A Course in Metric Spaces and Point-Set Topology

Nicolas Addington

Last updated October 27, 2025

Contents

1 Continuity, Convergence, and Metric Spaces

In a previous course you may have seen the definition of continuity and convergence in $\mathbb{R}^n$:

Definition 1.1.

(a) A map $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is *continuous* at a point $\mathbf{a} \in \mathbb{R}^m$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that for all $\mathbf{x} \in \mathbb{R}^m$ with $|\mathbf{x} - \mathbf{a}| < \delta$ we have $|f(\mathbf{x}) - f(\mathbf{a})| < \epsilon$.

(b) A sequence of points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots \in \mathbb{R}^n$ *converges* to a limit $\ell \in \mathbb{R}^n$ if for every $\epsilon > 0$ there is a natural number N such that for all $n \geq N$ we have $|\mathbf{x}_n - \ell| < \epsilon$.

You don't have to be completely at ease with these definitions, but if you haven't at least worked with them in one variable – that is, with functions $f\colon \mathbb{R} \to \mathbb{R}$ and sequences of numbers $x_1, x_2, x_3, \ldots \in \mathbb{R}$ – then you should take a rigorous advanced calculus course, called introductory real analysis in some places, before this one. My favorite book is Spivak's *Calculus* [5]. Oregon's Math 316–317 currently uses Abbott's *Understanding Analysis* [1].

In the definition above, $|\mathbf{x} - \mathbf{a}|$ means the length of the vector $\mathbf{x} - \mathbf{a}$, which we should understand as measuring the distance between $\mathbf{x}$ and $\mathbf{a}$ in $\mathbb{R}^m$, and similarly with $|f(\mathbf{x}) - f(\mathbf{a})|$ and $|\mathbf{x}_n - \ell|$. This course begins from the observation that when we study continuous maps and convergent sequences, we can forget almost everything we know about vectors and their geometry, and retain only the notion of distance. We introduce the following definition.

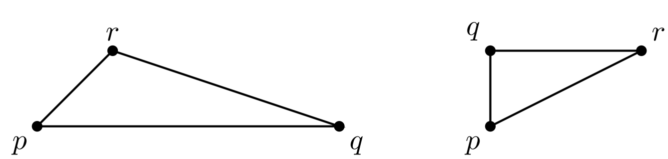
Definition 1.2. A *metric* on a set X is a function $d\colon X \times X \to \mathbb{R}_{\geq 0}$ such that

(a) $d(p,q) = d(q,p)$ for all $p, q \in X$,

(b) $d(p,q) = 0$ if and only if $p = q$, and

(c) $d(p,q) \leq d(p,r) + d(r,q)$ for all $p, q, r \in X$.

The last property is called the *triangle inequality*:



Now we can translate our definition of continuity and convergence in this abstract setting:

Definition 1.3. Let X and Y be sets equipped with metrics d_X and d_Y.

(a) A map $f\colon X \to Y$ is *continuous* at a point $p \in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that for all $q \in X$ with $d_X(p,q) < \delta$ we have $d_Y(f(p), f(q)) < \epsilon$.

(b) A sequence of points $p_1, p_2, p_3, \ldots \in X$ *converges* to a limit $\ell \in X$ if for every $\epsilon > 0$ there is a natural number N such that for all $n \geq N$ we have $d_X(p_n, \ell) < \epsilon$.

We will often say "let (X, d) be a metric space," which means that X is a set and d is a metric on it. A set will typically admit many metrics, or to put it another way, we can have many metric spaces with the same underlying set. Here are some of the examples that we will be interested in.

Example 1.4. *Three metrics on $\mathbb{R}^n$.*

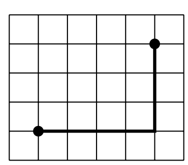
(a) The Euclidean metric: given two points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, we define

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}.$$

(b) The taxicab metric

$$d_1(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n|.$$

The name alludes to the driving distance between two points in a city whose roads are laid out on a grid, like Manhattan or Eugene.



(c) The square metric on $\mathbb{R}^n$:

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - y_1|,\ |x_2 - y_2|,\ \ldots,\ |x_n - y_n|\}.$$

The reason for the name will become clear in Exercise 1.1.

On $\mathbb{R}^1$ these three metrics are all the same, but when $n \geq 2$ they are all different: for example, they give different distances between $(0, 0, \ldots, 0)$ and $(1, 1, \ldots 1)$. We will eventually see, however, that they all have the same continuous maps and convergent sequences.

Example 1.5. *Interpolating between those three metrics, and a non-example.*

For $p > 0$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we can define

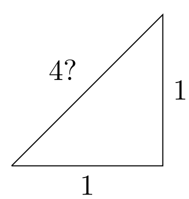
$$d_p(\mathbf{x}, \mathbf{y}) = \left(|x_1 - y_1|^p + |x_2 - y_2|^p + \cdots + |x_n - y_n|^p\right)^{1/p}.$$

We see that $p = 1$ gives the taxicab metric, $p = 2$ gives the Euclidean metric, and it is interesting to think about how letting $p \to \infty$ gives the square metric. We see that d_p satisfies the first two properties of Definition 1.2, and it turns out to satisfy the last property (the triangle inequality) if $p \geq 1$, although this is not obvious.

On the other hand, if $0 < p < 1$ then the triangle inequality fails: for example, with $p = 1/2$ we have

$$\begin{aligned}
d_{1/2}((0,0),(0,1)) &= \left(|0-0|^{1/2} + |0-1|^{1/2}\right)^2 = (0+1)^2 = 1,\\
d_{1/2}((0,1),(1,1)) &= \left(|0-1|^{1/2} + |1-1|^{1/2}\right)^2 = (0+1)^2 = 1,\\
d_{1/2}((0,0),(1,1)) &= \left(|0-1|^{1/2} + |0-1|^{1/2}\right)^2 = (1+1)^2 = 4,
\end{aligned}$$

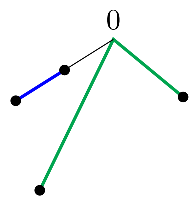
but 4 is bigger than $1 + 1$.



Example 1.6. To illustrate how permissive the definition of a metric is, we introduce the *SNCF metric* on $\mathbb{R}^2$:

$$d_*(p,q) = \begin{cases} d_2(p,q) & \text{if } p \text{ and } q \text{ lie on the same line} \\ & \text{through the origin, or} \\ d_2(p,0) + d_2(0,q) & \text{otherwise,} \end{cases}$$

where d_2 is the Euclidean metric. Here is a picture:

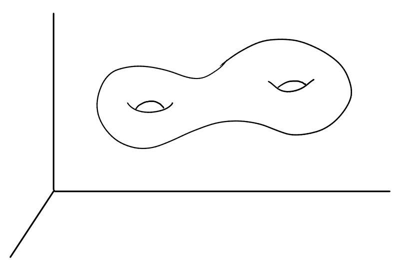


You should convince yourself that this satisfies the triangle inequality.

SNCF stands for *Société nationale des chemins de fer français*, the French national railway company. The joke is that if you want to go from, say, Dijon to Bordeaux, you might as well take the train up to Paris and back down. In the UK, it is called the British Rail metric; in Eugene, the LTD bus metric.

Example 1.7. *The induced metric on a subset.*

If Y is a subset of X, then a metric $d\colon X \times X \to \mathbb{R}$ induces a metric on Y, just by restricting the domain of d to $Y \times Y \subset X \times X$: that is, by declaring that the distance between two points in the subspace Y is the same as it was in X. So for example if $X = \mathbb{R}^3$ and Y is a surface,



then we can get metrics on Y by restricting the Euclidean metric, the taxicab metric, or the square metric from $\mathbb{R}^3$.

This surface carries other interesting metrics as well: for example, we could declare the distance between two points on the surface to be the length of the shortest path between them on the surface, which will be different from the shortest path through the ambient space. That metric belongs to a subject called Riemannian geometry, and is beyond the scope of these notes.

Example 1.8. *Some spaces of functions.*

Let $C([0,1])$ be the set whose elements are continuous, real-valued functions $f\colon [0,1] \to \mathbb{R}$. This set is much too big to visualize like we do $\mathbb{R}^2$ or $\mathbb{R}^3$, but we want to regard its elements as points in a huge space and consider different ways of measuring the distance between them. There are many interesting metrics on $C([0,1])$, for example the L^1 metric

$$d_1(f,g) = \int_0^1 |f(x) - g(x)|\, dx,$$

the L^2 metric

$$d_2(f,g) = \sqrt{\int_0^1 |f(x) - g(x)|^2\, dx},$$

the L^∞ or sup metric

$$d_\infty(f,g) = \sup_{x\in[0,1]} |f(x) - g(x)|,$$

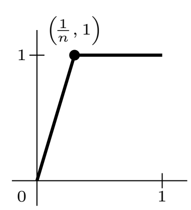
and the L^p metrics for $p \geq 1$, which interpolate between these:

$$d_p(f,g) = \left(\int_0^1 |f(x) - g(x)|^p\, dx\right)^{1/p}.$$

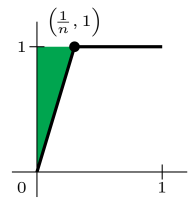
It is interesting to think about how these metrics are analogous to the metrics on $\mathbb{R}^n$ that we named d_1, d_2, d_∞, and d_p a moment ago. But whereas those metrics on $\mathbb{R}^n$ will all turn out to give the same continuous maps and convergent sequences, these ones do not:

1.8(a) *A sequence in $C([0,1])$ that converges in the L^1 metric but not in the sup metric.*

For $n = 1, 2, 3, \ldots$, consider the function $f_n\colon [0,1] \to \mathbb{R}$ that goes piecewise linearly from $f(0) = 0$ to $f(1/n) = 1$ to $f(1) = 1$:



I claim that the sequence $f_1, f_2, f_3, \dots$ converges to the constant function $g(x) = 1$ in the L^1 metric, but not in the sup metric. On the one hand, $d_1(f_n, g) = \int_0^1 |f_n - g|$ is the area of the triangle shown,



so that's $1/2n$, which goes to zero as $n \to \infty$, so $f_n \to g$ in the L^1 metric. On the other hand, $d_\infty(f_n, g) = \sup |f_n - g|$ is equal to 1 for all n, which does not go to zero as $n \to \infty$, so $f_n \not\to g$ in the sup metric.

You can check that the same happens with $f_n(x) = x^n$ and $g(x) = 0$.

You can unpack the definitions to see that a sequence of functions converges in the sup metric if and only if it converges *uniformly*.

1.8(b) *Integration gives a map from $C([0, 1])$ to $\mathbb{R}$ that is continuous in both the sup metric and the L^1 metric.*

Consider the map $\Phi\colon C([0, 1]) \to \mathbb{R}$ given by $\Phi(f) = \int_0^1 f(x)\, dx$.

