average of three nonnegative numbers, and is only zero if

$$f(1) = g(1) \text{ and } f(2) = g(2) \text{ and } f(3) = g(3)$$

— that is, if f and g are the same function. Thus positivity holds. We can also see that, in the formula for $d(f, g)$, it does not matter which of f and g is listed first. Thus symmetry holds.

It remains to prove the triangle inequality. Suppose that $f : S \rightarrow S$, $g : S \rightarrow S$, and $h : S \rightarrow S$ are three members of $\mathcal{M}(S)$. Then by the regular triangle inequality in \mathbb{R} we have that

$$\begin{aligned} |f(1) - h(1)| &\leq |f(1) - g(1)| + |g(1) - h(1)|, \\ |f(2) - h(2)| &\leq |f(2) - g(2)| + |g(2) - h(2)|, \text{ and} \\ |f(3) - h(3)| &\leq |f(3) - g(3)| + |g(3) - h(3)|. \end{aligned}$$

Therefore we have

$$\begin{aligned} d(f, h) &= \frac{|f(1) - h(1)| + |f(2) - h(2)| + |f(3) - h(3)|}{3} \\ &\leq \frac{|f(1) - g(1)| + |g(1) - h(1)| + |f(2) - g(2)| + |g(2) - h(2)| + |f(3) - g(3)| + |g(3) - h(3)|}{3} \\ &= \frac{|f(1) - g(1)| + |f(2) - g(2)| + |f(3) - g(3)|}{3} + \frac{|g(1) - h(1)| + |g(2) - h(2)| + |g(3) - h(3)|}{3} \\ &= d(f, g) + d(g, h), \end{aligned}$$

as desired.

We shall see another example of a set of functions being a metric space (with a different metric) in Exercise 2.3, and we shall see a much more substantial example in Section 4.6.

Reading Questions.

Reading Question 2.1.1. What is the distance between 2 and -7 if \mathbb{R} is equipped with the standard metric? What is the distance between 2 and -7 if \mathbb{R} is equipped with the discrete metric?

Reading Question 2.1.2. Combine the first and second triangle inequalities to show that, given any three points x, y, z in the metric space X ,

$$d(x, y) - d(y, z) \leq d(x, z) \leq d(x, y) + d(y, z).$$

Reading Question 2.1.3. For each of the following functions $\eta : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, show that η is not a metric by giving an explicit example of some part of Definition 2.1.1 being violated.

- (a) $\eta(x, y) = |x^2 - y^2|$.
- (b) $\eta(x, y) = x - y$.
- (c) $\eta(x, y) = |x - y|^2$.

Reading Question 2.1.4. This question is a follow-up to Example 2.1.6.

- (a) How many elements does the set $\mathcal{M}(S)$ have?
- (b) If $f(1) = 3$, $f(2) = 1$, and $f(3) = 2$, and $g = \text{id}_S$ (the identity function on S), what is $d(f, g)$?

2.2 Important metrics in \mathbb{R}^n

Given a set, there might be more than one sensible way to reckon the distance between two points in that set. Consider, for example, two friends with houses a few blocks apart in some city. These two friends might reasonably think of the distance between their two houses as the length of a straight line stretched between the houses; but they might also reasonably think of the distance between their houses as the distance they must actually travel (along roads or sidewalks) to get from one house to another. These two ways of measuring distance correspond to two different metrics.

Given any $n \in \mathbb{N}$, recall that \mathbb{R}^n is the set of n -dimensional points $x = (x_1, \dots, x_n)$ with real coordinates. There are many different metrics (that is, many different ways of measuring distance) that we can use in \mathbb{R}^n ; in this section we describe the most important.

Definition 2.2.1 (The Euclidean metric on \mathbb{R}^n). The *Euclidean metric* on \mathbb{R}^n is defined by the formula

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

In 2- and 3-dimensional space, the Euclidean metric yields the distance we are most familiar with — the distance we would measure with a ruler. The Euclidean distance $d(x, z)$ is the quantity we called $\|x - z\|_2$ in Section 1.6. Accordingly, we have already done most of the work to verify that the Euclidean metric is in fact a metric.

Proposition 2.2.2. *The Euclidean metric on \mathbb{R}^n is in fact a metric.*

Proof. Write d for the Euclidean metric, and suppose that

$$x = (x_1, \dots, x_n), \quad y = (y_1, \dots, y_n), \quad z = (z_1, \dots, z_n)$$

are any three points in \mathbb{R}^n .

Positivity: since

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

is the square root of a sum of nonnegative numbers, $d(x, y)$ is always nonnegative. $d(x, y)$ is zero if and only if $x_i - y_i = 0$ for all $i \in \{1, \dots, n\}$, which holds if and only if $x = y$.

Symmetry: since $(x_i - y_i)^2 = (y_i - x_i)^2$, we can see that our formula for $d(x, y)$ does not depend on the order in which x and y are written; thus $d(x, y) = d(y, x)$.

Triangle inequality: in Corollary 1.6.6 we showed that

$$\|x - z\|_2 \leq \|x - y\|_2 + \|y - z\|_2.$$

By the definition of d , the above inequality just says that

$$d(x, z) \leq d(x, y) + d(y, z),$$

as desired. □

Definition 2.2.3 (The sup metric on \mathbb{R}^n). The *sup metric* on \mathbb{R}^n is defined by the formula

$$d(x, y) = \max\{|x_1 - y_1|, \dots, |x_n - y_n|\}.$$

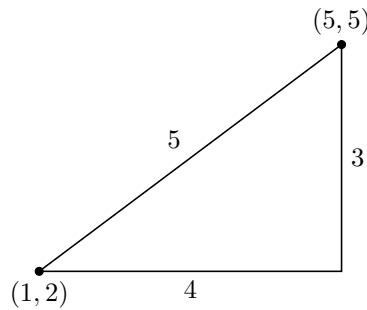


Figure 2.2. Two points $(1, 2)$ and $(5, 5)$ in \mathbb{R}^2 . The distance between the points is 4 in the sup metric, 5 in the Euclidean metric, and 7 in the ℓ_1 metric.

Definition 2.2.4 (The ℓ_1 metric on \mathbb{R}^n). The ℓ_1 metric on \mathbb{R}^n is defined by the formula

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

Of the Euclidean, sup, and ℓ_1 metrics, the Euclidean metric is probably the one that is most familiar to you; the other two, however, are also frequently used. The verifications that the sup and ℓ_1 metrics are actually metrics are left to Exercise 2.6 and Exercise 2.7, respectively.

Example 2.2.5. Consider the two points $(1, 2) \in \mathbb{R}^2$ and $(5, 5) \in \mathbb{R}^2$. The distance between these two points is

- $\sqrt{(5-1)^2 + (5-2)^2} = \sqrt{16+9} = 5$ in the Euclidean metric;
- $|5-1| + |5-2| = 4 + 3 = 7$ in the ℓ_1 metric; and
- $\max\{|5-1|, |5-2|\} = \max\{4, 3\} = 4$ in the sup metric;

see Figure 2.2. Recall our discussion at the beginning of the section. If $(1, 2)$ and $(5, 5)$ give the locations in the plane of two friends' houses, even though 5 is the length of a straight line (i.e. the Euclidean distance) between the houses, depending on the local geography the number 7 (i.e. the ℓ_1 distance) might better describe the actual distance one would have to walk from one house to the other.

Example 2.2.6 (The standard metric on \mathbb{C}). The set \mathbb{C} of complex numbers is the set of all numbers of the form $z = a + bi$, where $a, b \in \mathbb{R}$ and $i^2 = -1$. There is a bijection between \mathbb{C} and \mathbb{R}^2 whereby the number $z = a + bi$ corresponds to the point $(a, b) \in \mathbb{R}^2$; indeed, we usually visualize the complex number $z = a + bi$ by visualizing the point (a, b) in the plane. Any real number is also a complex number: for example, $17 = 17 + 0i$.

We define the algebraic operations on \mathbb{C} as follows:

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

and

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

(this latter formula comes from simply multiplying out the two terms and remembering that $i^2 = -1$). Note in particular that, if $(a + bi)$ is complex and s is real, we have

$$s(a + bi) = sa + sbi.$$

In algebraic terminology, these operations make \mathbb{C} into a *field* with additive identity 0 (which is the same as $0+0i$, and corresponds to the point $(0, 0)$ in \mathbb{R}^2) and multiplicative identity 1 (which is the same as $1 + 0i$, and corresponds to the point $(1, 0)$ in \mathbb{R}^2). The multiplicative inverse (i.e. the reciprocal) of the nonzero complex number $(a + bi)$ is

$$\frac{1}{a + bi} = \frac{a - bi}{a^2 + b^2} = \frac{1}{a^2 + b^2}(a - bi);$$

you can (and should) verify this by checking that

$$(a + bi) \text{ and } \frac{1}{a^2 + b^2}(a - bi)$$

multiply together to yield the number 1.

The standard metric on \mathbb{C} is given by the Euclidean distance between the corresponding points in \mathbb{R}^2 : that is, if $z = a + bi$ and $w = c + di$ are complex numbers, then

$$d(z, w) = \sqrt{(a - c)^2 + (b - d)^2}.$$

When $z \in \mathbb{C}$, by $|z|$ we mean the distance from z to the complex number 0: so if $z = a + bi$, then

$$|z| = \sqrt{a^2 + b^2}.$$

Similarly, the distance $d(z, w)$ between two complex numbers z and w is typically denoted $|z - w|$. Therefore, in this context, the triangle inequality reads as follows, where z, w, ξ are any three complex numbers:

$$|z - w| \leq |z - \xi| + |\xi - w|.$$

We can obtain as a special case the following version of the triangle inequality, which is an analog of what we called the “0th Euclidean triangle inequality” in Section 1.6:

$$|z + w| = |z - (-w)| \leq |z - 0| + |0 - (-w)| = |z| + |w|.$$

It is also true, as for real absolute values, that $|zw| = |z||w|$ and, if $w \neq 0$, that $|z/w| = |z|/|w|$.

Remark 2.2.7. The complex numbers have no standard order; thus, as a rule, saying that a number is greater than another presupposes that the numbers are real. In particular, when you are reading about or discussing a number z that is perhaps complex but someone says (for example) that $z \geq 0$, you should always understand this to mean that “ z is real and is greater than or equal to 0.”

We close this section with another infinite source of examples. If X is any metric space, then any nonempty subset $A \subseteq X$ is a metric space in its own right, with the same metric. So, for example, we will often want to think of the subsets $(0, 1) \subseteq \mathbb{R}$ or $\mathbb{Q} \subseteq \mathbb{R}$ as metric spaces in their own right — acting, essentially, as if the other points in \mathbb{R} did not exist.

Reading Questions.

Reading Question 2.2.1. Draw the set of all points that are distance 1 unit or less from the origin in \mathbb{R}^2 in the

- Euclidean metric;
- sup metric;
- ℓ_1 metric;
- discrete metric.

Reading Question 2.2.2. Compute the distance between the points $(2, -1, 3) \in \mathbb{R}^3$ and $(1, 1, -1) \in \mathbb{R}^3$ with the Euclidean metric, the sup metric, the ℓ_1 metric, and the discrete metric.

Reading Question 2.2.3. Suppose that x and y are two points in \mathbb{R}^n . If we measure the distance between x and y with the Euclidean, sup, and ℓ_1 metrics, are any of these distances guaranteed to be the largest? Are any guaranteed to be the smallest? Work with some examples until you are convinced of the answer; no proofs are required right now (but see Exercise 2.8 to follow up).

2.3 Open balls and open sets in metric spaces

As we have already suggested, our understanding of convergence and continuity will hinge on the notion of points being “close together.” We will find ourselves thinking more and more about sets that are described in terms of distances: sets of points whose members are all a certain distance apart, sets of points that can be approximated by certain other sets of points, and so on. We begin this process now.

Definition 2.3.1 (Open ball). Suppose that (X, d) is a metric space. Given any $r > 0$ and any point $p \in X$, the *open ball (in X with respect to d) of radius r about p* is the set $B_r(p)$ of points in X that are distance less than r from p :

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

When we wish to emphasize that we are using the metric d to define $B_r(p)$, we sometimes write $B_r(p)$ as $B_r^d(p)$.

Note that, no matter what the value of $r > 0$ is, $B_r(p)$ always contains the point p itself, since $d(p, p) = 0$.

Example 2.3.2. In \mathbb{R}^2 equipped with the Euclidean metric, the open ball $B_2((1, 1))$ is just the disk of radius 2 centered at the point $(1, 1)$. (The circle that forms the edge of the disk is not included in $B_2((1, 1))$, since points on that circle are distance *exactly* 2 from $(1, 1)$.) In \mathbb{R}^3 (again with the Euclidean metric), $B_4((0, 0, 0))$ is a solid sphere of radius 4 centered at the origin that does not include its outer “shell.” In \mathbb{R} , $B_3(1)$ is the open interval $(-2, 4)$. In general, open balls in \mathbb{R} are open intervals (see the Reading Questions).

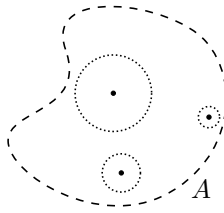


Figure 2.3. An open set.

Example 2.3.3. If we take as our metric space the set $X = [0, 1]$ in \mathbb{R} (equipped with the usual absolute value metric), then $B_{1/2}(3/4)$ is the set $(1/4, 1]$ — since this is the set of all points *in the metric space* X that are distance *less than* $1/2$ from the point $3/4$.

Remark 2.3.4. Some authors refer to open balls as *open neighborhoods* or simply *neighborhoods*. This terminology is not entirely standard: for some authors, a “neighborhood” is merely a set that *contains* an open ball.

Definition 2.3.5 (Open set). Let (X, d) be a metric space. A subset $A \subseteq X$ of X is *open* if it is empty or if, for every $x \in A$, there is some $\varepsilon > 0$ such that $B_\varepsilon(x)$ is contained entirely in A .

Note carefully that, in Definition 2.3.5, “ ε depends on x ” — that is, different points $x \in A$ may require different values of ε to have $B_\varepsilon(x)$ fit inside A .

Note that an entire metric space X itself is always open: given any $p \in X$ and any $\varepsilon > 0$, $B_\varepsilon(p)$ is by definition contained in X .

Figure 2.3 conveys the idea of an open set. The region A pictured (*not* including its border) is open, since around any point in the region we can fit an open ball of positive radius that is still in the set; in the figure, three such points and corresponding open balls are pictured. Roughly speaking, an open set has the feature that any point of the set can be “perturbed” and remain in the set. As we shall see, open sets play a crucial role in a careful understanding of continuity, since a function f is continuous if “perturbing” an input point p changes the corresponding output point very little (see Chapter 4).

Analytical Advice 2.3.6. You should develop the habit of drawing pictures like Figure 2.3 whenever possible. More generally, you should work hard to discover what best helps you keep an accurate but accessible grip on important ideas: cartoons, poems, songs, flashcards, anything.

The cognitive load involved in reading, writing, or thinking about mathematics can be enormous, and can quickly become insurmountable if you’re trying to think about too many unfamiliar ideas at once. Take the time to play with and digest ideas as you learn them, so you can think about them with some facility later. This way of working through mathematics is slow; but slow progress is always faster than no progress at all.

Example 2.3.7. Let the plane \mathbb{R}^2 be equipped with the Euclidean metric. Consider the “strict” first quadrant

$$Q = \{ (x, y) : x > 0 \text{ and } y > 0 \}.$$

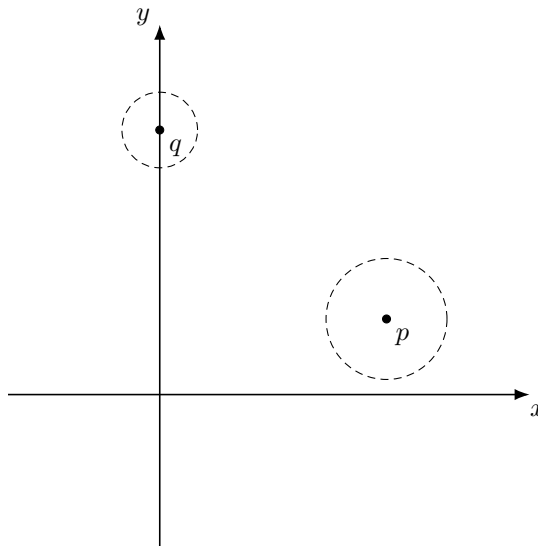


Figure 2.4. The “strict” first quadrant is open, the “semi-strict” first quadrant is not.

This is an open set. For choose any $p = (p_1, p_2) \in Q$. Then we can put an open ball around p that is contained in Q . This is plausibly illustrated in Figure 2.4; to be more rigorous, we reason as follows. Choose $0 < r < \min\{p_1, p_2\}$, and suppose that $(x, y) \in B_r(p)$. This means that

$$d((x, y), (p_1, p_2)) = \sqrt{(x - p_1)^2 + (y - p_2)^2} < r < \min\{p_1, p_2\},$$

which yields in particular that

$$|x - p_1| = \sqrt{(x - p_1)^2} < r < p_1 \quad \text{and} \quad |y - p_2| = \sqrt{(y - p_2)^2} < r < p_2.$$

Since $|x - p_1| < p_1$ and $p_1 > 0$, we have $x > 0$ also; similarly, since $|y - p_2| < p_2$ and $p_2 > 0$, we have $y > 0$ also. Thus $(x, y) \in Q$. Since (x, y) is an arbitrary member of $B_r(p)$, we conclude that $B_r(p) \subseteq Q$.

On the other hand, the “semi-strict” first quadrant

$$\bar{Q} = \{(x, y) : x \geq 0 \text{ and } y \geq 0\}$$

is *not* open: for if we choose $q = (q_1, q_2) \in \bar{Q}$ with (say) $q_1 = 0$, any open ball of positive radius about q contains points that are not in \bar{Q} ; see Figure 2.4 again.

In Definition 2.3.5, $B_\varepsilon(x)$ is an open ball *in the metric space* X . In general, whether a set is open or not depends on the context provided by the underlying metric space. For example, the interval $[0, 1]$, viewed as a subset of \mathbb{R} , is not open since there is no open interval in \mathbb{R} centered at 1 that is contained in $[0, 1]$; but $[0, 1]$ *is* open if we consider ourselves as working in the metric space $[0, 1]$ in its own right, since an entire metric space is always open. Sometimes we say that a set A is “open in X ,” instead of saying merely that it is “open,” to emphasize the underlying metric space we have in mind. That said, it turns out that this issue is usually not that troublesome, since we rarely change our underlying metric in the middle of a proof or train of thought.

So far in this section we have used the word “open” to describe two types of objects: open *balls* and open *sets*. It is by no means true that every open set is an open ball; but every open ball *is* an open set.

Theorem 2.3.8 (Open balls are open). *Open balls are open — that is, any open ball in the metric space X is actually an open set in X .*

Before launching into the proof, we make a few preparatory remarks.

Theorem 2.3.8 is a “for all”-type statement: for all open balls in X , the open ball is actually an open set. Accordingly (recall Analytical Advice 1.2.13), we begin the proof by choosing and naming some open ball in X ; we accomplish this by choosing and naming the ball’s center and radius. Once we have this open ball to work with, we need to prove *another* “for all”-type statement to show that this ball is open. Accordingly, once we have chosen our open ball, we will choose and name some point inside it. It is only then, in some sense, that the real work of the proof will begin.

You will find it helpful to consult Figure 2.5 while reading the proof. Again, you should cultivate the habit of drawing pictures, diagrams and cartoons for yourself whenever possible. A drawing like Figure 2.5 is very helpful both for keeping track of notation and for building intuition.

Proof of Theorem 2.3.8. Let (X, d) be a metric space. Choose any point $p \in X$ and any $r > 0$, and consider the open ball $B_r(p)$. Let $x \in B_r(p)$ be given; we need to show that there is some open ball about x that is entirely contained in $B_r(p)$.

Note that $d(x, p) < r$ by the definition of $B_r(p)$; thus $r - d(x, p) > 0$. Choose ε satisfying $0 < \varepsilon < r - d(x, p)$.

We claim that $B_\varepsilon(x) \subseteq B_r(p)$: for if we choose any $y \in B_\varepsilon(x)$, the triangle inequality yields that

$$d(y, p) \leq d(y, x) + d(x, p) < \varepsilon + d(x, p) < r - d(x, p) + d(x, p) = r.$$

Concentrating on the first and last terms in this string of inequalities, we see that we have shown that $d(y, p) < r$, which means that $y \in B_r(p)$. Since $y \in B_\varepsilon(x)$ was arbitrary, we have shown that $B_\varepsilon(x) \subseteq B_r(p)$ (see Figure 2.5). This completes the proof. \square

Theorem 2.3.9. *Any union of open sets is open. The intersection of any finite collection of open sets is open.*

Proof. Consider a collection $\{U_\alpha : \alpha \in J\}$ of open sets. (Here J is some index set — perhaps finite, perhaps infinite, perhaps even uncountably infinite.) Write $U = \bigcup_{\alpha \in J} U_\alpha$. Then given any $x \in U$, $x \in U_\alpha$ for some α ; since U_α is open, there is some $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U_\alpha \subseteq U$. Thus U is open.

Now let U_1, \dots, U_K be some finite collection of open sets, and write $U = \bigcap_{k=1}^K U_k$. If U is empty, we’re done (the empty set is open). Otherwise, given $x \in U$, for each $1 \leq k \leq K$ there is some $\varepsilon_k > 0$ such that $B_{\varepsilon_k}(x) \subseteq U_k$. Choose $0 < \varepsilon < \min\{\varepsilon_1, \dots, \varepsilon_K\}$; then $B_\varepsilon(x)$ is contained in each of the U_k and hence contained in U . Thus U is open. \square

Theorem 2.3.10. *A set is open if and only if it is a union of open balls.*

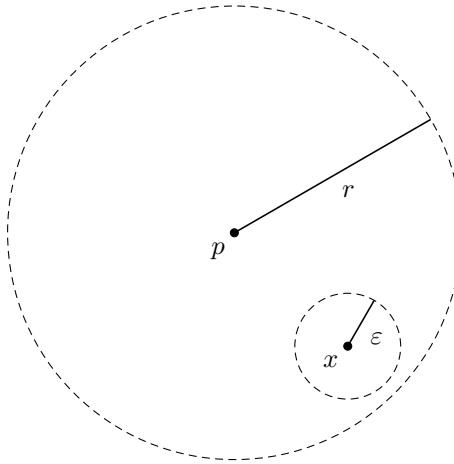


Figure 2.5. The proof that open balls are open.