First I claim that Φ is continuous if $C([0, 1])$ is given the sup metric (and the target space $\mathbb{R}$ is given the usual Euclidean metric). Unpacking the definitions, this means that for every $f \in C([0, 1])$ and every $\epsilon > 0$, there is a $\delta > 0$ such that for all $g \in C([0, 1])$ with $d_\infty(f, g) < \delta$ we have $|\Phi(f) - \Phi(g)| < \epsilon$. So let $f \in C([0, 1])$ and $\epsilon > 0$ be given, and take $\delta = \epsilon$. If $d_\infty(f, g) < \delta$, then for all $x \in [0, 1]$ we have $|f(x) - g(x)| < \delta$, so

$$|\Phi(f) - \Phi(g)| = \left| \int_0^1 f(x)\, dx - \int_0^1 g(x)\, dx \right|$$
$$= \left| \int_0^1 (f(x) - g(x))\, dx \right| \le \int_0^1 |f(x) - g(x)|\, dx < \int_0^1 \delta\, dx = \delta = \epsilon,$$

which is what we wanted.

Next I claim that Φ is also continuous in the L^1 metric. Let $f \in C([0, 1])$ and $\epsilon > 0$ be given, and take $\delta = \epsilon$. If $d_1(f, g) < \delta$, then

$$|\Phi(f) - \Phi(g)| \le \int_0^1 |f(x) - g(x)|\, dx = d_1(f, g) < \delta = \epsilon,$$

which is what we wanted.

1.8(c) *Evaluation at $x = 0$ gives a map from $C([0,1])$ to $\mathbb{R}$ that is continuous in the sup metric but not in the L^1 metric.*

Consider the map $\Psi\colon C([0,1]) \to \mathbb{R}$ given by $\Psi(f) = f(0)$.

On the one hand, I claim that Ψ is continuous in the sup metric. Let $f \in C([0,1])$ and $\epsilon > 0$ be given, and take $\delta = \epsilon$. If $d_\infty(f,g) < \delta$, then for all $x \in [0,1]$ we have $|f(x) - g(x)| < \delta$, so

$$|\Psi(f) - \Psi(g)| = |f(0) - g(0)| < \delta = \epsilon,$$

as desired.

On the other hand, I claim that Ψ is *not* continuous in the L^1 metric. Take the sequence $f_1, f_2, f_3, \ldots$ from part (a) above, which converges in the L^1 metric to the constant function g. If Ψ were continuous in the L^1 metric, then the sequence $\Psi(f_1), \Psi(f_2), \Psi(f_3), \ldots$ in $\mathbb{R}$ would converge to $\Psi(g)$ by Exercise 1.11 below, but in fact we have $\Psi(f_n) = 0$ for all n while $\Psi(g) = 1$.

Our definition of the sup metric implicitly uses the fact that a continuous function on $[0,1]$ is bounded: otherwise $\sup|f - g|$ might be infinite. You have probably seen this proved in a first course in real analysis; we will assume it for now, and prove it properly when we come to compactness.

Example 1.9. There are many other function spaces. For example, we can take the set $C^1([0,1])$ of functions $f\colon [0,1] \to \mathbb{R}$ that are continuously differentiable, meaning that the derivative f' exists and is continuous, and give it the metric

$$d(f,g) = \sup|f - g| + \sup|f' - g'|,$$

called the C^1 *metric*. Or we can give it one of the *Sobolev metrics*

$$d(f,g) = \Big((d_p(f,g))^p + (d_p(f',g'))^p\Big)^{1/p},$$

where d_p is the L^p metric from Example 1.8 and $p \geq 1$. More generally, we can take the set $C^k([0,1])$ of functions whose first k derivatives exist and are continuous, and give it a similarly-defined C^k metric, Sobolev metrics, and many others. We can generalize further to functions of several variables, and on and on. These function spaces are useful in studying solutions to partial differential equations.

We conclude this section with a key fact relating continuous maps and convergent sequences, leaving part of the proof as an exercise.

Proposition 1.10. *Let (X, d_X) and (Y, d_Y) be metric spaces. Then a map $f\colon X \to Y$ is continuous at a point $\ell \in X$ if and only if for every sequence $p_1, p_2, p_3, \ldots$ in X that converges to ℓ, the sequence $f(p_1), f(p_2), f(p_3), \ldots$ in Y converges to $f(\ell)$.*

Proof. One direction – if f is continuous at ℓ then it takes any sequence converging to ℓ to a sequence converging to $f(\ell)$ – is Exercise 1.11 below. For the other direction, let us argue that if f is *not* continuous at ℓ, then there is a sequence $p_1, p_2, \ldots \in X$ that converges to ℓ, but $f(p_1), f(p_2), \ldots \in Y$ does not converge to $f(\ell)$.

To say that f is not continous at ℓ means that there is an $\epsilon > 0$ such that for every $\delta > 0$, there is a point $p \in X$ with $d_X(p, \ell) < \delta$ but $d_Y(f(p), f(\ell)) \geq \epsilon$. So for each $n = 1, 2, 3, \ldots$, we can take $\delta = 1/n$ and get a point p_n such that $d_X(p_n, \ell) < 1/n$ but $d(f(p_n), f(\ell)) \geq \epsilon$. Then the sequence $p_1, p_2, p_3, \ldots$ converges to ℓ, but $f(p_1), f(p_2), f(p_3), \ldots$ does not converge to $f(\ell)$. (You may want to write out the details of these last two claims as well.) $\square$

Exercises.

1.1. (a) For each of the three metrics in Example 1.4, sketch the open ball of some radius $r > 0$ around the origin in $\mathbb{R}^2$:

$$B_r(0) = \{(x, y) \in \mathbb{R}^2 : d((x, y),\, 0) < r\}.$$

(b) For one of the three metrics (your choice), prove or give a counter-example to the following statement: a sequence of points (x_1, y_1), (x_2, y_2), (x_3, y_3), $\ldots \in \mathbb{R}^2$ converges to a limit (x, y) if and only if $x_n \to x$ and $y_n \to y$ separately, as sequences in $\mathbb{R}$ with the usual metric.

(c) Why is

$$d(\mathbf{x}, \mathbf{y}) = \min\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|\}$$

not a metric on $\mathbb{R}^n$ if $n \geq 2$?

1.2. The SNCF metric on $\mathbb{R}^2$ was introduced in Example 1.6.

(a) Give an example of a sequence that converges in the Euclidean metric but not in the SNCF metric.

(b) Prove that every sequence that converges in the SNCF metric converges in the Euclidean metric.

1.3. Consider the following silly metric on $\mathbb{R}^2$:

$$d((x_1, y_1),\, (x_2, y_2)) = \begin{cases} |y_1 - y_2| & \text{if } x_1 = x_2 \\ |y_1 - y_2| + 1 & \text{if } x_1 \neq x_2. \end{cases}$$

(a) Prove that d is a metric, that is, it has the three properties listed in Definition 1.2.

(b) Sketch the open balls of radius $1/2$, 1, and 2 around the origin in this metric.

(c) Give an example of a sequence that converges in the Euclidean metric d_2 but not in our silly metric d.

(d) Prove that every sequence that converges in d also converges d_2.

1.4. Let X be any set, and let d_X be the *discrete metric*

$$d_X(p, q) = \begin{cases} 0 & \text{if } p = q, \text{ or} \\ 1 & \text{if } p \neq q. \end{cases}$$

(a) Prove that d_X is a metric.

(b) Let (Y, d_Y) be another metric space (not necessarily discrete). Prove that every map $f\colon X \to Y$ is continuous.

(c) Prove that a sequence $p_1, p_2, p_3, \ldots \in X$ converges in the discrete metric if and only if it is eventually constant.

1.5. In Example 1.8(a) we saw a sequence in $C([0, 1])$ that converges in the L^1 metric but not in the sup metric. Prove that the reverse cannot happen: every sequence that converges in the sup metric converges in the L^1 metric.

1.6. Let (X, d) be a metric space. Prove the *reverse triangle inequality*:

$$|d(p, q) - d(p, r)| \leq d(q, r)$$

for all $p, q, r \in X$. Include an appropriate picture.

1.7. Let (X, d_X) and (Y, d_Y) and (Z, d_Z) be metric spaces. Let $f\colon X \to Y$ be continuous at a point $p \in X$, and let $g\colon Y \to Z$ be continuous at $f(p)$. Prove that $g \circ f$ is continuous at p.

1.8. Let (X, d) be a metric space, and fix a point $a \in X$. Prove that the map $f\colon X \to \mathbb{R}$ given by $f(p) = d(p, a)$ is continuous.

Hint: Use the reverse triangle inequality from Exercise 1.6.

1.9. Let (X, d) be a metric space, and let $p_1, p_2, p_3, \ldots \in X$ be a sequence that converges to a point $\ell \in X$.

(a) Prove that if $p_n \to \ell'$ for another point $\ell' \in X$, then $\ell' = \ell$.
Hint: Prove that $d(\ell', \ell) < \epsilon$ for every $\epsilon > 0$, using the triangle inequality. Then argue that this implies $\ell' = \ell$.

(b) Let $q_1, q_2, q_3, \ldots$ be another sequence. Prove that if $d(p_n, q_n) \to 0$ as $n \to \infty$, then $q_n \to \ell$.

1.11. (One direction of Proposition 1.10.) Let (X, d_X) and (Y, d_Y) be metric spaces, let $p_1, p_2, p_3, \ldots$ be a sequence that converges to a point ℓ in X, and let $f\colon X \to Y$ be continuous at ℓ. Prove that the sequence $f(p_1), f(p_2), f(p_3), \ldots$ converges to $f(\ell)$ in Y.

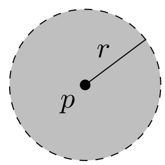
1.12. Let
$$W = \{1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \ldots, 0\}$$
with the metric induced from the usual one on $\mathbb{R}$. Let (X, d_X) be another metric space. Given a sequence $p_1, p_2, p_3 \ldots \in X$ and a point $\ell \in X$, prove that $p_n \to \ell$ if and only if the map $f\colon W \to X$ defined by
$$\begin{cases} f(\frac{1}{n}) = p_n, \\ f(0) = \ell \end{cases}$$
is continuous.

2 Interior, Closure, and Boundary

Definition 2.1. Let (X, d) be a metric space. The *open ball* of radius r around a point $p \in X$ is

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

In $\mathbb{R}^2$ with the Euclidean metric, this looks like



and Exercises 1.1(a) and 1.3(b) asked you to sketch open balls in some other metrics.

Definition 2.2. Let (X, d) be a metric space and $A \subset X$ be a subset.

(a) The *interior* of A, denoted $\operatorname{int} A$ or sometimes A°, is

$$\{p \in X : B_r(p) \subset A \text{ for some } r > 0\}.$$

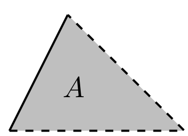
(b) The *closure* of A, denoted $\bar{A}$ or sometimes $\operatorname{cl} A$, is

$$\{p \in X : B_r(p) \cap A \neq \varnothing \text{ for all } r > 0\}.$$

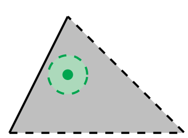
(c) The *boundary* of A, denoted ∂A or sometimes $\operatorname{bd} A$, is $\bar{A} \setminus \operatorname{int} A$.

We see that $\operatorname{int} A \subset A \subset \bar{A}$.