Proof. Certainly a union of open balls is open by Theorem 2.3.9 (since open balls are themselves open sets by Theorem 2.3.8). Conversely, suppose that U is an open set. Then, given any $x \in U$, there is some $\varepsilon(x) > 0$ such that $B_{\varepsilon(x)}(x) \subseteq U$. Then the set $\bigcup_{x \in U} B_{\varepsilon(x)}(x)$ is contained in U since every $B_{\varepsilon(x)}(x)$ is contained in U . On the other hand, U is contained in $\bigcup_{x \in U} B_{\varepsilon(x)}(x)$ since the latter set contains every element x in U . Thus

$$U = \bigcup_{x \in U} B_{\varepsilon(x)}(x),$$

and we have expressed U as a union of open balls. \square

Reading Questions.

Reading Question 2.3.1. In this section we said that “in general, open balls in \mathbb{R} are open intervals.” Suppose that $p \in \mathbb{R}$ and that $r > 0$. Express the open ball $B_r(p)$ in \mathbb{R} as an open interval.

Reading Question 2.3.2. Redraw Figure 2.5 and put the point y where it belongs, where y is as in the proof of Theorem 2.3.8.

Reading Question 2.3.3. Let $X = [0, 1] \subseteq \mathbb{R}$, equipped with the standard (absolute value) metric inherited from \mathbb{R} . Draw or describe the open ball in X about $1/4$ of radius $1/2$.

Reading Question 2.3.4. True, false, or depends on the metric space (and explain): one-point sets are open.

Reading Question 2.3.5. Show that the intersection of *infinitely many* open sets need not be open.

Reading Question 2.3.6. Where \mathbb{R} has the standard metric,

- Show that $(0, \infty)$ (that is, the set of all positive real numbers) is open;
- Find an open subset of \mathbb{R} that is not an interval.

2.4 Closed sets and limit points

In this section we define the so-called *closed* sets and learn some of their properties. We saw in Section 2.3 that open intervals are open sets in \mathbb{R} ; we shall see in Example 2.4.6 that closed intervals are closed sets in \mathbb{R} . The distinction between closed intervals and open intervals is that closed intervals contain their “edges” while open intervals do not; this intuitive picture captures the distinction between closed and open sets more generally.

Recall that, if C and D are any two sets, $C \setminus D$ denotes the set consisting of members of C that are not in D :

$$C \setminus D = \{x \in C : x \notin D\}.$$

If $C \subseteq D$, then $C \setminus D = \emptyset$; at the other extreme, if $C \cap D = \emptyset$, then $C \setminus D = C$.

Definition 2.4.1 (Complement). If X is any set and $A \subseteq X$, the *complement of A in X* is the set

$$X \setminus A = \{x \in X : x \notin A\}.$$

When the set X is clear from context, we usually refer to the complement of A in X simply as “the complement of A ” and denote it by A^c . In this situation, we often describe A^c verbally as “the set of points that are not in A ,” leaving it implicit that all points under consideration belong to the underlying set X .

Example 2.4.2. The complement of $[0, 1]$ in \mathbb{R} is $(-\infty, 0) \cup (1, \infty)$. The complement of \mathbb{Z} in \mathbb{R} is $\cup_{k \in \mathbb{Z}} (k, k + 1)$. If X is any metric space, the complement of X in X is \emptyset .

If A is a subset of some underlying set X , then $(A^c)^c = A$.

If A and B are subsets of some underlying set X , then $(A \cup B)^c$ is the set of points that are neither in A nor in B — otherwise put, the set of points that are both not in A and also not in B . Thus we have

$$(A \cup B)^c = A^c \cap B^c.$$

Similarly, $(A \cap B)^c$ is the set of points that are not in both A and B — otherwise put, the set of points that are either not in A or not in B . Thus we have

$$(A \cap B)^c = A^c \cup B^c.$$

These formulas extend to more general collections of sets. If J is any “index set” (whether finite or infinite) and A_α is a set for each $\alpha \in J$, then

$$\left(\bigcup_{\alpha \in J} A_\alpha \right)^c = \bigcap_{\alpha \in J} (A_\alpha^c) \quad \text{and} \quad \left(\bigcap_{\alpha \in J} A_\alpha \right)^c = \bigcup_{\alpha \in J} (A_\alpha^c).$$

[These expressions both denote the set of elements that are in *none* of the sets A_α .]

[These expressions both denote the set of elements that are in *at least one* of the sets A_α^c .]

These equalities are known as *DeMorgan’s laws*.

Remark 2.4.3. In many areas of mathematics, the complement of a set A is traditionally denoted \bar{A} . In analysis, however, this notation is reserved for the so-called closure of A (which we shall define in just a moment) and so we use A^c to denote the complement.

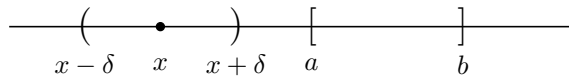


Figure 2.6. Closed intervals are closed.

Definition 2.4.4 (Closed set). Suppose that (X, d) is a metric space and that $A \subseteq X$. The set A is *closed* if the complement of A in X , $A^c = X \setminus A$, is open.

Example 2.4.5. In Example 2.3.7 we learned that the set

$$Q = \{ (x, y) : x > 0 \text{ and } y > 0 \}$$

is open in \mathbb{R}^2 (where \mathbb{R}^2 has the Euclidean metric). Therefore Q^c is closed in \mathbb{R}^2 . Q^c consists of all points (x, y) for which either $x \leq 0$ or $y \leq 0$ (or both).

Example 2.4.6 (Closed intervals are closed). Consider the closed interval $[a, b] \subseteq \mathbb{R}$ (where \mathbb{R} has the standard metric). Then $[a, b]$ is a closed set. For choose any $x \in [a, b]^c$. If $x < a$, choose $0 < \delta < a - x$ and observe that the open interval $(x - \delta, x + \delta)$ (that is, the open ball $B_\delta(x)$) is contained in $[a, b]^c$ (see Figure 2.6). If $x > b$, choose $0 < \delta < x - b$ and observe again that $(x - \delta, x + \delta) = B_\delta(x) \subseteq [a, b]^c$. We have shown that $[a, b]^c$ is open and therefore that $[a, b]$ is closed.

Some people are initially tempted to believe, upon learning Definition 2.4.4, that every subset of a metric space is either open or closed, but not both. This is *not true*. Since X and the empty set \emptyset are always open, Definition 2.4.4 tells us that X and \emptyset are both always closed as well. Thus it is possible for a set to be both open and closed at the same time (though for some metric spaces — including the real line — the empty set and the whole space are the only subsets with this property; see Exercise 2.37). On the other hand, sets that are *neither* open nor closed are, in most metric spaces, very common.

Example 2.4.7 (A set that is neither open nor closed). Consider the interval $(0, 1]$ in \mathbb{R} (where \mathbb{R} has the standard metric). The point 1 lies in $(0, 1]$, but any open interval $(1 - r, 1 + r)$ centered at 1 contains points that are not in $(0, 1]$; therefore $(0, 1]$ is not open. On the other hand, the point 0 lies in $(0, 1]^c$, but any open interval $(-r, r)$ centered at 0 contains points that are in $(0, 1]$; therefore $(0, 1]^c$ is not open either, and so $(0, 1]$ is not closed.

Example 2.4.8. Consider the plane \mathbb{R}^2 , equipped with the Euclidean metric. Write A for the x -axis. Then A is closed in \mathbb{R}^2 . For choose any $(p, q) \in A^c$; by the definition of the x -axis, $q \neq 0$. Then there is an open ball around (p, q) that is completely contained in A^c . (You should, right now, draw a picture that shows this vividly.) For choose $0 < r < |q|$. Then if $(x, y) \in B_r((p, q))$, we have

$$\begin{aligned} d((x, y), (p, q)) &= \sqrt{(x - p)^2 + (y - q)^2} < r < |q| \\ \implies \sqrt{(y - q)^2} &< r < |q| \\ \implies |y - q| &< r < |q|. \end{aligned}$$

The last line above says that the distance from y to q is less than the distance from q to 0; we conclude that y cannot be equal to zero and so see that $(x, y) \in A^c$. Thus A^c is open, and therefore A is closed.

DeMorgan's laws and Theorem 2.3.9 yield the following theorem.

Theorem 2.4.9. *A finite union of closed sets is closed. An arbitrary intersection of closed sets is closed.*

Definition 2.4.10 (Intersecting sets). We say that two sets A and B *intersect* if $A \cap B \neq \emptyset$.

Definition 2.4.11 (Closure). Suppose that (X, d) is a metric space and that $A \subseteq X$. The *closure* of A (in X) is the set denoted \bar{A} and defined as follows:

$$\bar{A} = \{x \in X : \text{for all } \varepsilon > 0, B_\varepsilon(x) \cap A \text{ is not empty}\}.$$

Otherwise put, \bar{A} consists of all points $x \in X$ such that, given any $\varepsilon > 0$, $B_\varepsilon(x)$ intersects A .

Looking at Definition 2.4.11, we see that there are a few different ways that a point can be in the closure \bar{A} of A . First, any point $x \in A$ is clearly in \bar{A} , since given any $\varepsilon > 0$ the intersection of $B_\varepsilon(x)$ and A contains (at least) the point x itself. If there is some $\varepsilon > 0$ such that this intersection contains *only* the point x itself, x is called an *isolated* point of A .

Definition 2.4.12 (Isolated point). A point x of $A \subseteq X$ is called an *isolated point* of A if there is some $\varepsilon > 0$ such that $B_\varepsilon(x) \cap A = \{x\}$.

If x is in \bar{A} but is not an isolated point of A , then, whether x is in A or not, for every $\varepsilon > 0$ the intersection $B_\varepsilon(x) \cap A$ contains some point *other than* x . In this case we say that the point x is a *limit point* of A .

Definition 2.4.13 (Limit point). Let $A \subseteq X$. The point $x \in X$ is a *limit point* of A if, given any $\varepsilon > 0$, the open ball $B_\varepsilon(x)$ intersects A in a point other than x itself: that is,

$$\text{for all } \varepsilon > 0, B_\varepsilon(x) \cap (A \setminus \{x\}) \neq \emptyset.$$

Limit points are also sometimes called *cluster points* or *accumulation points*.

If $x \in A$ but x is not a limit point of A , then x is an isolated point of A ; therefore *every point of A is either a limit point of A or an isolated point of A , but not both*. We emphasize, though, that a limit point of A might or might not be a member of A .

Thus \bar{A} consists (potentially) of three kinds of points: isolated points of A (which always belong to A), limit points of A that also belong to A (sometimes called *non-isolated* points of A), and limit points of A that do not belong to A . Thus we can say that \bar{A} *consists of the union of A and all limit points of A* ; this description of \bar{A} is sometimes used as its definition. We can also say that \bar{A} consists of the union of all isolated points of A and all limit points of A .

Given a set A , it is precisely the elements of \bar{A} that can be “approximated arbitrarily closely” by elements of A . We shall see that closures are important in understanding convergence (see Chapter 3).

Example 2.4.14. Let $X = \mathbb{R}$, equipped with the standard metric.

Consider the set $A = \{0\} \cup (1, 2]$. Then the following hold.

- The point 0 is an isolated point of A : for instance, the interval $(-\frac{1}{2}, \frac{1}{2})$ (which is just the open ball of radius $\frac{1}{2}$ about 0) intersects A only in the single point $\{0\}$.
- The point 1 is a limit point of A that is not in A : given any $\varepsilon > 0$, the open ball $(1 - \varepsilon, 1 + \varepsilon)$ of radius ε about 1 intersects A .
- Every point of $(1, 2]$ is both a member of A and a limit point of A : given any $x \in (1, 2]$ and any $\varepsilon > 0$, the open ball $(x - \varepsilon, x + \varepsilon)$ of radius ε about x intersects A in some point of A (infinitely many, in fact) other than x .

Thus $\bar{A} = \{0\} \cup [1, 2]$.

Example 2.4.15. Consider the set $B = [0, 1] \subseteq \mathbb{R}$. The following hold.

- Every point of B is a limit point of B : given any $x \in B$ and any $\varepsilon > 0$, the open ball $(x - \varepsilon, x + \varepsilon)$ intersects B in some point of B (infinitely many, in fact) other than x .
- B has no limit points that do not belong to B : for since B^c is open (recall Example 2.4.6), if $p \notin B$ there is some $\varepsilon > 0$ such that the open ball $(p - \varepsilon, p + \varepsilon)$ does not intersect B .

Thus $\bar{B} = B$.

Example 2.4.16. Consider the set $C \subseteq \mathbb{R}$, where

$$C = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\} = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\}.$$

The following hold.

- Every point of C is isolated (see Reading Questions), and so no point of C is a limit point of C .
- The point 0 is a limit point of C : for given any $\varepsilon > 0$, by the little reciprocals lemma (Theorem 1.5.17) there is some $n \in \mathbb{N}$ such that $0 < \frac{1}{n} < \varepsilon$, which tells us that the open ball $(-\varepsilon, \varepsilon)$ of radius ε about 0 intersects C .
- C has no limit points other than 0: for choose any point p that is neither in C nor equal to 0. If $p < 0$, then the open ball of radius $|p|/2$ about p contains only negative numbers and so does not intersect C . If $p > 1$, then the open ball of radius $(p-1)/2$ about p contains only numbers greater than 1 and so does not intersect C . If $p \in (0, 1)$, then p actually lies between two “successive” members of C — that is, p is in an open interval of the form

$$\left(\frac{1}{k+1}, \frac{1}{k} \right)$$

that does not intersect C . Since open intervals are open sets (Theorem 2.3.8), there is some open interval centered at p that is contained in $(\frac{1}{k+1}, \frac{1}{k})$ and so does not intersect C .

Thus $\bar{C} = C \cup \{0\}$.

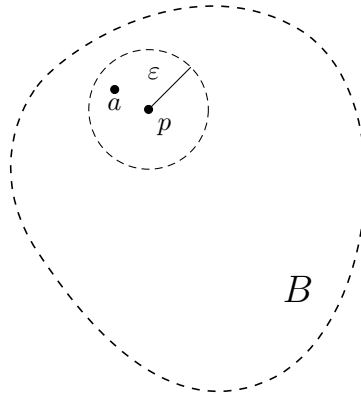


Figure 2.7. Lemma 2.4.18.

Example 2.4.17. Consider the set $\mathbb{Z} \subseteq \mathbb{R}$. The following hold.

- Every point of \mathbb{Z} is isolated: for if $k \in \mathbb{Z}$, $(k - \frac{1}{2}, k + \frac{1}{2})$ intersects \mathbb{Z} only in the point $\{k\}$.
- \mathbb{Z} has no limit points: for choose any $p \notin \mathbb{Z}$. Then p lies in an open interval of the form $(k, k + 1)$, where k is an integer. Again because open intervals are open sets, there is some open interval centered at p that is contained in $(k, k + 1)$ and so does not intersect \mathbb{Z} .

Thus $\overline{\mathbb{Z}} = \mathbb{Z}$.

Let A and B be two sets in some metric space. It is perfectly possible for B and \overline{A} to intersect even if B and A do not. (Consider, for example, $B = [1, 2] \subseteq \mathbb{R}$ and $A = (0, 1) \subseteq \mathbb{R}$: B does not intersect A , but B does contain the limit point 1 of A .) However, if B is open, the situation is different because, roughly speaking, if p is in B then points “close” to p are in B also. More precisely: suppose that A and B are subsets of a metric space (X, d) , that B is open, that $p \in \overline{A}$, and that $p \in B$. Then there is some $\varepsilon > 0$ for which $B_\varepsilon(p) \subseteq B$; but by the definition of closure, there is some $a \in A$ (perhaps equal to p , perhaps not) that is in $B_\varepsilon(p)$ and hence in B (see Figure 2.7). We have proven the following lemma.

Lemma 2.4.18. *Suppose that (X, d) is a metric space and that $B \subseteq X$ is open. If B and \overline{A} intersect, then B and A intersect.*

We can now answer an important question: are closures closed? That is, if A is a set and \overline{A} is its closure, is \overline{A} itself a closed set? The answer is yes.

Theorem 2.4.19 (Closures are closed). *Let (X, d) be a metric space and let $A \subseteq X$. The closure \overline{A} of A in X is closed in X .*

Proof. Choose $q \in (\overline{A})^c$. This means that there is some $\varepsilon > 0$ such that $B_\varepsilon(q) \cap A = \emptyset$ — that is, $B_\varepsilon(q)$ is contained in A^c . Since $B_\varepsilon(q)$ is open (Theorem 2.3.8) and $B_\varepsilon(q)$ and A do not intersect, Lemma 2.4.18 tells us that $B_\varepsilon(q)$ and \overline{A} also do not intersect. Thus $B_\varepsilon(q)$ is actually contained in $(\overline{A})^c$ as well! It follows that $(\overline{A})^c$ is open. \square

Here is our main theorem for this section.

Theorem 2.4.20. *Let (X, d) be a metric space. A set $A \subseteq X$ is closed if and only if $A = \bar{A}$.*

Analytical Advice 2.4.21. Informally, we can describe Theorem 2.4.20 as saying that “the closed sets are the sets that contain all their limit points.” Informal shorthand like this can be helpful.

Proof of Theorem 2.4.20. Since \bar{A} always contains A , the statement that $A = \bar{A}$ is equivalent to the statement that A contains \bar{A} . Therefore what we will actually prove is

$$A \text{ is closed} \iff A \text{ contains } \bar{A}.$$

Suppose that p lies in \bar{A} , but that $p \notin A$ — that is, that $p \in \bar{A} \cap A^c$. By the definition of closure, $B_\varepsilon(p)$ intersects A for every $\varepsilon > 0$ — otherwise put, there is *no* $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq A^c$. We conclude that A^c is *not* open and so A is *not* closed. We have proven (the contrapositive of)

$$A \text{ is closed} \implies A \text{ contains } \bar{A}.$$

Conversely, suppose that A contains \bar{A} , and pick any point $q \in A^c$. Since A contains \bar{A} and q is not in A , q is not in \bar{A} either. Thus there is *some* $\varepsilon > 0$ such that $B_\varepsilon(q)$ does not intersect A — that is, $B_\varepsilon(q) \subseteq A^c$. Thus A^c is open, and so A is closed. \square

Analytical Advice 2.4.22 (Memorization). So far in this chapter we have met an almost dizzying array of definitions, examples, and facts. You will be able to master analysis much more easily if you do your best to commit them (or a large number of them) to memory. If you constantly have to interrupt yourself to look something up or laboriously call something to mind, thinking through new arguments becomes almost impossible.

Reading Questions.

Reading Question 2.4.1. Let X be the open interval $(0, 1)$, viewed as a metric space in its own right (equipped with the standard metric), and let $A = [1/2, 1)$. Is A closed in X ? Explain.

Reading Question 2.4.2. Describe what it means for a point p to *not* be a limit point of a set A .

Reading Question 2.4.3. Give examples of the following:

- (a) A metric space X such that *every* point of X is non-isolated;
- (b) A metric space X such that *every* point of X is isolated;
- (c) A metric space X that has both isolated and non-isolated points.

Reading Question 2.4.4. Find the closures of the following subsets of \mathbb{R} :

- (a) $(0, 1)$
- (b) $(0, 1]$
- (c) $[0, 1]$
- (d) \mathbb{N}
- (e) $(0, 1) \cup \{5\}$.

Reading Question 2.4.5. Recall the following subset $C \subseteq \mathbb{R}$ discussed in Example 2.4.16:

$$C = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\} = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\}.$$

Show that every point of C is isolated.

Reading Question 2.4.6. Find a subset of \mathbb{R} that is neither open nor closed.

Reading Question 2.4.7. Find a subset of \mathbb{R} that is both open and closed.

Reading Question 2.4.8. Explain: if X has the discrete metric, then every subset of X is both open and closed.

2.5 Interior, closure, and boundary

We have seen that, given a subset A in a metric space X , A might be neither open nor closed. In Section 2.4, though, we found that each subset A has a closed set — namely, its closure \bar{A} — that is closely associated with it. In this section we introduce an open set that is similarly associated with A : the so-called *interior* of A . The closure and interior of a set A tell us a lot about A itself.

Definition 2.5.1 (Interior point; interior). Suppose that X is a metric space and that $A \subseteq X$. The point $p \in A$ is called an *interior point of A* if there is some $r > 0$ such that $B_r(p) \subseteq A$.

The union of all the interior points of A is called the *interior of A* and is written A° :

$$A^\circ = \{p \in A : B_r(p) \subseteq A \text{ for some } r > 0\}.$$

By definition, $A^\circ \subseteq A$. Suppose that A° is actually equal to A . Then, since every point of A is an interior point of A there is, for every $a \in A$, some $r > 0$ such that $B_r(a) \subseteq A$; this is precisely the definition of A being open. Conversely, if A is open, then every point of A is an interior point. We have established the following proposition (compare to Theorem 2.4.20).

Proposition 2.5.2. *A subset A in a metric space X is open if and only if $A = A^\circ$.*

Similarly to how \bar{A} is a closed set (Theorem 2.4.19), A° is open.