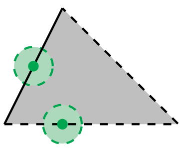
Example 2.3. Let $X = \mathbb{R}^2$ with the Euclidean metric, and let A be the triangular region shown below, which is filled in and contains one edge but not the other two:



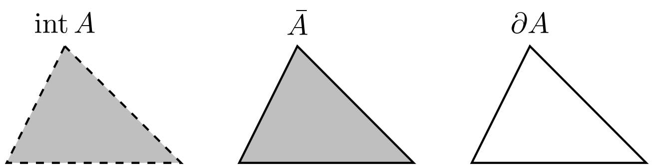
The point shown below is in the interior of A, because we can find a little ball around it that stays within A:



The two points shown below are in the closure of A, and in fact in the boundary of A, because every ball around them meets A, but no ball around them, however small, stays within A:

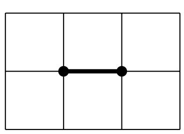


Thus the interior, closure, and boundary of A are as shown:



Example 2.4. The interior, closure, and boundary depend not only on the geometry of A itself, but on how it sits in the ambient space X. For example, if we take $A = [0, 1]$ inside $\mathbb{R}$ with the usual metric,

then the interior is the open interval $(0, 1)$, the closure is the same as A, and the boundary is the two points 0 and 1. But if we take it in $\mathbb{R}^2$ with the Euclidean metric,



then the interior is empty, the closure is the same as A, and the boundary is also the same as A.

We will study this issue more closely in §8, especially Exercise 8.1(e)–(g).

Example 2.5. Let $X = \mathbb{R}$ with the usual metric, and let $A = \mathbb{Q}$. An open ball in this case is just an open interval $(x - r, x + r)$, and you have probably seen in a first course in real analysis that every open interval contains both rational and irrational numbers. Thus the interior of A is empty, the closure is all of $\mathbb{R}$, and the boundary is all of $\mathbb{R}$.

Proposition 2.6. *The complement of the interior is the closure of the complement, and vice versa: that is, if (X, d) is a metric space and $A \subset X$, then*

$$X \setminus \operatorname{int} A = \overline{X \setminus A} \quad \textit{and} \quad X \setminus \bar{A} = \operatorname{int}(X \setminus A).$$

Proof. The two claims are equivalent: replacing A with $X \setminus A$ and taking complements turns one into the other. To prove the first equality, observe that each of the following statements is equivalent to the next:

- $p \in X \setminus \operatorname{int} A$.
- $p \notin \operatorname{int} A$.
- $B_r(p) \not\subset A$ for all $r > 0$.
- $B_r(p) \cap (X \setminus A) \neq \varnothing$ for all $r > 0$.
- $p \in \overline{X \setminus A}$. $\square$

Exercise 2.2 asks you to prove that that

$$\partial A = \partial(X \setminus A) = \bar{A} \cap \overline{X \setminus A}.$$

Here are some basic properties of interiors. Exercise 2.3 asks you to prove the analogous properties of closures, and Exercise 2.5 explores analogues with unions and intersections of infinitely many subsets.

Proposition 2.7. *Let (X, d) be a metric space, and let $A, B \subset X$.*

(a) If $A \subset B$ then $\operatorname{int} A \subset \operatorname{int} B$.

(b) $\operatorname{int}(A \cap B) = \operatorname{int} A \cap \operatorname{int} B$.

(c) $\operatorname{int} A \cup \operatorname{int} B \subset \operatorname{int}(A \cup B)$.
(But Example 2.8 below shows that the inclusion can be strict.)

(d) $\operatorname{int}(\operatorname{int} A) = \operatorname{int}(A)$.

Proof.

(a) If $p \in \operatorname{int} A$ then there is an $r > 0$ such that $B_r(p) \subset A$, so $B_r(p) \subset B$, so $p \in \operatorname{int} B$.

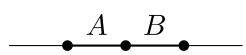
(b) We have $A \cap B \subset A$, so $\operatorname{int}(A \cap B) \subset \operatorname{int} A$ by part (a), and similarly $\operatorname{int}(A \cap B) \subset \operatorname{int} B$; thus $\operatorname{int}(A \cap B) \subset \operatorname{int} A \cap \operatorname{int} B$.

For the reverse inclusion, suppose that $p \in \operatorname{int} A$ and $p \in \operatorname{int} B$. Then there is an $r > 0$ such that $B_r(p) \subset A$, and an $s > 0$ such that $B_s(p) \subset B$. Let $t = \min\{r, s\}$, which is still positive; then $B_t(p) \subset B_r(p) \subset A$ and $B_t(p) \subset B_s(p) \subset B$, so $B_t(p) \subset A \cap B$, so $p \in \operatorname{int}(A \cap B)$ as desired.

(c) We have $A \subset A \cup B$, so $\operatorname{int} A \subset \operatorname{int}(A \cup B)$ by part (a), and similarly $\operatorname{int} B \subset \operatorname{int}(A \cup B)$; thus $\operatorname{int} A \cup \operatorname{int} B \subset \operatorname{int}(A \cup B)$. □

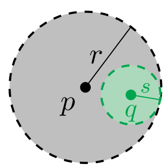
(d) We have $\operatorname{int} A \subset A$, so $\operatorname{int}(\operatorname{int} A) \subset \operatorname{int} A$. For the reverse inclusion, let $p \in \operatorname{int}(A)$, so there is an $r > 0$ such that $B_r(p) \subset A$. Taking interiors, we get $\operatorname{int} B_r(p) \subset \operatorname{int} A$. Now Proposition 2.9 below gives $B_r(p) \subset \operatorname{int} A$, so $p \in \operatorname{int}(\operatorname{int}(A))$.

Example 2.8. It need not be true that $\operatorname{int} A \cup \operatorname{int} B = \operatorname{int}(A \cup B)$. For example, let $X = \mathbb{R}$ with the usual metric, let $A = [0, 1]$, and let $B = [1, 2]$, so $A \cup B = [0, 2]$. Then $\operatorname{int}(A \cup B) = (0, 2)$ is bigger than $\operatorname{int} A \cup \operatorname{int} B = (0, 1) \cup (1, 2)$.



Proposition 2.9. *Let (X, d) be a metric space. For every $p \in X$ and $r > 0$, the open ball $B_r(p)$ equals its own interior.*

Proof. We want to prove that for every $q \in B_r(p)$ there is an $s > 0$ such that $B_s(q) \subset B_r(p)$.



Let q be given, and let $s = r - d(p, q)$, which is positive because $d(p, q) < r$. If $q' \in B_s(q)$ then

$$d(p, q') \leq d(p, q) + d(q, q') < d(p, q) + s = r,$$

so $q' \in B_r(p)$. Thus $B_s(q) \subset B_r(p)$, as required. □

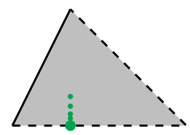
We conclude this section with another characterization of the closure:

Proposition 2.10. *Let (X, d) be a metric space, and let $A \subset X$. Then $\bar{A}$ is the set of points $p \in X$ for which there is a sequence $a_1, a_2, a_3, \ldots \in A$ that converges to p.*

Proof. Suppose that $p \in \bar{A}$. Then for each $n \in \{1, 2, 3, \ldots\}$, the intersection $B_{1/n}(p) \cap A$ is not empty, so we can choose a point a_n in the intersection. We see that $a_1, a_2, a_3, \ldots$ is a sequence in A, and that $d(a_n, p) < 1/n$ which goes to zero as $n \to \infty$, so the sequence converges to p.

Conversely, suppose there is a sequence $a_1, a_2, a_3, \ldots \in A$ that converges to p. Then for every $r > 0$ there is an N such that for all $n \geq N$ we have $d(a_n, p) < r$. In particular $B_r(p) \cap A$ contains a_N, so it is not empty. □

Here is an illustration of Proposition 2.10 in the context of Example 2.3:



Exercises.

2.1. Sketch each subset of $\mathbb{R}^2$ and find its closure, interior, and boundary in the Euclidean metric:

(a) $A_1 = \{(x, y) : 0 < x \leq 1, 0 \leq y < 1\}$
(b) $A_2 = \{(x, y) : 0 < x \leq 1, y = 0\}$
(c) $A_3 = \{(x, y) : x \in \mathbb{Q} \text{ or } y \in \mathbb{Q}\}$
(d) $A_4 = \{(x, y) \in \mathbb{R}^2 : x \neq 0 \text{ or } y = 0\}$

2.2. Let (X, d) be a metric space and $A \subset X$. Use Proposition 2.6 to prove that

$$\partial A = \partial(X \setminus A) = \bar{A} \cap \overline{X \setminus A}.$$

2.3. Prove the analogue of Proposition 2.7 for closures, without appealing to Proposition 2.6.

(a) If $A \subset B$ then $\bar{A} \subset \bar{B}$.
(b) $\bar{A} \cup \bar{B} = \overline{A \cup B}$.
(c) $\overline{A \cap B} \subset \bar{A} \cap \bar{B}$.
Give an example to show that the inclusion can be strict.
(d) $\bar{\bar{A}} = \bar{A}$.

2.4. Define the closed ball

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

(a) Prove that $\bar{B}_r(p)$ equals its own closure.
(b) Prove that $\overline{B_r(p)} \subset \bar{B}_r(p)$: that is, the closure of the open ball is contained in the closed ball. But give an example to show that the inclusion can be strict.
Hint: For the proof, you may quote Exercise 2.3(a). For the example, you might take $X = \mathbb{Z}$ with the usual metric inherited from $\mathbb{R}$, or any set with a discrete metric (Exercise 1.4).

2.5. Let I be a set, and suppose that for each $i \in I$ we have a subset $A_i \subset X$. Do not assume that I is countable.

(a) Prove that $\bigcup_{i \in I} \operatorname{int} A_i \subset \operatorname{int}(\bigcup_{i \in I} A_i)$.

(b) Prove that $\operatorname{int}(\bigcap_{i \in I} A_i) \subset \bigcap_{i \in I} \operatorname{int} A_i$.
Give an example to show that the inclusion can be strict.

(c) Prove that $\overline{\bigcap_{i \in I} A_i} \subset \bigcap_{i \in I} \bar{A}_i$.

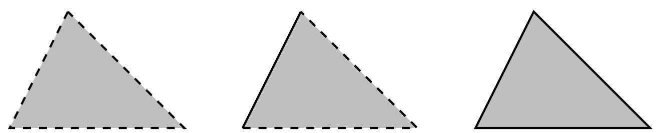
(d) Prove that $\bigcup_{i \in I} \bar{A}_i \subset \overline{\bigcup_{i \in I} A_i}$.
Give an example to show that the inclusion can be strict.

2.6. We saw in Example 2.8 that the inclusion $\operatorname{int} A \cup \operatorname{int} B \subset \operatorname{int}(A \cup B)$ can be strict. Prove however that if $\bar{A} \cap \bar{B} = \varnothing$ then $\operatorname{int} A \cup \operatorname{int} B = \operatorname{int}(A \cup B)$.

3 Open and Closed

Let (X, d) be a metric space and $A \subset X$. In the last section we saw that $\operatorname{int} A \subset A \subset \bar{A}$. We say that A is *open* if $A = \operatorname{int} A$, and *closed* if $A = \bar{A}$. Thinking about the boundary $\partial A = \bar{A} \setminus \operatorname{int} A$, we see that A is closed if it contains all of its boundary, and open if it contains none of its boundary.