Theorem 2.5.3 (The interior is open). *Suppose that X is a metric space and that $A \subseteq X$. Then A° is an open set in X .*

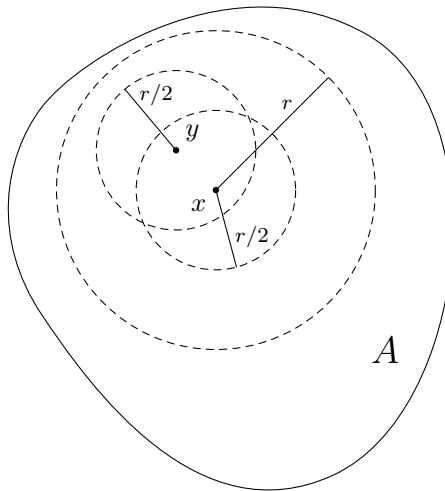


Figure 2.8. A° is open.

Proof. Suppose that p is *not* in the interior of A . This means that, for every $\varepsilon > 0$, $B_\varepsilon(p)$ intersects A^c ; this in turn means that $p \in \overline{A^c}$. Therefore $p \in (A^\circ)^c$ if and only if $p \in \overline{A^c}$ — otherwise put, $(A^\circ)^c = \overline{A^c}$. Since the closure of any set is closed (Theorem 2.4.19), we conclude that $(A^\circ)^c$ is closed and hence that A° is open. \square

A second proof of Theorem 2.5.3 is suggested by Figure 2.8. Choose $x \in A^\circ$. By the definition of interior, there is some $r > 0$ such that $B_r(x) \subseteq A$. Now choose any $y \in B_{r/2}(x)$. The triangle inequality shows that $B_{r/2}(y)$ is contained in $B_r(x)$ and hence in A , since if $z \in B_{r/2}(y)$ we have

$$d(x, z) \leq d(x, y) + d(y, z) < r/2 + r/2 = r.$$

Thus y is in A° ; but since y was an arbitrary point of $B_{r/2}(x)$ this shows that $B_{r/2}(x) \subseteq A^\circ$, and so x is actually an interior point of A° . Thus A° is equal to its own interior and so is open by Proposition 2.5.2.

Analytical Advice 2.5.4. Multiple proofs of a single theorem often contribute to our understanding. Our first proof of Theorem 2.5.3 relied on the fact that closures are closed, while the second was (in my view) a shade more complicated but relied more purely on the definitions of interior point and open set.

Note that, for any set $A \subseteq X$, we always have $A^\circ \subseteq A \subseteq \overline{A}$. The points that are in \overline{A} but not in A° are, loosely speaking, on the “edge” of A — they are either limit points or isolated points of A , but are not interior to A . This intuitive picture motivates the following definition.

Definition 2.5.5 (Boundary). Suppose that X is a metric space and that $A \subseteq X$. The set $\overline{A} \setminus A^\circ$ is called the *boundary of A* and is written ∂A .

The following theorem is of basic importance.

Theorem 2.5.6 (Facts about interior, closure and boundary). *Suppose that X is a metric space and that $A \subseteq X$. We have the following.*

- (a) $(A^\circ)^c = \overline{A^c}$ (the complement of the interior is the closure of the complement).
- (b) $(A^c)^\circ = \overline{A}^c$ (the interior of the complement is the complement of the closure).
- (c) $x \in \partial A$ if and only if $B_\varepsilon(x)$ intersects both A and A^c for every $\varepsilon > 0$.
- (d) $\partial A = \partial(A^c)$.

Proof. See Exercise 2.22 (though observe that we proved part (a) in the proof of Theorem 2.5.3). □

Note that part (c) of the above theorem explains why the boundary is called the boundary. Part (c) also shows that any set A induces a division of X into three pairwise disjoint subsets: A° , $(A^c)^\circ$, and ∂A (any one of which might be empty in any particular case).

Theorem 2.5.7. *Suppose that X is a metric space and that $A \subseteq X$. Then*

- (a) A° is precisely the union of all the open sets contained in A ;
- (b) \overline{A} is precisely the intersection of all the closed sets containing A .

Remark 2.5.8. The slogan version of Theorem 2.5.7 is that “ A° is the largest open set contained in A , and \overline{A} is the smallest closed set containing A .”

Proof of Theorem 2.5.7. Write U for the union of all the open sets contained in A . Since A° is itself an open set contained in A (Theorem 2.5.3), we certainly have that $A^\circ \subseteq U$. Conversely, choose any $p \in U$. Then $p \in V$ for some open set V contained in A , just by the definition of U . Since V is open, $B_r(p) \subseteq V \subseteq A$ for some $r > 0$. Thus p is an interior point of A ; since p was an arbitrary point of U , we have that $U \subseteq A^\circ$. This proves part (a).

Similarly, write I for the intersection of all closed sets containing A . Since \overline{A} is a closed set containing A (Theorem 2.4.19), $I \subseteq \overline{A}$. Conversely, let $p \in \overline{A}$, and choose an arbitrary closed set W containing A . Since $p \in \overline{A}$, any open ball around p intersects A and hence W . Thus $p \in \overline{W}$; since W is closed, $W = \overline{W}$ (Theorem 2.4.20) and so $p \in W$. We have shown that $\overline{A} \subseteq W$ for any closed set W containing A ; thus $\overline{A} \subseteq I$ also. This completes the proof. □

Reading Questions.

Reading Question 2.5.1. For each of the following subsets of \mathbb{R} (where \mathbb{R} has its usual metric), give the interior, the closure, and the boundary.

- (a) $(0, 1)$
- (b) $(0, 1]$
- (c) $[0, 1]$

- (d) \mathbb{Z}
 (e) $\{1/n : n \in \mathbb{N}\}$
 (f) $\{x \in \mathbb{R} : x > 0\}$.

Reading Question 2.5.2. Suppose that X is a metric space and that $A \subseteq X$. Suppose that $U \subseteq X$ is an open set and that $V \subseteq X$ is a closed set.

- (a) True or false: if $U \subseteq A$, then $U \subseteq A^o$ also.
 (b) True or false: if $A \subseteq V$, then $\bar{A} \subseteq V$ also.
 (c) True or false: if $A \subseteq U$, then $\bar{A} \subseteq U$ also.
 (d) True or false: if $V \subseteq A$, then $V \subseteq A^o$ also.

2.6 Dense subsets

In different branches of mathematics we have various ways of characterizing a “large” subset of a given set: for example, we might say that A is a “large” subset of X if its complement is finite, or if choosing a random member of X yields a member of A with high probability. In this section, we introduce one notion (not the only one!) of a subset being “large” that is useful in analysis.

Definition 2.6.1 (Dense). Let X be a metric space. A subset $A \subseteq X$ is called *dense in* X if $\bar{A} = X$.

Definition 2.6.1 says that, if A is dense in X , every point of X is either a member of A or a limit point of A . Observe that a metric space X is always dense in itself. Just restating Definition 2.4.11 yields the following alternative characterization, which is sometimes useful: A is dense in X if and only if, given any point $p \in X$ and any $r > 0$, $B_r(p) \cap A$ is not empty.

The following proposition furnishes us with one of our most important examples of a dense subset.

Proposition 2.6.2. *The rational numbers \mathbb{Q} are dense in \mathbb{R} .*

Proof. Let x be any real number and let r be any positive number; we need to show that $B_r(x) = (x - r, x + r)$ contains a rational number. Invoking the little reciprocals lemma (Theorem 1.5.17), let us choose an $n \in \mathbb{N}$ such that $0 < 1/n < r$, and consider the set

$$M_n = \left\{ \frac{k}{n} : k \in \mathbb{Z} \right\}.$$

Observe that M_n is closed: the complement of M_n is a union of open intervals of the form $\left(\frac{k}{n}, \frac{k+1}{n}\right)$. Note too that all the members of M_n are rational, and that “successive” members of M_n are $1/n$ units apart.

If we imagine that the interval $(x - r, x + r)$ does not intersect M_n , then the largest member of M_n that is less than x and the smallest member of M_n that is greater than x are successive members of M_n but are at least $2r > 2/n$ units apart, which is ridiculous. □

Remark 2.6.3. If X is a metric space and $A \subseteq X$ is dense, then any element p of X can be “approximated with arbitrary precision” by a member a of A — that is, we can find a member of A that is as close as we want to p . Doing this can be especially useful if the elements of A are in some way familiar or easy to understand. Most people have a better intuitive grasp of the rationals than the reals, for example, so it’s nice to know that any real number can be approximated by a rational.

Reading Questions.

Reading Question 2.6.1. Draw the set M_n defined in the proof of Proposition 2.6.2 for a couple of choices of n .

Reading Question 2.6.2. Give an example of a subset of \mathbb{Q} (other than \mathbb{Q} itself) that is dense in \mathbb{R} .

Reading Question 2.6.3. Give an example of a dense subset of \mathbb{R} that is *not* contained in \mathbb{Q} .

Reading Question 2.6.4. True or false: the complement of a dense set is not dense.

2.7 Equivalent metrics

We have already seen that there are many different metrics that we can put on a set. As we will learn, it turns out that many of the core ideas in analysis (convergence of sequences and continuity of functions, for example) don’t depend as much on precisely how we measure distance in a metric space as on which subsets of the metric space are open. In this section we define what it means for two metrics to be “equivalent,” and show that two equivalent metrics on a set generate exactly the same open subsets.

In this section, since we shall be talking about multiple metrics at once, we are going to need the somewhat more elaborate notation for open balls introduced immediately after Definition 2.3.1: if X is equipped with the metric d (and perhaps some other metrics as well), we write $B_r^d(x)$ for the open ball about x of radius r with respect to the metric d .

Definition 2.7.1 (Equivalent metrics). Suppose that X is a set equipped with two metrics d and ρ . The metrics d and ρ are *equivalent* if the following holds. Given any $x \in X$ and any $\varepsilon > 0$, there is some $\delta_1 > 0$ such that

$$B_{\delta_1}^{\rho}(x) \subseteq B_{\varepsilon}^d(x)$$

and some $\delta_2 > 0$ such that

$$B_{\delta_2}^d(x) \subseteq B_{\varepsilon}^{\rho}(x).$$

Intuitively, this definition says that we “can fit some ρ -ball inside any given d -ball and can fit some d -ball inside any given ρ -ball.”

Example 2.7.2. Consider the plane \mathbb{R}^2 . The open balls in \mathbb{R}^2 with respect to the Euclidean metric are discs; the open balls in \mathbb{R}^2 with respect to the sup metric are squares. We can readily convince ourselves with a picture that any disc contains a square with the same center and that any square contains a disc with the same center (see Figure 2.9), and thus that the Euclidean and sup metrics are equivalent in \mathbb{R}^2 .

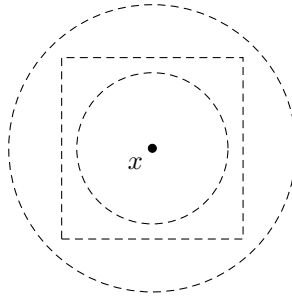


Figure 2.9. The Euclidean and sup metrics are equivalent in \mathbb{R}^2 .

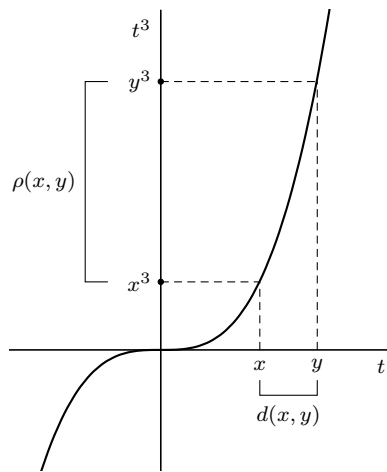


Figure 2.10. The distances $d(x, y)$ and $\rho(x, y)$.

It is not too hard to incorporate the basic idea in Figure 2.9 into a rigorous proof, or to generalize it; in fact, we have the following.

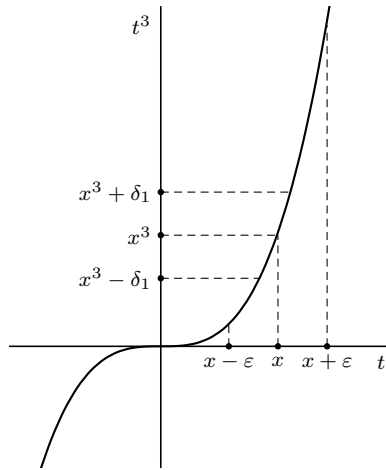
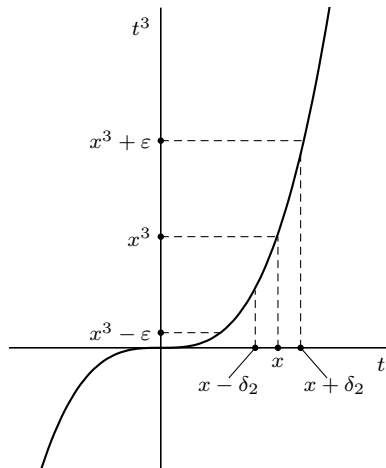
Proposition 2.7.3. *The Euclidean metric, sup metric, and ℓ_1 metric on \mathbb{R}^n are all equivalent.*

Proof. See the Reading Questions and Exercise 2.31. □

Example 2.7.4. Consider the metric on \mathbb{R} given by $\rho(x, y) = |x^3 - y^3|$. (You are asked to verify in Exercise 2.5 that ρ is indeed a metric.) For this example, we denote the standard metric on \mathbb{R} by d .

The metrics d and ρ are equivalent. We prove (a generalization of) this statement in Exercise 5.28; for now, we content ourselves with a pictorial illustration of the idea.

If we graph the function $t \mapsto t^3$ and plot two numbers $x, y \in \mathbb{R}$ on the horizontal axis, then the (ordinary) distance between x and y on the horizontal axis is $d(x, y)$ and the (ordinary) distance between x^3 and y^3 on the vertical axis is $\rho(x, y)$. See Figure 2.10.

Figure 2.11. $B_{\delta_1}^{\rho}(x) \subseteq B_{\epsilon}^d(x)$.Figure 2.12. $B_{\delta_2}^d(x) \subseteq B_{\epsilon}^{\rho}(x)$.

Now, let $x \in \mathbb{R}$ and $\epsilon > 0$ be given. Then we can find some $\delta_1 > 0$ such that, if $|x^3 - y^3| < \delta_1$, then we must have $|x - y| < \epsilon$ (such a δ_1 is illustrated in Figure 2.11). This is the same as saying that $\rho(x, y) < \delta_1$ implies that $d(x, y) < \epsilon$, which in turn is the same as saying that $B_{\delta_1}^{\rho}(x) \subseteq B_{\epsilon}^d(x)$.

Again let $x \in \mathbb{R}$ and $\epsilon > 0$ be given. Then we can find some $\delta_2 > 0$ such that, if $|x - y| < \delta_2$, then we must have $|x^3 - y^3| < \epsilon$ (such a δ_2 is illustrated in Figure 2.12). This is the same as saying that $d(x, y) < \delta_2$ implies that $\rho(x, y) < \epsilon$, which in turn is the same as saying that $B_{\delta_2}^d(x) \subseteq B_{\epsilon}^{\rho}(x)$. Thus d and ρ are equivalent.

Analytical Advice 2.7.5. The existence of the numbers δ_1 and δ_2 as described in Example 2.7.4 is completely plausible from Figures 2.11 and 2.12; but we do not

regard these figures alone as constituting a proof. Carefully proving the existence of δ_1 and δ_2 , while not a trivial task, is absolutely doable and we will build up the tools to do so over the next few chapters.

Even once it is easy for you to formulate such proofs, you may never find them as illuminating as good pictures (I don't). It turns out, though, that part of what we learn as we carefully build up a rigorous theory is how to recognize when the intuition conferred by pictures is not misleading; we come to view pictures as *illustrating* abstract arguments as much as *motivating* them. As you draw pictures — hopefully lots of them — during the (life-long!) process of learning real analysis, watch for this change!

While the open *balls* for two equivalent metrics are typically different, the open *sets* are the same.

Theorem 2.7.6. *Suppose that X is a set with metrics d and ρ . Then d and ρ are equivalent if and only if the open sets in (X, d) are exactly the same as the open sets in (X, ρ) .*

The first part of the proof of Theorem 2.7.6 is illustrated in Figure 2.13; you will find it helpful to refer to the figure as you read the proof. In the figure, the open balls in the metric d are drawn as disks, while the open balls in the metric ρ are drawn as squares.

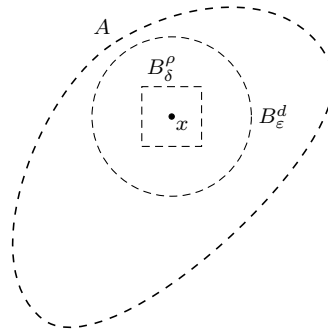


Figure 2.13. Illustration of the first half of the proof of Theorem 2.7.6.

Proof of Theorem 2.7.6. Assume that d and ρ are equivalent. Suppose that A is open with respect to d , and choose $x \in A$. Then there is some $\epsilon > 0$ such that $B_\epsilon^d(x) \subseteq A$. By the definition of equivalent metric, there is some $\delta > 0$ such that $B_\delta^\rho(x) \subseteq B_\epsilon^d(x) \subseteq A$; but since $x \in A$ was arbitrary this shows that A is open with respect to ρ . Essentially the same argument shows that if A is open with respect to ρ , then A is open with respect to d .

Now suppose that the open sets in (X, d) are exactly the same as the open sets in (X, ρ) (that is, $A \subseteq X$ is open in (X, d) if and only if it is open in (X, ρ)). Let $\epsilon > 0$ be given. The open ball $B_\epsilon^d(x)$ is open in (X, d) (Theorem 2.3.8) and so is open in (X, ρ) ; thus there is some $\delta_1 > 0$ for which $B_{\delta_1}^\rho(x) \subseteq B_\epsilon^d(x)$. Essentially the same argument shows that there is some $\delta_2 > 0$ for which $B_{\delta_2}^d(x) \subseteq B_\epsilon^\rho(x)$. Thus d and ρ are equivalent. \square

As mentioned above, in many situations in analysis it is which sets are open, rather than the precise manner of measuring distance, that matters. This turns out to mean that, when faced with multiple equivalent metrics, we can often use whichever one is most convenient; this will become clearer in future chapters.

Remark 2.7.7. There is another notion of “equivalence” between metrics that is frequently used. If the set X is equipped with two metrics d and ρ , the metrics d and ρ are *strongly equivalent* if there are two numbers $c_1 > 0$ and $c_2 > 0$ such that, if x and y are any two points in X , then

$$c_1 d(x, y) \leq \rho(x, y) \leq c_2 d(x, y).$$

(What we are calling “equivalent metrics” some authors call “topologically equivalent metrics,” and what we are calling “strongly equivalent metrics” some authors call simply “equivalent metrics.” Whenever you encounter the term *equivalent metrics*, read carefully to see which sense is meant.)

We explore some of the properties of strongly equivalent metrics in Exercise 2.41, including the fact that two strongly equivalent metrics are equivalent. It is possible, though, for two metrics to be equivalent without being strongly equivalent (Exercise 5.32).

We close this section with a brief discussion of metrics on Cartesian products of metric spaces. Suppose that X and Y are metric spaces with metrics d_X and d_Y respectively. If we want to put a metric on the set $X \times Y$, how should we do it? There is no really standard way to do this; note that, after all, we have already discussed several different metrics on \mathbb{R}^2 , which is the same set as $\mathbb{R} \times \mathbb{R}$. We do, however, usually insist that, if $(x_1, y_1) \in X \times Y$ and $(x_2, y_2) \in X \times Y$ are to be regarded as “close,” it should be true *both* that x_1 is close to x_2 in X *and* that y_1 is close to y_2 in Y . Accordingly, we make the following definition.

Definition 2.7.8 (Product metric). Let $n \in \mathbb{N}$, and let $X_1, X_2, X_3, \dots, X_n$ be n metric spaces. Write d_k for the metric on X_k . The *sup product metric* on $X_1 \times X_2 \times \dots \times X_n$ is the metric ρ on $X_1 \times X_2 \times \dots \times X_n$ given by

$$\rho((x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n)) = \max_{k=1}^n \{d_k(x_k, y_k)\}.$$

A *product metric* on $X_1 \times X_2 \times \dots \times X_n$ is any metric that is equivalent to ρ .

So, under the terminology of Definition 2.7.8, the sup metric is the sup product metric on \mathbb{R}^n ; by Proposition 2.7.3, the sup metric, the ℓ_1 metric, and the Euclidean metric are all product metrics on \mathbb{R}^n .

Remark 2.7.9. The term “sup product metric” is *not* standard. At least for Cartesian products of finitely many metric spaces, though, the product metrics (as we have defined them) are the most usual choices of metrics.

Reading Questions.

Reading Question 2.7.1. Draw more pictures like Figure 2.9 to illustrate Proposition 2.7.3 in \mathbb{R}^2 .

Reading Question 2.7.2. Find two metrics on \mathbb{R} that are *not* equivalent.

2.8 Normed vector spaces

In analysis, the most important metric spaces are also vector spaces. If X is a vector space, then the points of X are also called *vectors*. In a vector space X , two vectors can be added together to yield a new vector, and any vector can be multiplied by a *scalar* to yield a new vector. (A scalar is a real number if X is a *real vector space* and a complex number if X is a *complex vector space*.) For example, \mathbb{R}^n is a real vector space for any $n \in \mathbb{N}$; in Section 1.6, we described how addition and scalar multiplication work in \mathbb{R}^n . If you have not studied vector spaces before, while reading this book you can think about \mathbb{R}^n in place of general vector spaces without being misled.

In Section 2.2 we saw that the Euclidean distance $d(x, y)$ between two vectors x and y in \mathbb{R}^n is the same as $\|x - y\|_2$, the “length” or “size” of the vector $x - y$. It turns out that many important metrics on vector spaces have basically the same interpretation; this is the topic of this section.

Definition 2.8.1 (Norm). Suppose that X is a real or complex vector space. A *norm* on X is a function $\nu : X \rightarrow \mathbb{R}$ satisfying the following properties, where x and y are vectors (i.e. elements of X), s is a scalar, and we are writing $\nu(x) = \|x\|$.