Example 3.1. Let us revisit Example 2.3 involving triangles in $\mathbb{R}^2$ with the Euclidean metric. Of the three subsets below, the one on the left is open but not closed, the one on the right is closed but not open, and the one in the middle is neither open nor closed, because it contains part of its boundary but not all.

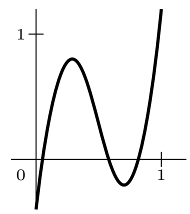


An open set is often denoted by the letter U, from the German *Umgebung*, neighborhood, as in the early texts by Hausdorff [4, VII §1] and Tietze [6]. Looking at the definition of interior, we see that $U \subset X$ is open if and only if for every $p \in U$ there is an $r > 0$ such that $B_r(p) \subset U$. Thus an open subset is one where we can move around a little bit without going outside.

A closed set is often denoted by the letter F, from the French *fermé*, closed. From Proposition 2.10 we see that $F \subset X$ is closed if and only if for every sequence $p_1, p_2, p_3, \ldots \in F$ that converges to a limit $\ell \in X$, we have $\ell \in F$. Thus a closed subset is one where we can take the limit of a sequence without going outside.

Example 3.2. The four subsets of $\mathbb{R}^2$ given in Exercise 2.1 are neither open nor closed in the Euclidean metric. In a typical metric space you should have the feeling that "most" subsets are neither open nor closed.

Example 3.3. Let $X = C^1([0, 1])$, the set of continuously differentiable functions, with the C^1 metric from Example 1.9. The subset consisting of functions with simple roots – that is, those for which $f'(x) \neq 0$ whenver $f(x) = 0$ – turns out to be open. Exercise 3.3 asks you to prove this, but it agrees with our intuition that if we take a function with simple roots and wiggle it a bit, then it still has simple roots:



This is a first taste of what is called *transversality.*

But the set of functions with simple roots is not closed: for example, the functions

$$f_n(x) = \left(x - \tfrac{1}{2}\right)^2 - \tfrac{1}{n}$$

all have simple roots (at $\frac{1}{2} \pm \frac{1}{\sqrt{n}}$), but they converge to $g(x) = (x - \frac{1}{2})^2$ which has a double root at $\frac{1}{2}$.

Example 3.4. Let $X = C([0,1])$ with the sup metric, and fix a subset $S \subset [0,1]$. The subset consisting of functions that vanish on S – that is, those with $f(x) = 0$ for all $x \in S$ – turns out to be closed in the sup metric. Exercise 3.4 asks you to prove this.

But the same set of functions is not closed in the L^1 metric: in Example 1.8(a) we saw a sequence of functions f_n that all vanish at 0, but converge in the L^1 metric to a function g that does not vanish at 0.

Thus we see that the notions of open and closed, interior, closure, and boundary depend on the metric, and may be change if we choose a different metric.

Example 3.5. In a discrete metric (Exercise 1.4), every subset is both open and closed.

Here are some basic properties of closed sets. Exercise 3.5 asks you to prove the analogous properties of open sets.

Proposition 3.6. *If $F, G \subset X$ are closed, then the union $F \cup G$ is again closed.*

Proof. Because F and G are closed we have $F = \bar{F}$ and $G = \bar{G}$. By Exercise 2.3(b) we have $\bar{F} \cup \bar{G} = \overline{F \cup G}$. Thus $F \cup G = \overline{F \cup G}$, so $F \cup G$ is closed. $\square$

By induction, if we have finitely many closed sets $F_1, F_2, \ldots, F_n \subset X$, then their union $F_1 \cup F_2 \cup \cdots \cup F_n$ is again closed. But we could have a countable collection of closed sets whose union is not closed: for example, if we set $F_n = [\frac{1}{n}, 1] \subset \mathbb{R}$ for $n = 1, 2, 3, \ldots$, then each F_n is closed in the usual metric, but

$$F_1 \cup F_2 \cup F_3 \cup \cdots = (0, 1]$$

is not closed.

On the other hand, an arbitrary (even uncountable) intersection of closed sets is closed:

Proposition 3.7. *Let I be a set, and suppose that for each $i \in I$ we have a closed set $F_i \subset X$. Then the intersection $\bigcap_{i \in I} F_i$ is again closed.*

Proof. We must prove that $\bigcap_{i \in I} F_i = \overline{\bigcap_{i \in I} F_i}$. We know that $\bigcap F_i \subset \overline{\bigcap F_i}$. For the reverse inclusion, Exercise 2.5(c) gives $\overline{\bigcap F_i} \subset \bigcap \bar{F}_i$, but each F_i is closed, so $\bar{F}_i = F_i$, so $\overline{\bigcap F_i} \subset \bigcap F_i$. $\square$

Proposition 3.8. *For any subset $A \subset X$:*

(a) The interior $\operatorname{int} A$ *is the largest open subset contained in A, in the sense that any open subset contained in A is contained in* $\operatorname{int} A$.

(b) The closure $\bar{A}$ is the smallest closed subset containing A, in the sense that any closed subset containing A contains $\bar{A}$.

Proof.

(a) Proposition 2.7(d) says that $\operatorname{int} A$ is open. If $U \subset A$ then $\operatorname{int} U \subset \operatorname{int} A$ by Proposition 2.7(a), and if U is open then $\operatorname{int} U = U$.

(b) Exercise 2.3(d) says that $\bar{A}$ is closed. If $A \subset F$ then $\bar{A} \subset \bar{F}$ by Exercise 2.3(a), and if F is closed then $F = \bar{F}$. $\square$

Open and closed sets are dual to one another:

Proposition 3.9. *A subset $A \subset X$ is closed if and only if the complement $X \setminus A$ is open.*

Proof. By definition, A is closed if and only if $A = \bar{A}$, which is true if and only if $X \setminus A = X \setminus \bar{A}$, which is true if and only if $X \setminus A = \operatorname{int}(X \setminus A)$ by Proposition 2.6, which is true if and only if $X \setminus A$ is open by definition. $\square$

Exercises.

3.1. In Example 1.4 we saw three different metrics on $\mathbb{R}^2$. Prove one of the following:

(a) A subset $A \subset \mathbb{R}^2$ is open in the Euclidean metric if and only if it is open in the taxicab metric.

(b) A subset $A \subset \mathbb{R}^2$ is open in the Euclidean metric if and only if it is open in the square metric.

(c) A subset $A \subset \mathbb{R}^2$ is open in the taxicab metric if and only if it is open in the square metric.

3.2. Let $X = \mathbb{Q}$ with the metric induced from the usual one on $\mathbb{R}$: that is, $d(x, y) = |x - y|$ for all $x, y \in \mathbb{Q}$, but we're thinking about $\mathbb{Q}$ in itself and forgetting about the rest of $\mathbb{R}$.

(a) Prove that the subset

$$\{x \in \mathbb{Q} : x^2 < 1\}$$

is open but not closed.

(b) Prove that the subset

$$\{x \in \mathbb{Q} : x^2 < 2\}$$

is both open and closed.
(You may use the fact that $\sqrt{2}$ is irrational without proving it.)

3.3. Let $A \subset C^1([0, 1])$ be the set of functions with simple roots as in Example 3.3. Prove that A is open in the C^1 metric.

Hint: For a given $f \in A$, take the ball of radius

$$r = \inf_{x \in [0,1]} \left(|f(x)| + |f'(x)|\right).$$

3.4. Let $S \subset [0, 1]$, and let $A \subset C([0, 1])$ be the set of continuous functions that vanish on S as in Example 3.4. Prove that A is closed in the sup metric.

Hint: You could prove this directly from the definitions, or you could use Proposition 2.10 and Proposition 1.10 together with Example 1.8(c), which is stated for evaluation at $x = 0$ but which we can see is equally valid for evaluation at any $x \in [0, 1]$.

3.5. Without using Proposition 3.9,

(a) Prove that if $U, V \subset X$ are open, then the intersection $U \cap V$ is again open.

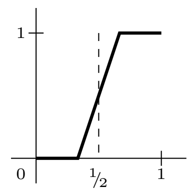
(b) Give an example of a metric space (X, d) and countably many open sets $U_1, U_2, U_3, \ldots \subset X$ such that their intersection $U_1 \cap U_2 \cap U_3 \cap \cdots$ is not open.

(c) Let I be a set, and suppose that for each $i \in I$ we have an open set $U_i \subset X$. Prove that the union $\bigcup_{i \in I} U_i$ is again open.

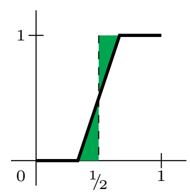
(Don't assume that the index set I is countable!)

4 Completeness

We have seen that a subset $F \subset X$ is closed if it contains all the limits it should – or at least, all the limits that X knows about. Sometimes, however, the metric space X itself has sequences that seem like they ought to converge, but the limit is missing. For example, take $X = C([0,1])$ with the L^1 metric, and for $n = 2, 3, 4, \ldots$, consider the function $f_n \colon [0,1] \to \mathbb{R}$ that goes piecewise linearly from $f_n(0) = 0$ to $f_n(\frac{1}{2} - \frac{1}{n}) = 0$ to $f_n(\frac{1}{2} + \frac{1}{n}) = 1$ to $f_n(1) = 1$:



As $n \to \infty$, we see that $f_n(x)$ looks more and more the step function $g(x)$ that jumps from 0 to 1 at $x = 1/2$. The sequence does not converge to g in the sup metric, but it would in the L^1 metric, if g were in $C([0,1])$: the distance between f_n and g would be the area of the two triangles shown,



which is $\frac{1}{2n} \to 0$. But g is not in $C([0,1])$. We will name this problem by saying that the L^1 metric on $C([0,1])$ is *incomplete*.

For a more familiar example, the sequence

$$1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ 1.41421,\ 1.414213,\ \ldots \in \mathbb{Q}$$

wants to approach a limit, namely $\sqrt{2}$, but the limit is missing, so we have to complete the rational numbers $\mathbb{Q}$ to get the real numbers $\mathbb{R}$.

To formalize this, we introduce the definition of a Cauchy sequence,* which tries to say that the sequence converges without referring to the limit, because the limit might be missing from our metric space.

Definition 4.1. A sequence $p_1, p_2, p_3, \ldots$ in a metric space (X, d) is *Cauchy* if for every $\epsilon > 0$ there is a natural number N such that for all $m, n \geq N$ we have $d(p_m, p_n) < \epsilon$.

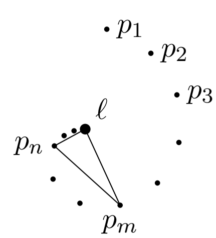
In contrast to Definition 1.3(b) of a convergent sequence, where the tails of the sequence get arbitrarily close to a limit point ℓ, here the tails just get close to themselves.

Proposition 4.2. *Let (X, d) be a metric space. If a sequence $p_1, p_2, p_3, \ldots$ converges to a limit ℓ, then it is Cauchy.*

Proof. Let $\epsilon > 0$ be given. Because the sequence converges to ℓ, there is a natural number N such that $n \geq N$ implies $d(p_n, \ell) < \epsilon/2$. Thus if $m, n \geq N$ then the triangle inequality gives

$$d(p_m, p_n) \leq d(p_m, \ell) + d(\ell, p_n) < \epsilon/2 + \epsilon/2 = \epsilon. \qquad \square$$

This picture attempts to illustrate the preceding proof:



Definition 4.3. A metric space is *complete* if every Cauchy sequence converges.