- Positivity: $\|x\| \geq 0$, and $\|x\| = 0$ if and only if x is the zero vector;
- Scaling: $\|sx\| = |s|\|x\|$ for any scalar s and any vector x ;
- Triangle inequality (for norms): $\|x + y\| \leq \|x\| + \|y\|$ for any two vectors x and y .

A vector space equipped with a norm is called a *normed vector space*.

We think of $\|x\|$ as a measure of the “length” or “size” of the vector x .

Compare the triangle inequality for norms to the “0th Euclidean triangle inequality” in Section 1.6, and the corresponding inequality for complex numbers introduced in Example 2.2.6.

Some important norms on \mathbb{R}^n are

- The Euclidean norm (recall Section 1.6):

$$\|(x_1, \dots, x_n)\|_2 = \sqrt{\sum_{i=1}^n x_i^2}.$$

- The sup norm:

$$\|(x_1, \dots, x_n)\|_\infty = \max\{|x_1|, \dots, |x_n|\}.$$

- The ℓ_1 norm:

$$\|(x_1, \dots, x_n)\|_1 = \sum_{i=1}^n |x_i|.$$

These examples illustrate the basic link between norms and metrics, which you might have already guessed.

Definition 2.8.2 (Metric induced by a norm). Let X be a normed vector space with norm ν , and write $\nu(x) = \|x\|$. The function $d : X \times X \rightarrow \mathbb{R}$ given by

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

is called the *metric induced by the norm ν* .

An obviously important fact is that the metric induced by a norm is indeed a metric.

Proposition 2.8.3. *Let X be a normed vector space with norm ν , and write $\nu(x) = \|x\|$. Let $d : X \times X \rightarrow \mathbb{R}$ be given by*

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

Then d is a metric.

Proof. See Exercise 2.33. □

Note that our usual metrics on \mathbb{R}^n are induced by norms: the Euclidean norm induces the Euclidean metric, the sup norm induces the sup metric, and the ℓ_1 norm induces the ℓ_1 metric.

We repeat that, in a normed vector space (with metric induced by the norm), $\|x - y\|$ defines the distance from x to y ; so $\|x - y\|$ is another way to write the quantity $d(x, y)$. (In particular, $\|x\| = \|x - 0\|$ is the distance from x to the zero vector.) At this point you might be feeling a little annoyed: isn't all this just two sets of notation for exactly the same thing?

It is admittedly true that, very often, viewing a metric as induced by a norm does not add anything extra. But metrics induced by norms do have some nice additional properties that general metrics don't have. These properties hinge on the fact that points in a normed vector space can be added and multiplied by scalars; this is not true of general metric spaces. (Look back, for instance, at the metric space in Example 2.1.6. It's not at all clear what the sum of two elements in that metric space would mean.)

For example, in a normed vector space, if x and y are vectors and s is a scalar, by the scaling property of norms we have

$$\|sx - sy\| = s\|x - y\|.$$

Written in terms of the metric induced by the norm, this says that

$$d(sx, sy) = sd(x, y).$$

This is certainly not a property of metrics in general (consider the discrete metric on \mathbb{R} , for example). A similar feature of metrics induced by norms is that they are *translation invariant*: if x , y , and p are vectors, then

$$\|(x + p) - (y + p)\| = \|x - y\|.$$

Written in terms of the metric induced by the norm, this says that

$$d((x + p), (y + p)) = d(x, y).$$

Again, this is not a property of metrics in general.

We devote the rest of this section to another nice property of metrics induced by norms.

Definition 2.8.4 (Closed ball). Let (X, d) be a metric space, with $p \in X$ and $r > 0$. We define the *closed ball of radius r about p* to be the set

$$B_r^{\text{closed}}(p) = \{x \in X : d(p, x) \leq r\}.$$

In various applications, it is preferable to deal with closed balls rather than open balls. It seems at first blush like closed balls should be the closures of the corresponding open balls: that is,

$$B_r^{\text{closed}}(p) = \overline{B_r(p)}.$$

Unfortunately, this is not always the case. It is true that closed balls are always closed (see Exercise 2.34); so since $\overline{B_r(p)}$ is a closed set containing $B_r(p)$, by Theorem 2.5.7 we always have that $\overline{B_r(p)} \subseteq B_r^{\text{closed}}(p)$. However, equality need not always hold (see the Reading Questions).

We do, however, have the following theorem, which says that $B_r^{\text{closed}}(p) = \overline{B_r(p)}$ in normed vector spaces. (We repeat that most important metric spaces in applications are actually normed vector spaces.)

Theorem 2.8.5. *Suppose that X is a normed vector space with norm ν , and write $\nu(x) = \|x\|$. Let d be the metric induced by the norm. Then, given any $p \in X$ and $r > 0$, $B_r^{\text{closed}}(p) = \overline{B_r(p)}$.*

Proof. Choose $p \in X$ and $r > 0$. As remarked before the statement of the theorem, it follows from Exercise 2.34 and Theorem 2.5.7 that $\overline{B_r(p)} \subseteq B_r^{\text{closed}}(p)$; we wish to show that $B_r^{\text{closed}}(p) \subseteq \overline{B_r(p)}$.

We repeat that $B_r^{\text{closed}}(p)$ is the set of points that are distance less than or equal to r from the point p . If $d(p, x) < r$, then certainly $x \in B_r(p) \subseteq \overline{B_r(p)}$; it therefore remains only to show that any point x with $d(p, x) = r$ also lies in $\overline{B_r(p)}$. See Figure 2.14.

Given $\varepsilon > 0$, choose $\delta \in (0, 1)$ such that $\delta r < \varepsilon$. Consider the point $y = (1 - \delta)x + \delta p$. (It is right here that we are using the fact that X is a vector space: the point y is obtained by adding together scalar multiples of two elements of X !) We claim that $y \in B_\varepsilon(x)$ and also that $y \in B_r(p)$. Once the claim is proven, we will have established that $B_\varepsilon(x)$ intersects $B_r(p)$ for every $\varepsilon > 0$, and hence that x is in $\overline{B_r(p)}$; this will complete the proof.

Step 1: $y \in B_\varepsilon(x)$. We have

$$d(y, x) = \|y - x\| = \|(1 - \delta)x + \delta p - x\| = \|\delta(p - x)\| = \delta\|p - x\| = \delta r < \varepsilon.$$

Step 2: $y \in B_r(p)$. We have

$$d(y, p) = \|y - p\| = \|(1 - \delta)x + \delta p - p\| = \|(1 - \delta)(x - p)\| = (1 - \delta)\|x - p\| = (1 - \delta)r < r.$$

This completes the proof. \square

Reading Questions.

Reading Question 2.8.1. In this section, we introduced three norms on \mathbb{R}^n . Which of these yields the usual absolute value function in the case that $n = 1$?

Reading Question 2.8.2. Find a metric space with a metric that is *not* induced by a norm, and explain.

Reading Question 2.8.3. Consider \mathbb{R}^2 , with the sup metric. Give an explicit example of a point that is in $B_1^{\text{closed}}((0, 0))$ but not in $B_1((0, 0))$.

Reading Question 2.8.4. Let X be the real line \mathbb{R} , but equipped with the discrete metric. Give an explicit example of a point p and a radius r such that $\overline{B_r(p)} \neq B_r^{\text{closed}}(p)$.

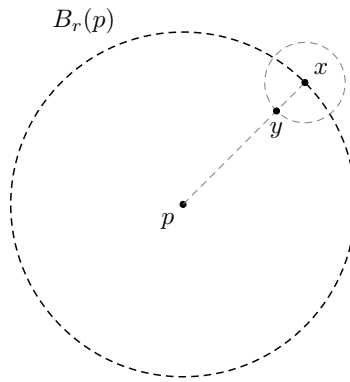


Figure 2.14. Illustration of the proof of Theorem 2.8.5.

2.9 A brief note about conventions

We have seen that there are many metrics that can be put on a given set, and in any particular instance it might not be clear which metric is meant — especially in \mathbb{R}^n , where the three main (equivalent, norm-induced) metrics we have introduced really are each frequently used. Nevertheless, there are some conventions worth mentioning. You should always, unless explicitly notified otherwise, assume that \mathbb{R} has the standard metric induced by absolute value. You should always assume that \mathbb{C} has the Euclidean metric that we introduced in Example 2.2.6. You should always assume that any Cartesian product of a finite number of metric spaces, whether it is arising in an abstract setting or is something concrete like \mathbb{R}^n , has a product metric.

2.10 Exercises

Exercises for Section 2.1.

Exercise 2.1. Recall that, if S is a set, by $|S|$ we (usually) mean the number of elements in S .

Let \mathcal{F} be the set of *finite* subsets of \mathbb{N} . Given $A, B \in \mathcal{F}$, define

$$d(A, B) = |A \setminus B| + |B \setminus A|$$

— that is, $d(A, B)$ is the number of elements that are in one, but not both, of A or B . Show that d is a metric. [Hint: use a Venn diagram for the triangle inequality.]

Exercise 2.2. Show that the discrete metric is in fact a metric.

Exercise 2.3. Let $A = \{1, 2, 3, 4, 5\}$, and let $\mathcal{M}(A)$ be the set of functions from A to itself. Given $f, g \in \mathcal{M}(A)$, define

$$d(f, g) = \max\{|f(x) - g(x)| : x \in A\}.$$

- If $f \in \mathcal{M}(A)$ is the constant function whose value is always 3 and $g = id_A$, what is $d(f, g)$?
- Show that d is a metric on $\mathcal{M}(A)$.

Exercise 2.4. Consider the function $\eta : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$\eta(x, y) = (x - y)^3.$$

Decide whether η is a metric on \mathbb{R} , and prove.

Exercise 2.5. Consider the function $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$\rho(x, y) = |x^3 - y^3|.$$

Prove that ρ is a metric on \mathbb{R} . (This metric is discussed in Example 2.7.4; see also Exercise 5.28.)

Exercises for Section 2.2.

Exercise 2.6. Show that the sup metric on \mathbb{R}^n is actually a metric.

Exercise 2.7. Show that the ℓ_1 metric on \mathbb{R}^n is actually a metric.

Exercise 2.8. This exercise is a follow-up to Reading Question 2.2.3. Suppose that x and y are two points in \mathbb{R}^n . If we measure the distance between x and y with the Euclidean, sup, and ℓ_1 metrics, is any of these distances guaranteed to be the largest? Is any guaranteed to be the smallest? Figure out the answer, and prove.

Exercise 2.9. Verify the following facts about \mathbb{C} given in Example 2.2.6: the formula for the multiplicative inverse of z ; that $|zw| = |z||w|$ for any two complex numbers z and w ; and that, if $w \neq 0$, then $|z/w| = |z|/|w|$.

Exercises for Section 2.3.

Exercise 2.10. Suppose that \mathbb{R}^2 has the sup metric. Show that the set

$$S = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > x_1\}$$

is open.

Exercise 2.11. Show that the set $\mathbb{Q} \subseteq \mathbb{R}$ of rational numbers is not open.

Exercise 2.12. Suppose that (X, d) is a metric space and that $Y \subseteq X$. View (Y, d) as a metric space in its own right. Prove: $A \subseteq Y$ is open in Y if and only if there is some set U , open in X , such that $A = Y \cap U$.