Proposition 4.4. $\mathbb{R}$ *is complete in the usual metric.*

You have probably seen this proved in a first course in real analysis, but we review the proof, which relies on the fact that a bounded set in $\mathbb{R}$ has a supremum (least upper bound) and an infimum (greatest lower bound). That is, the completeness of $\mathbb{R}$ as a metric space follows from its completeness as an ordered set.

---

*Named for Augustin-Louis Cauchy, 1789–1857, and pronounced ko-shee, not kaw-shee or kow-shee. It's difficult for English speakers to avoid stressing one syllable or the other, so American speakers tend say ko-SHEE, while British speakers tend to say KO-shee. The same thing happens with ballet, garage, and many other words borrowed from French.

Proof of Proposition 4.4. Let $x_1, x_2, x_3, \ldots \in \mathbb{R}$ be a Cauchy sequence.

Exercise 4.3 asks you to prove that a Cauchy sequence is bounded: thus there is an $M \in \mathbb{R}$ such that $|x_n| \leq M$ for all n.

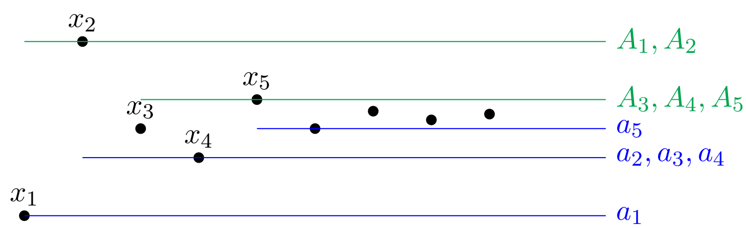
Take the limit superior and limit inferior

$$\ell = \liminf_{n\to\infty} x_n \qquad \text{and} \qquad L = \limsup_{n\to\infty} x_n,$$

which are defined as follows. For $n = 1, 2, 3, \ldots$, let

$$a_n = \inf\{x_n, x_{n+1}, x_{n+2} \ldots\} \qquad \text{and} \qquad A_n = \sup\{x_n, x_{n+1}, x_{n+2}, \ldots\},$$

which exist because our sequence is bounded. This picture attempts to illustrate how the a_ns and A_ns are related to the x_ns in a particular example:



We see that

$$-M \leq a_n \leq x_n \leq A_n \leq M. \tag{4.1}$$

We define

$$\ell = \sup\{a_1, a_2, a_3 \ldots\} \qquad \text{and} \qquad L = \sup\{A_1, A_2, A_3, \ldots\}.$$

The sequence $a_1, a_2, a_3, \ldots$ is non-decreasing, because each term is the infimum of a smaller set than the one before, and a bounded non-decreasing sequence converges to its supremum, so $a_n \to \ell$ as $n \to \infty$. Similarly we have $A_n \to L$ as $n \to \infty$.

We want to prove that the sequence $x_1, x_2, x_3, \ldots\cdot$ converges. By inequalities (4.1) and the squeeze theorem, it is enough to prove that $\ell = L$, or equivalently, that $|A_n - a_n| \to 0$ as $n \to \infty$. Let $\epsilon > 0$ be given. Because $x_1, x_2, \ldots$ is a Cauchy sequence, there is an N such that for all $m, n \geq N$ we have $|x_m - x_n| < \epsilon/3$. Because a_N was defined as an infimum, there is some $m \geq N$ such that $x_m < a_N + \epsilon/3$; otherwise $a_N + \epsilon/3$ would have been a greater lower bound for the set $\{x_N, x_{N+1}, \ldots\}$. Similary, there is an $n \geq N$ such that $x_n > A_N - \epsilon/3$. Now by the triangle inequality we have

$$|A_N - a_N| \leq |A_N - x_n| + |x_n - x_m| + |x_m - a_N| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

Because the As are non-increasing and the as are non-decreasing, for all $\nu \geq N$ we have $|A_\nu - a_\nu| \leq |A_N - a_N| < \epsilon$, which is what we wanted. $\square$

Example 4.5. On the other hand, $\mathbb{R}$ is *not* complete in the metric

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

because the sequence $1, 2, 3, \ldots$ is Cauchy but does not converge. To see that it is Cauchy, note that the sequence $\arctan 1$, $\arctan 2$, $\arctan 3, \ldots$ converges to $\pi/2$ in the usual metric, so it is Cauchy in the usual metric, so for every $\epsilon > 0$ there is an N such that if $m, n \geq N$ then $|\arctan m - \arctan n| < \epsilon$. To see that it does not converge, note that the distance from $1, 2, 3, \ldots$ to any $\ell \in \mathbb{R}$ approaches $\pi/2 - \arctan(\ell)$, and in particular does not go to zero.

If we wanted to make $\mathbb{R}$ complete in this metric, we would need to add two more points at $\pm\infty$. We will return to the idea of completion at the end of the section.

One can show that $\mathbb{R}$ with this metric has the same continuous functions, convergent sequences, and open and closed sets as it does in the usual metric. Later we will see that continuity and convergence can be defined purely in terms of open sets, but this example shows that completeness cannot.

Example 4.6. $\mathbb{R}^n$ is complete in any of the metrics from Example 1.4. To see this, let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be a sequence in $\mathbb{R}^n$ that is Cauchy under one of the three metrics. Exercise 1.1(b) asked you to prove that the sequence of points in $\mathbb{R}^2$ converges in one of the three metrics if and only if each of its coordinates converges in the usual metric, and the same proof applies to $\mathbb{R}^n$. So if we let the i^{th} coordinate of $\mathbf{x}_n$ be called $x_{n,i}$, then it is enough to prove that the sequence $x_{1,i}, x_{2,i}, x_{3,i}, \ldots$ is Cauchy in $\mathbb{R}$ in the usual metric. This follows from the fact that

$$|x_{m,i} - x_{n,i}| \leq d(\mathbf{x}_m, \mathbf{x}_n),$$

in any of the three metrics.

Example 4.7. *Poincaré's hyperbolic ball and half-space.*

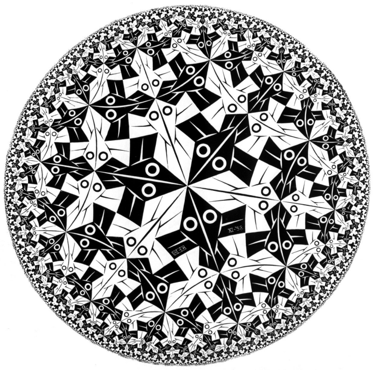
Consider the open unit ball

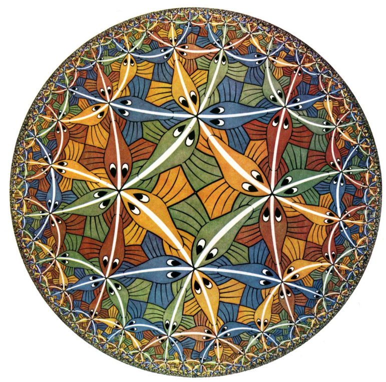
$$B = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < 1\}.$$

The Euclidean metric that B inherits as a subset of $\mathbb{R}^n$ is not complete: a sequence that approaches a point on the boundary sphere will be Cauchy, but will not converge. On the other hand, the *hyperbolic metric*

$$d(\mathbf{x}, \mathbf{y}) = \cosh^{-1}\left(1 + 2\frac{|\mathbf{x} - \mathbf{y}|^2}{(1 - |\mathbf{x}|^2)(1 - |\mathbf{y}|^2)}\right)$$

turns out to be complete. Here are two images from a series of four by M. C. Escher set in the hyperbolic disc; any two fish have the same length in the hyperbolic metric.





As in Example 4.5, the hyperbolic and Euclidean metrics have the same continuous functions, convergent sequences, and open and closed sets, but one is complete while the other is not.

Similarly, the upper half-space

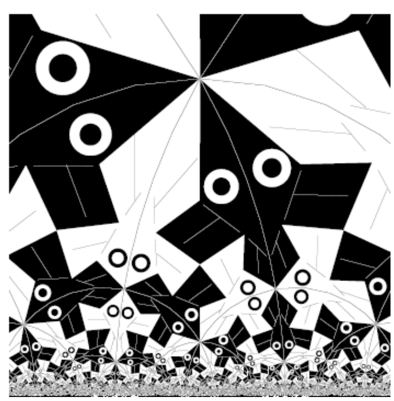
$$H = \{\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_n > 0\}$$

is not complete in the Euclidean metric, but is complete in the hyperbolic metric

$$d(\mathbf{x}, \mathbf{y}) = 2 \sinh^{-1} \frac{|\mathbf{x} - \mathbf{y}|}{2\sqrt{x_n y_n}}.$$

If you've seen how linear fractional transformations (Möbius transformations) with real coefficients act on the upper half of the complex plane, you may be interested to know that they preserve this hyperbolic metric.

There is a bijection between the ball and the half-space that turns one hyperbolic metric into the other. The image of Escher's first print under that isometry, reproduced from [3, Fig. 5], is shown at right.



Hyperbolic spaces are important in differential geometry, complex analysis, and number theory, and historically in demonstrating the logical independence of Euclid's parallel postulate from his other four.

Example 4.8. Given a prime number p, the *p-adic metric* on $\mathbb{Q}$ is defined as follows. For $x, y \in \mathbb{Q}$ with $x \neq y$, we can write

$$|x - y| = p^v \cdot \frac{r}{s},$$

where $v, r, s \in \mathbb{Z}$ and r and s are not divisible by p. Then we define $d_p(x, y) = p^{-v}$, or if $x = y$ then $d_p(x, y) = 0$.* Exercise 4.2 asks you to check that this is a metric and explore some numerical examples.

Convergent sequences in the p-adic metric can look bizarre: for example, if $p = 2$, then the sequence

$$1, 3, 7, 15, \ldots, 2^n - 1, \ldots$$

converges to -1, because the 2-adic distance between $2^n - 1$ and -1 is 2^{-n}, which goes to zero as $n \to \infty$. The metric is not complete: for example, you could prove by hand that the sequence of partial sums of the series

$$1 + p + p^4 + p^9 + \cdots + p^{n^2} + \cdots$$

are Cauchy but fail to converge to a limit in $\mathbb{Q}$, or later, after we have proved the Baire category theorem, we could use it to argue that a complete metric space containing $\mathbb{Q}$ in this metric must be uncountable (Exercise 5.5). The completion of $\mathbb{Q}$ in this metric is called $\mathbb{Q}_p$, the p-adic numbers.

I will digress to discuss where the p-adic metric and its completion come from, but you can skip ahead if you like. They were developed by Hensel, who wanted to do something in number theory analogous to working on a power series "order by order." For example, we can study the equation $f(x)^2 = 1 - x$ by looking for a solution of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots.$$

Looking at the constant terms, we find that a_0 could be 1 or -1. If we choose $a_0 = 1$, then looking at the linear terms we find that a_1 must be $-\frac{1}{2}$, looking at the quadratic terms we find that a_2 must be $-\frac{1}{8}$, and so on. (Later we could ask about the convergence of the power series that we obtain.) Hensel's idea is that if we want to study an equation like $x^2 = 2$, we can look for a solution that's a formal power series in a prime p, say $p = 7$,

$$x = a_0 + a_1 \cdot 7 + a_2 \cdot 7^2 + \cdots$$

*This is unrelated to the metrics called d_p in Examples 1.5 and 1.8.

with coefficients $a_i \in \{0, 1, \ldots, 6\}$. Again we can work order by order. Reducing modulo 7, we find that a_0 could be 3 or 4, because these are the two solutions to $a_0^2 \equiv 2 \pmod 7$. If we choose $a_0 = 3$, then reducing modulo $7^2 = 49$ we find that a_1 must be 1, reducing modulo $7^3 = 343$ we find that a_2 must be 2, and so on. The sequence of partial sums of the series

$$3 + 1 \cdot 7 + 2 \cdot 7^2 + \cdots$$

is $3, 10, 108, \ldots$ which is not Cauchy in the Euclidean metric – indeed, it goes to infinity. But it Cauchy is in the 7-adic metric, and the limit in the completion $\mathbb{Q}_p$ satisfies $x^2 = 2$. Questions of convergence for the power series correspond approximately to questions of whether the solution can be finagled back into $\mathbb{Q}$; this one cannot, because $\sqrt{2}$ is irrational.

An outstanding result in this area is the Hasse–Minkowski theorem, which states that a homogeneous quadratic equation in several variables, say $x^2 - xy + y^2 = 3z^2$, has a solution with coordinates in $\mathbb{Q}$ if and only if it has a solution in $\mathbb{R}$ and solutions in $\mathbb{Q}_p$ for every prime p.

By now we have gotten pretty far from real analysis or topology; the point is that while most of the examples we've considered are more or less closely related to the Euclidean metric on $\mathbb{R}$, the ideas we're developing have applications in radically different mathematical settings.

Example 4.9. The L^1 metric on $C([0, 1])$ is not complete, as we can see using the sequence $f_1, f_2, f_3, \ldots$ introduced at the beginning of the section: Exercise 4.1 asks you to prove that it is Cauchy, but let us sketch an argument that it does not converge to any limit $\ell \in C([0, 1])$. We see that the restriction of f_n to an interval of the from $[0, \frac{1}{2} - \delta]$ converges to the constant function 0, and the restriction to $[\frac{1}{2} + \delta, 1]$ converges to the constant function 1. So if f_n converged to a limit ℓ, then ℓ would take the value 0 on the half-open interval $[0, \frac{1}{2})$, and 1 on $(\frac{1}{2}, 1]$, so it could not be continuous at $x = 1/2$.

As with $\mathbb{Q}$ in $\mathbb{R}$, one can embed $C([0, 1])$ into a larger set of functions and extend the L^1 metric to a complete metric on that larger set, although this is far from staightforward. First, one has to to decide which discontinuous functions to include, which requires developing the Lebesgue integral. Second, for the step function g that we want to be the limit, the value at $x = 1/2$ is not uniquely determined: we could set $g(1/2) = 0$, or $g(1/2) = 1$, or $g(1/2) = 1/2)$, or any other value, and $\int |f_n - g|$ will go to zero in any case; so one ends up working not with actual functions but with equivalence classes of functions. You can learn all this in a course on measure theory.

While $C([0,1])$ is not complete in the L^1 metric, the next three propositions show that it is complete in the sup metric. We continue to assume the fact that every continuous function on $[0,1]$ is bounded.

Proposition 4.10. *Let (X,d) be a complete metric space, and let $Y \subset X$. Then Y is complete in the induced metric if and only if Y is closed in X.*

Proof. First suppose that Y is closed, and let $p_1, p_2, p_3, \ldots$ be a Cauchy sequence in Y. By definition of the induced metric (Example 1.7), the sequence is also Cauchy in X, and because X is complete, it converges to a limit $\ell \in X$. Because Y is closed, we have $\ell \in Y$. Thus Y is complete.

Conversely, suppose that Y is complete, and let $p_1, p_2, p_3, \ldots$ be a sequence in Y that converges to a limit $\ell \in X$. By Proposition 4.2, the sequence is Cauchy, so it also converges to a limit $\ell' \in Y$. By Exercise 1.9, we have $\ell = \ell'$, so $\ell \in Y$. Thus Y is closed. $\square$

Proposition 4.11. *Let X be a set, and let $B(X)$ be the set of bounded functions $f \colon X \to \mathbb{R}$. The sup metric on $B(X)$, defined by*

$$d(f,g) = \sup_{p \in X} |f(p) - g(p)|,$$

is complete.

Proof. Let $f_1, f_2, f_3, \ldots \in B(X)$ be a sequence that is Cauchy in the sup metric. For each $p \in X$ we have $|f_m(p) - f_n(p)| \leq d(f_m, f_n)$, so the sequence $f_1(p), f_2(p), f_3(p), \ldots$ is also Cauchy in the usual metric on $\mathbb{R}$. Because $\mathbb{R}$ is complete, this sequence converges, so we can define a function $\ell \colon X \to \mathbb{R}$ by $\ell(p) = \lim_{n \to \infty} f_n(p)$.

Let us argue that $\ell \in B(X)$, that is, that ℓ is bounded. Again by Exercise 4.3, there is an $M \in \mathbb{R}$ such that $d(f_n, 0) = \sup |f_n| \leq M$ for all n, so $|f_n(p)| \leq M$ for all n and all $p \in X$. Letting $n \to \infty$, we get $|\ell(p)| \leq M$ for all $p \in X$, as desired.

Last we argue that $f_n \to \ell$ in the sup metric. Let $\epsilon > 0$ be given. Because the sequence $f_1, f_2, f_3, \ldots$ is Cauchy, there is an N such that if $m, n \geq N$ then $d(f_m, f_n) < \epsilon/2$, so $|f_m(p) - f_n(p)| < \epsilon/2$ for all $p \in X$. Letting $m \to \infty$, we get $|\ell(p) - f_n(p)| \leq \epsilon/2$ for all $n \geq N$ and all $p \in X$. Thus $d(\ell, f_n) \leq \epsilon/2 < \epsilon$. $\square$

Proposition 4.12. *Continue to let $B(X)$ denote the set of bounded functions on a set X with the sup metric d. For any metric d_X on X, the set of bounded functions on X that are continuous with respect to d_X is a closed subset of $B(X)$.*

Proof. Let $f_1, f_2, f_3, \ldots$ be a sequence of bounded, continuous functions converging in the sup metric to a bounded function ℓ. We want to prove that ℓ is continuous.

Let $p \in X$ and $\epsilon > 0$ be given. Because $f_n \to \ell$ in the sup metric, there is an N such that if $n \geq N$ then $\sup |f_n - \ell| < \epsilon/3$, and in particular $|f_N - \ell| < \epsilon/3$. Because f_N is continuous, there is a $\delta > 0$ such that if $d_X(p,q) < \delta$ then $|f_N(p) - f_N(q)| < \epsilon/3$. Thus if $d_X(p,q) < \delta$ then

$$\begin{aligned} &|\ell(p) - \ell(q)| \\ &\qquad \leq |\ell(p) - f_N(p)| + |f_N(p) - f_N(q)| + |f_N(q) - \ell(q)| \\ &\qquad\qquad < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon, \end{aligned}$$

where the first inequality is the triangle inequality in $\mathbb{R}$. □

Why completeness? You might ask why we care whether a metric space is complete. One major reason is the Banach fixed-point theorem, also called the contraction mapping theorem, which we prove next. It is crucial to proving the existence and uniqueness of solutions to ordinary differential equations, the inverse function theorem, and the implicit function theorem. Another reason is the Baire category theorem, which is the topic of Section

Theorem 4.13 (Banach fixed-point theorem)**.** *Let (X, d) be a complete metric space, let $f\colon X \to X$, and suppose there is a "Lipschitz constant" $r \in [0,1)$ such that for all $p, q \in X$ we have*

$$d(f(p), f(q)) \leq r \cdot d(p,q). \tag{4.2}$$

Then f has a unique fixed point, that is, there is a unique point $p \in X$ such that $f(p) = p$.

We can see that it is really necessary to have $r < 1$: if $X = \mathbb{R}$ with the usual metric, then a translation like $f(x) = x + 1$ satisfies (4.2) with $r = 1$, but does not have a fixed point. Exercise 4.5 asks you to explore how the theorem can fail if the hypotheses are weakened in other ways.

Proof of Theorem 4.13. Uniqueness is easy: if $f(p) = p$ and $f(q) = q$, then then

$$d(p,q) = d(f(p), f(q)) \leq r \cdot d(p,q),$$

so $d(p,q) = 0$, so $p = q$.

For existence, choose any point $p_0 \in X$, and define a sequence of points by repeatedly applying f:

$$p_1 = f(p_0) \qquad p_2 = f(p_1) \qquad p_3 = f(p_2) \qquad \ldots .$$

Let us use the geometric series to argue that this sequence is Cauchy.

Set

$$D = d(p_0, p_1).$$

Then we have

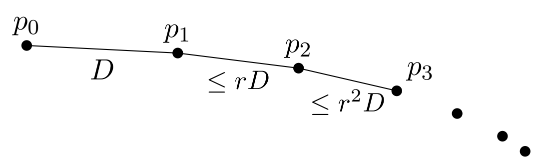
$$d(p_1, p_2) = d(f(p_0), f(p_1)) \le r \cdot d(p_0, p_1) = rD,$$

and

$$d(p_2, p_3) = d(f(p_1), f(p_2)) \le r \cdot d(p_1, p_2) = r^2 D,$$

and similarly

$$d(p_n, p_{n+1}) \le r^n D.$$



If we have some integer N and $n \ge m \ge N$, then by the triangle inequality,

$$\begin{aligned} d(p_m, p_n) &\le d(p_m, p_{m+1}) + d(p_{m+1}, p_{m+2}) + \cdots + d(p_{n-1}, p_n) \\ &\le r^m D + r^{m+1} D + \cdots + r^{n-1} D = \frac{r^m - r^n}{1 - r} \cdot D \le \frac{r^N}{1 - r} \cdot D, \end{aligned}$$

where in the third step we have summed a geometric series.

Now let $\epsilon > 0$ be given. Because $r < 1$, we can choose an integer N such that

$$\frac{r^N}{1 - r} \cdot D < \epsilon,$$

so if $m, n \ge N$ then $d(p_m, p_n) < \epsilon$. Thus the sequence is Cauchy, as claimed.