Exercise 2.13. Let A be a nonempty subset of \mathbb{R} . Define the set

$$\mathcal{N}(A) = \{p \in \mathbb{R} : p \text{ is neither an upper bound of } A \text{ nor a lower bound of } A\}.$$

Prove that $\mathcal{N}(A)$ is open.

Exercises for Section 2.4.

Exercise 2.14. Consider \mathbb{R}^2 equipped with the sup metric. Show that the x -axis is a closed set.

Exercise 2.15. Supply the details in the proof of Theorem 2.4.9.

Exercise 2.16. Let (X, d) be a metric space.

- (a) Prove that any one-point set is closed.
- (b) Prove that any finite set is closed.

Exercise 2.17. Find an example of a subset $A \subseteq \mathbb{R}$ for which every point of A is isolated, but A is not closed.

Exercise 2.18. Let (X, d) be a metric space, and let $A \subseteq X$. Show that, if p is a limit point of A , then given any $\varepsilon > 0$ there are *infinitely many* distinct members of A in $B_\varepsilon(p)$.

Exercise 2.19. Let (X, d) be a metric space. Show that if X has no isolated points, then X must be infinite.

Exercise 2.20. Let (X, d) be a metric space, and let $A \subseteq X$. Write $L(A)$ for the set of limit points of A . Show that $L(A)$ is a closed set.

Exercise 2.21. Let (X, d) be a metric space. If $S \subseteq X$, write $L(S)$ for the set of limit points of S . Prove or disprove each of the following statements.

- (a) For any two subsets $A \subseteq X$ and $B \subseteq X$, $L(A \cap B) \subseteq L(A) \cap L(B)$.
- (b) For any two subsets $A \subseteq X$ and $B \subseteq X$, $L(A) \cap L(B) \subseteq L(A \cap B)$.

Exercises for Section 2.5.

Exercise 2.22. Prove Theorem 2.5.6.

Exercise 2.23. Let (X, d) be a metric space, and let $A \subseteq X$. Show that ∂A , the boundary of A , is a closed set.

Exercise 2.24. What are the interior, closure, and boundary of the empty set?

Exercise 2.25. Prove that, if p is a limit point of ∂A , then p is a limit point of A as well.

Exercises for Section 2.6.

Exercise 2.26. Prove that $\mathbb{Q} \subseteq \mathbb{R}$ has empty interior.

Exercise 2.27. Find an example of a subset of \mathbb{R} (other than \mathbb{R} itself) that is both open and dense.

Exercise 2.28. Suppose that X is a metric space and that $A \subseteq X$ is dense in X . Show that, if X is infinite, then A must be infinite also.

Exercise 2.29. Suppose that X is a metric space, and that $A \subseteq X$. Decide whether each of the following statements is true or false, and prove:

- (a) If A° is empty, then A^c is dense.
- (b) If A^c is dense, then A° is empty.

Exercise 2.30. Suppose that X is a metric space and that $B \subseteq A \subseteq X$. Prove that, if B is dense in A and A is dense in X , then B is dense in X .

Exercises for Section 2.7.

Exercise 2.31. Prove Proposition 2.7.3.

Exercise 2.32. Consider the following function $\rho : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$:

$$\rho(x, y) = \begin{cases} |x - y|, & \text{if } |x - y| \leq 1; \\ 1, & \text{otherwise.} \end{cases}$$

- (a) Prove that ρ is a metric on \mathbb{R} .
 (b) Prove or disprove: ρ is equivalent to the standard metric on \mathbb{R} .

Exercises for Section 2.8.

Exercise 2.33. Prove Proposition 2.8.3.

Exercise 2.34. Prove that $B_r^{\text{closed}}(p)$ is always a closed set.

Exercise 2.35. Suppose that X is a normed vector space, and suppose that x and y are two points in the open ball $B_r(p)$. Show that, for any $t \in [0, 1]$,

$$tx + (1 - t)y \in B_r(p).$$

(This result says that the open ball $B_r(p)$ is *convex* — that is, the line segment joining any two points in $B_r(p)$ lies entirely in $B_r(p)$.)

Additional exercises for Chapter 2.

Exercise 2.36. Let (X, d) be a metric space, and let $A \subseteq X$. Let us say that A is *blurry* if ∂A , the boundary of A , has nonempty interior — that is, if there is some $p \in (\partial A)^\circ$. [Examples: \mathbb{Q} is a blurry subset of \mathbb{R} since $\partial \mathbb{Q}$ is all of \mathbb{R} ; $[0, 1]$ is not a blurry subset of \mathbb{R} since $\partial[0, 1] = \{0, 1\}$.] (Note: *blurry* is not a standard term.)

Prove that, if A is blurry, then A is neither open nor closed.

Exercise 2.37. Prove that the only subsets of \mathbb{R} that are *both* open and closed are \mathbb{R} and \emptyset . [Hint: imagine that $A \subseteq \mathbb{R}$ is neither empty nor all of \mathbb{R} , but is both open and closed. Then there are numbers $a \in A$ and $b \notin A$; without loss of generality, we can assume that $a < b$. Now consider the following set: $U(a) = \{r : r > a \text{ and } (a, r) \subseteq A\}$. Since A is open, $U(a)$ is nonempty; since $b \notin A$, $U(a)$ is bounded above. Therefore $s = \sup U(a)$ exists. So where does s lie — in A or in A^c ?]

Exercise 2.38. Fix the writing in the following proof by adding necessary quantifiers.

Proposition: Suppose that (X, d) is a metric space and that $A \subseteq X$ is a set such that $A = \partial A$. Then A^c is dense.

Proof. Saying that A^c is dense means that $\overline{A^c} = X$. By Definition 2.4.11, then, it is enough to prove that $B_r(p) \cap A^c$ is nonempty. If $p \in A^c$, this is clearly true since $p \in B_r(p) \cap A^c$; if $p \in A$, since $A = \partial A$ this is true by part (c) of Theorem 2.5.6. \square

Exercise 2.39. Fix the writing in the following proof by adding necessary quantifiers.

Proposition: Suppose that (X, d) is a metric space and that $A \subseteq X$ is a finite set. Then every point of A is isolated.

Proof. We need to show that $B_r(p) \cap A = \{p\}$ for $p \in A$. To see that is true, we can take

$$0 < r < \min_{q \in A \setminus \{p\}} d(p, q)$$

and observe that, for $q \in A \setminus \{p\}$, $q \notin B_r(p)$. \square

Exercise 2.40. Fix the writing in the following proof by adding necessary quantifiers.

Proposition: Suppose that $S \subseteq \mathbb{R}$ is a nonempty bounded subset. Then $\sup S \in \bar{S}$.

Proof. If $\sup S \in S$, we're done. Otherwise the no-gap lemma tells us that

$$S \cap (\sup S - r, \sup S)$$

is nonempty, and since

$$(\sup S - r, \sup S) \subseteq B_r(\sup S)$$

we have that $S \cap B_r(\sup S)$ is nonempty. Thus $\sup S$ is a limit point of S and so belongs to \bar{S} . \square

Exercise 2.41. Strongly equivalent metrics. If the set X is equipped with two metrics d and ρ , the metrics d and ρ are *strongly equivalent* if there are two numbers $c_1 > 0$ and $c_2 > 0$ such that, if x and y are any two points in X , then

$$c_1 d(x, y) \leq \rho(x, y) \leq c_2 d(x, y)$$

(recall Remark 2.7.7).

(a) Show that, if d and ρ are strongly equivalent with notation as above, then there are two numbers $b_1 > 0$ and $b_2 > 0$ such that

$$b_1 \rho(x, y) \leq d(x, y) \leq b_2 \rho(x, y).$$

(The point is that while it might seem at first glance that d and ρ do not play symmetric roles in the definition of strong equivalence, they actually do.)

(b) Prove that if d and ρ are strongly equivalent, then they are equivalent. (The converse does *not* hold; see Exercise 5.32.)

(c) Suppose that X is a vector space equipped with two norms ν_d and ν_ρ , and write d and ρ for the metrics induced by ν_d and ν_ρ , respectively. Prove that, if d and ρ are equivalent, they are in fact strongly equivalent.

(d) Show that the Euclidean metric, sup metric, and ℓ_1 metric on \mathbb{R}^n are all strongly equivalent.

Exercise 2.42. Let d be any of the Euclidean, ℓ_1 , or sup metrics on \mathbb{R}^n , and let ρ be any of the Euclidean, ℓ_1 , or sup metrics on \mathbb{R}^n . Let $p \in \mathbb{R}^n$, and let $r > 0$ be given. Find (for each of the combinations of possibilities for which metrics d and ρ are) a number $R > 0$ such that

$$B_r^d(p) \subseteq B_R^\rho(p).$$

[Hint: The mere existence of such an R can be deduced from part (d) of Exercise 2.41. It is possible, though, to derive much more detailed information on exactly how large we need R to be (in terms of r) to make the above inclusion true; the answer depends on exactly which metrics d and ρ are, as well as on the dimension n . You may well discover much of this information in the course of proving Proposition 2.7.3 in Exercise 2.31.]

Exercise 2.43. The p -adic metric. This problem introduces a very interesting metric, called the p -adic metric, on the set of rational numbers \mathbb{Q} . This metric is a useful tool in number theory. For more information see, for example, the book [10].

Let the prime number p be given and fixed. Given any nonzero integer a , the p -adic valuation of a is denoted $v_p(a)$ and is defined to be the largest nonnegative integer n such that p^n divides a . (So, for example, if $p = 3$, $v_p(36) = 2$ since $36 = 2 \times 2 \times 3 \times 3$; $v_p(-15) = 1$, since $-15 = 3 \times (-5)$; and $v_p(7) = 0$, since 3 does not divide 7.)

We extend this function to the rationals as follows. Given any nonzero $x = a/b \in \mathbb{Q}$, we define the p -adic valuation of x , $v_p(x)$, by the formula $v_p(x) = v_p(a) - v_p(b)$.

- (a) Show that $v_p(x)$ does not depend on “how x is written” — that is, if $x = a/b = c/d$, then $v_p(a) - v_p(b) = v_p(c) - v_p(d)$.
- (b) Show that $v_p(x + y) \geq \min(v_p(x), v_p(y))$.

Given any nonzero $x \in \mathbb{Q}$, we now define the p -adic absolute value of x , $|x|_p$, by the formula

$$|x|_p = p^{-v_p(x)}.$$

We also set $|0|_p = 0$.

- (c) Show that the formula $d_p(x, y) = |x - y|_p$ defines a metric (called the p -adic metric) on \mathbb{Q} .
- (d) For $p = 3$, find $d_p(3, 5)$ and $d_p(\frac{11}{4}, \frac{1}{2})$.
- (e) The rational numbers are certainly a vector space over the field \mathbb{Q} (that is, if the scalars are allowed to be elements of \mathbb{Q}). In the sense of Section 2.8, the function $x \mapsto |x|_p$ is *not quite a norm* on \mathbb{Q} (at least if we view rational “scalars” as having the standard absolute value), since it is *not* in general true that, given $x, y \in \mathbb{Q}$, $|xy|_p = |x||y|_p$. Can you replace this false equality with a true one, and prove?