Because the metric is complete, there is an $\ell \in X$ such that $p_n \to \ell$. It remains to prove that $f(\ell) = \ell$. Exercise 4.4 asks you to prove that f is continuous. Thus it preserves limits by Exercise 1.11, and we can write

$$f(\ell) = f\left(\lim_{n\to\infty} p_n\right) = \lim_{n\to\infty} f(p_n) = \lim_{n\to\infty} p_{n+1} = \ell. \qquad \square$$

To give an idea of the power of the Banach fixed-point theorem, I'll digress to sketch the proof of existence of solutions to ordinary differential equations, but again you can skip ahead if you like. To take a concrete example, the motion of a pendulum is described by

$$\theta''(t) = -a\sin\theta(t) - b\theta'(t)$$

where θ is the angle that the pendulum makes with the vertical, a is a constant related to the length of the pendulum, and b is a constant related to friction. First we repackage this second-order differential equation as a system of first-order differential equations by introducing the angular speed ω and writing

$$\theta'(t) = \omega(t) \qquad\qquad \omega'(t) = -a\sin\theta(t) - b\omega(t),$$

We continue repackaging: if we let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$F(x,y) = (y,\ -a\sin x - by),$$

and we let $f\colon [0,1] \to \mathbb{R}^2$ be the vector-valued function whose components are $\theta(t)$ and $\omega(t)$, then we want to solve $f'(t) = F(f(t))$ with some initial condition $f(0) = C \in \mathbb{R}^2$.

Using the fundamental theorem of calculus, we can see that $f'(t) = F(f(t))$ and $f(0) = C$ if and only if

$$f(t) = C + \int_0^t F(f(s))\,ds.$$

And we observe that a solution to this integral equation is the same as a point $f \in C([0,1])^2$ that's a fixed point of the map Φ from $C([0,1])^2$ to itself given by $\Phi(f) = C + \int_0^t F(f(s))\,ds$. Now using the fact that F is continuously differentiable, we could do some serious analysis and cook up a Lipschitz constant r for Φ as in (4.2); then the Banach fixed-point theorem gives a solution to our original differential equation. The same method works for any differential equation.

Moreover, the proof of the theorem yields a good algorithm for solving differential equations numerically: start with a constant function $f(t) = C$, and hit it with Φ over and over until you get close enough to a solution. You can learn about this in a course in differential equations or numerical methods, but for now let me just say that it's fun to see how applying this method to the differential equation $f'(t) = f(t)$, $f(0) = 1$ yields the power series for e^t.

The Banach fixed-point theorem is also used to prove the inverse function theorem and the implicit function theorem.

Completion. We conclude this section by sketching how every metric space (X, d_X) can be embedded into a complete metric space $(\hat{X}, d_{\hat{X}})$. In examples, we usually have a more concrete way to construct the completion – we can construct $\mathbb{R}$ via Dedekind cuts, or $\mathbb{Q}_p$ via formal Laurent series in p, or $L^p([0,1])$ via Lebesgue integrable functions – but it is reassuring to know that a completion exists in any abstract situation.

We define $\hat{X}$ to be the set of equivalence classes of Cauchy sequences in X, where the equivalence relation is

$$\{p_n\} \sim \{p'_n\} \quad \Longleftrightarrow \quad \lim_{n\to\infty} d_X(p_n, p'_n) = 0. \tag{4.3}$$

Then we define

$$d_{\hat{X}}(\{p_n\}, \{q_n\}) = \lim_{n\to\infty} d_X(p_n, q_n). \tag{4.4}$$

There are many things to check:

(a) The relation (4.3) is an equivalence relation:

- Reflexive: $\{p_n\} \sim \{p_n\}$.
- Symmetric: if $\{p_n\} \sim \{p'_n\}$ then $\{p'_n\} \sim \{p_n\}$.
- Transitive: if $\{p_n\} \sim \{p'_n\}$ and $\{p'_n\} \sim \{p''_n\}$ then $\{p_n\} \sim \{p''_n\}$.

(b) The limit in (4.4) always exists.

(c) The function $d_{\hat{X}}$ defined in (4.4) is well-defined with respect to the equivalence relation (4.3): that is, if $\{p_n\} \sim \{p'_n\}$ and $\{q_n\} \sim \{q'_n\}$, then

$$d_{\hat{X}}(\{p_n\}, \{q_n\}) = d_{\hat{X}}(\{p'_n\}, \{q'_n\}).$$

(d) The function $d_{\hat{X}}$ defined in (4.4) is a metric, with the three properties given in Definition 1.2.

(e) The metric $d_{\hat{X}}$ on $\hat{X}$ is complete.

(f) While $\hat{X}$ does not literally contain X as a subset, the map $i\colon X \to \hat{X}$ that sends a point p to the constant sequence $p, p, p, \ldots$ preserves distances, so it identifies X with subset of $\hat{X}$.

(g) $\hat{X}$ is the smallest complete set that contains X, in the following sense: if (Y, d_Y) is a complete metric space and $j\colon X \to Y$ is a distance-preserving map, then there is a unique distance-preserving map $\hat{\jmath}\colon \hat{X} \to Y$ such that $\hat{\jmath} \circ i = j$.

Many of these are straightforward to prove: for example, for (c) we can write

$$\begin{aligned} d_{\hat{X}}(\{p_n\},\{q_n\}) &= \lim_{n\to\infty} d_X(p_n,q_n) \\ &\le \lim_{n\to\infty}\Big(d_X(p_n,p_n') + d_X(p_n',q_n') + d_X(q_n,q_n')\Big) \\ &= 0 + d_{\hat{X}}(\{p_n'\},\{q_n'\}) + 0, \end{aligned}$$

where the second step used the triangle inequality, and similarly

$$d_{\hat{X}}(\{p_n'\},\{q_n'\}) \le d_{\hat{X}}(\{p_n\},\{q_n\}),$$

so the two are equal. Exercise 4.6 asks you to prove (b). Here is a complete proof of (e), which is much harder than the others, but also terribly boring.

Proof of (e). Suppose we are given a Cauchy sequence in $\hat{X}$, that is, a Cauchy sequence of Cauchy sequences in X, where the second "Cauchy" is with respect to d_X and the first is with respect to $d_{\hat{X}}$. Let the first sequence be called $p_{1,1}, p_{1,2}, p_{1,3}, \ldots$, the second $p_{2,1}, p_{2,2}, p_{2,3}, \ldots$, and so on. We must construct a sequence $q_1, q_2, \ldots \in X$, prove that it is Cauchy with respect to d_X, and prove that our sequence of sequences converges to it with respect to $d_{\hat{X}}$

First let us define q_k. Because the sequence $p_{k,1}, p_{k,2}, p_{k,3}, \ldots$ is Cauchy with respect to d_X, we can apply the definition of Cauchy with $\epsilon = 1/k$ and get an integer N_k such that if $m, n \ge N_k$ then $d_X(p_{k,m}, p_{k,n}) < 1/k$. Then we set $q_k = p_{k,N_k}$.

Next let us argue that the sequence $q_1, q_2, q_3, \ldots$ is Cauchy with respect to d_X. Let $\epsilon > 0$ be given. Because the sequence $\{p_{1,n}\}, \{p_{2,n}\}, \{p_{3,n}\}, \ldots \in \hat{X}$ is Cauchy with respect to $d_{\hat{X}}$, there is an integer K such that if $k, l \ge K$ then

$$d_{\hat{X}}(p_{k,n}, p_{l,n}) < \epsilon/3. \tag{4.5}$$

Moreover we can increase K if necessary to get $1/K < \epsilon/3$. Fix some $k, l \ge K$. Recalling the definition of $d_{\hat{X}}$ from (4.4), we see that (4.5) says that there is an integer N such that if $n \ge N$ then

$$d_X(p_{k,n}, p_{l,n}) < \epsilon/3.$$

Set $n = \max(N_k, N_l, N)$. Then we have

$$\begin{aligned} d_X(q_k, q_l) &= d_X(p_{k,N_k},\, p_{l,N_l}) \\ &\le d_X(p_{k,N_k},\, p_{k,n}) + d_X(p_{k,n},\, p_{l,n}) + d_X(p_{l,n},\, p_{l,N_l}) \\ &< 1/k + \epsilon/3 + 1/l \le 1/K + \epsilon/3 + 1/K < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon, \end{aligned}$$

as desired.

Finally let us argue that $\{p_{1,n}\}, \{p_{2,n}\}, \{p_{3,n}\}, \dots \to \{q_n\}$ with respect to $d_{\hat{X}}$. Let $\epsilon > 0$ be given. Because $\{q_n\}$ is Cauchy with respect to d_X, there is an integer K such that if $k, l \geq K$ then $d_X(q_k, q_l) < \epsilon/2$. Increase K if necessary to get $1/K < \epsilon/2$. If $k \geq K$ then

$$d_{\hat{X}}(\{p_{k,n}\}, \{q_n\}) = \lim_{n\to\infty} d_X(p_{k,n}, q_n) \leq \lim_{n\to\infty} d_X(p_{k,n}, q_k) + \lim_{n\to\infty} d_X(q_k, q_n)$$

The first term of the sum on the right is

$$\lim_{n\to\infty} d_X(p_{k,n}, p_{k,N_k}),$$

where N_k is the one we chose in the previous paragraph; by construction, if $n \geq N_k$ then $d_X(p_{k,n}, p_{k,N_k}) < 1/k$, so the limit is $\leq 1/k < \epsilon/2$. For the second term, if $n \geq K$ then $d_X(q_k, q_n) < \epsilon/2$, so the limit is $\leq \epsilon/2$. Thus the sum is $< \epsilon$, as desired. □

Exercises.

4.1. Prove that the sequence of piecewise-linear functions $f_2, f_3, f_4, \dots \in C([0,1])$ introduced at the beginning of the section is Cauchy in the L^1 metric.

4.2. Let $X = \mathbb{Q}$, let p be a prime number, and let d_p be the p-adic meric defined in Example 4.8.

(a) Write down some rational numbers, and compute the 2-adic distance between them.

(b) Prove that d_p is a metric (for any prime p).

(c) Consider a "formal power series" in p,

$$a_0 + a_1 p + a_2 p^2 + a_3 p^3 + \cdots,$$

where the coefficients a_n are integers with $0 \leq a_n < p$. Let $s_0, s_1, s_2, \dots \in \mathbb{Z}$ be the sequence of partial sums of this series. Prove that it is Cauchy in the p-adic metric.

4.3. Let $p_1, p_2, p_3, \dots$ be a Cauchy sequence in a metric space (X, d). Prove that the sequence is bounded, meaning that there is a point $q \in X$ and a radius $R > 0$ such that $p_n \in B_R(q)$ for all n. (In fact, for any $q \in X$ you can find such a radius R, and in particular for $X = \mathbb{R}^n$ you can take $q = 0$.)

Hint: Start by applying the definition of Cauchy with $\epsilon = 1$ to get an N such that if $m, n \geq N$ then $d(p_m, p_n) < 1$, and in particular $d(p_N, p_n) < 1$.

4.4. Let (X, d) be a metric space, let $f\colon X \to X$, suppose there is a "Lipschitz constant" $r \in [0, 1)$ such that for all $p, q \in X$ we have

$$d(f(p), f(q)) \leq r \cdot d(p, q).$$

Prove that f is continuous.

Hint: Take $\delta = \epsilon$.

4.5. Here are two examples of how Theorem 4.13 can fail if the hypotheses are weakened.

(a) Instead of asking for a uniform constant $r < 1$ such that $d(f(p), f(q)) \leq r \cdot d(p, q)$ for all $p, q \in X$, we might just have asked that $d(f(p), f(q)) < d(p, q)$ whenever $p \neq q$. But let $X = [1, \infty)$ with the usual metric, and let $f\colon X \to X$ be defined by $f(x) = x + 1/x$; prove that f satisfies this weaker condition, but does not have a fixed point.

(b) The theorem can also fail if the space is not complete: let $X = [1, \infty) \cap \mathbb{Q}$ with the usual metric, and let $f\colon X \to X$ be defined by $f(x) = x/2 + 1/x$; prove that f satisfies the hypothesis of the theorem with of $r = 1/2$, but does not have a fixed point.

4.6. Prove item (b) from the laundry list of things to check about the completion of a metric space: if $\{p_n\}$ and $\{q_n\}$ are Cauchy sequences in X, then the limit

$$\lim_{n\to\infty} d_X(p_n, q_n)$$

exists.

Hint: Prove that $d_X(p_n, q_n)$ is a Cauchy sequence in $\mathbb{R}$ with its usual metric. The reverse triangle inequality (Exercise 1.6) may be useful.

5 The Baire Category Theorem

This section is devoted to the Baire category theorem, the gist of which is that a countable collection of "thin" subsets cannot cover very much of a complete metric space. We begin with two definitions.

Definition 5.1. Let (X, d) be a metric space, and let $A \subset X$.

(a) A is *dense* in X if $\bar{A} = X$, or equivalently, if every non-empty open subset of X intersects A. Some older texts call this *everywhere dense*.

(b) A is *nowhere dense* if $\operatorname{int} \bar{A} = \varnothing$, or equivalently (by Proposition 2.6), if $X \setminus \bar{A}$ is dense in X.

Example 5.2.

(a) In $\mathbb{R}$ with the usual metric, both the rational numbers $\mathbb{Q}$ and the irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ are dense.

(b) The Weierstrass approximation theorem says that in $C([0, 1])$ with the sup metric, the set of polynomials is dense: that is, any continuous function on $[0, 1]$ can be approximated as closely as you like by a polynomial. The proof is beyond the scope of these notes.

(c) A point in $\mathbb{R}^n$ is nowhere dense, as is a line in $\mathbb{R}^n$ when $n \geq 2$. Their complements are open and dense.

Theorem 5.3 (Baire category theorem)**.** *Let (X, d) be a complete metric space, and let $A_1, A_2, A_3, \ldots \subset X$ be a countable collection nowhere dense subsets. Then their union $A_1 \cup A_2 \cup A_3 \cup \cdots$ has empty interior, or equivalently (by Proposition 2.6), its complement is dense in X.*

Thus, for example, the complement of any countable subset of $\mathbb{R}^n$ is dense, as is the complement of a countable union of lines in $\mathbb{R}^n$ when $n \geq 2$. The theorem can also be formulated as follows:

Theorem 5.4 (Baire category theorem, alternate version)**.** *Let (X, d) be a complete metric space, and let $U_1, U_2, U_3, \ldots \subset X$ be a countable collection of open, dense subsets. Then the intersection $U_1 \cap U_2 \cap U_3 \cap \cdots$ is again dense.*

Exercise 5.1 asks you to prove that Theorems 5.3 and 5.4 are equivalent. Exercise 5.2 asks you to explore how Theorem 5.4 can fail if the hypotheses are weakened.

The name of the theorem has nothing to do with category theory, but comes from Baire's original terminology [2, p. 65]: he defined a set *of the first category* as one that can be written as a countable union of nowhere dense sets, and a set *of the second category* as one that cannot.* This sort of naming was widespread in 19th and early 20th century mathematics – Bessel functions of the first and second kinds, elliptic integrals of the first, second, and third kinds, differentials of the first, second, and third kinds on Riemann surfaces – but has fallen out of favor because it's hard to remember which kind is which. Modernly, a set of the first category is sometimes called *meager*, and the complement of a meager set is sometimes called *co-meager* or *residual*.

Before proving the theorem, let me give some idea of how it is used in practice.

(a) Exercise 5.4 asks you to use the Baire category theorem to prove that a complete metric space with no "isolated points" is uncountable. Exercise 5.5 asks you to deduce that the p-adic metric on $\mathbb{Q}$ is incomplete.

(b) In $C([0,1])$ with the sup metric, the set of nowhere-differentiable functions – that is, functions $f(x)$ such that the derivative

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}$$

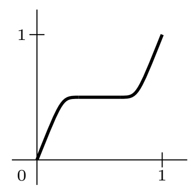
does not exist for any x – is dense. The proof is more involved than we want to discuss here, but one starts by letting F_n be the set of functions $f \in C([0,1])$ for which there is some $x \in [0,1]$ such that for all $y \in [0,1]$ with $y \neq x$ we have

$$\left| \frac{f(y) - f(x)}{y - x} \right| \leq n.$$

With a little effort we can see that if f is differentiable at some point $x \in [0,1]$, then f is in some F_n, so the set of somewhere-differentiable functions is contained in the union $F_1 \cup F_2 \cup \cdots$. With more effort we can prove that F_n is closed, and that its complement is dense. Thus the complement of $F_1 \cup F_2 \cup \cdots$ is dense by the Baire category theorem, and this is contained in the set of nowhere-differentiable functions, so the latter is dense as well.

*Although later authors sometimes define a set of the second category as one whose complement is of the first category.

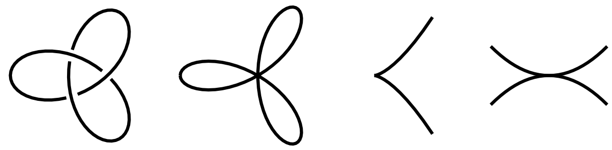
(c) Some versions of Sard's theorem rely on the Baire category theorem. To state Sard's theorem, we can warm up with a continuously differentiable function $f\colon \mathbb{R} \to \mathbb{R}$. A *critical point* of f is a point $x \in \mathbb{R}$ with $f'(x) = 0$, the same meaning as in a first calculus course. On the other hand, a *critical value* of f is a point $y \in \mathbb{R}$ such that $y = f(x)$ for some critical point $x \in \mathbb{R}$, and a *regular value* is one that is not a critcal value. The set of critical points could very well have non-empty interior,



but Sard's theorem states that the set of critical values is meager, so the set of regular values is dense by the Baire category theorem. In several variables, a critical point of a continuously differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is one where the matrix of partial derivatives is not invertible, and a critical point of $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is one where the rank of matrix of partial derivatives is less than n; then critical and regular values are defined in the same way, and Sard's theorem states that if f is $n - m + 1$ times continuously differentiable then the set of critical points is meager, so the set of regular values is dense.

(A more common formulation of Sard's theorem states that the set of critical values has *measure zero*, which is not the same as being meager, but it also implies that the set of regular values is dense. Infinite-dimensional versions of Sard's theorem must be phrased in terms of meager sets, however, because one cannot speak about measure zero in infinite dimensions.)

Sard's theorem is used all the time in differential topology, especially when dealing with questions of transversality. For example, if we have a knot in $\mathbb{R}^3$, meaning a copy of the circle embedded via a continuously differentiable map, we can use Sard's theorem to get a linear projection to $\mathbb{R}^2$ that gives a "knot diagram" where the crossings are only transverse double points, not triple points or cusps or worse:



yes no no no

(d) As an application to algebraic geometry, let us use the Baire category theorem to prove that among all circles in $\mathbb{R}^2$, the ones with no rational points are dense. A circle is, of course, a set described by an equation of the form $(x-a)^2 + (y-b)^2 = r^2$, so we can identify the set of all circles with the set

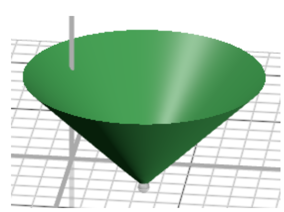
$$X = \{(a, b, r) \in \mathbb{R}^3 : r \geq 0\}.$$

A *rational point* is just an element of $\mathbb{Q}^2 \subset \mathbb{R}^2$. It is easy to write down circles that have no rational points: for example $x^2 + y^2 = \pi$ or, more subtly, $x^2 + y^2 = 3$. Let us argue that the set of circles that *do* have rational points form a meager subset of X, and thus the ones that don't are dense.

To see this, fix a point $(x, y) \in \mathbb{Q}^2$, and let $F_{x,y} \subset X$ be the set of all circles that contain (x, y): that is,

$$F_{x,y} = \{(a, b, r) \in X : (x-a)^2 + (y-b)^2 = r^2\}.$$

With some thought, we can see that this forms a cone in (a, b, r)-space, with its vertex at $(x, y, 0)$:



In particular, $F_{x,y}$ is closed, but its interior is empty, so it is nowhere dense. Now the set of all circles with a rational point is

$$\bigcup_{(x,y) \in \mathbb{Q}^2} F_{x,y},$$

a countable union of nowhere dense sets, so by the Baire category theorem, its complement is dense.

A much more sophisticated result in the same vein is the Noether–Lefschetz theorem. Consider the set of all polynomials $f(x, y, z)$ whose terms have degree at most 4: so we allow terms like x^4 and $12x^3y$ and $-xy^2z$ and also lower-degree terms. Such a polynomial turns out to have 35 coefficients, so we can identify the set of all such polynomials with $\mathbb{R}^{35}$, or if we allow complex coefficients, $\mathbb{C}^{35}$. Each polynomial determines a "quartic surface" in $\mathbb{R}^3$ or $\mathbb{C}^3$. The theorem says roughly that the polynomials defining surfaces that contain unexpectedly many curves form a meager subset of $\mathbb{R}^{35}$ or $\mathbb{C}^{35}$, and in particular, for a dense subset of polynomials, the only way to get algebraic curves on the corresponding surface is by intersecting with another surface. In fact the theorem holds for higher-degree surfaces as well; but surfaces of degree 3 always contain lines, which are not obtained by intersecting with another surface.

Now we turn to the proof of Theorem 5.4. We will need a generalization of the nested interval theorem, which in turn needs the following definition.

Definition 5.5. Let (X, d) be a metric space. The *diameter* of a non-empty subset $A \subset X$ is

$$\operatorname{diam}(A) = \sup_{p,q \in A} d(p, q),$$

which could be ∞ if A is unbounded. The diameter of the empty set is 0.

Proposition 5.6. *Let X be a complete metric space, and suppose that $F_1 \supset F_2 \supset F_3 \supset \cdots$ is a nested sequence of non-empty, closed subsets with* $\operatorname{diam}(F_n) \to 0$ *as $n \to \infty$. Then the intersection $F_1 \cap F_2 \cap F_3 \cap \cdots$ is not empty.*

Proof. Because every F_n is non-empty, we can choose a point $p_n \in F_n$ for each n. First I claim that the sequence $p_1, p_2, \ldots$ is Cauchy. To see this, let $\epsilon > 0$ be given. Because $\operatorname{diam}(F_n) \to 0$ as $n \to \infty$, there is an integer N such that if $n \geq N$ then $\operatorname{diam}(F_n) < \epsilon$, and in particular $\operatorname{diam}(F_N) < \epsilon$. If $m, n \geq N$, then p_m and p_n are both in F_N (because $F_m \subset F_N$ and $F_n \subset F_N$), so $d(p_m, p_n) \leq \operatorname{diam}(F_N) < \epsilon$. Thus the sequence is Cauchy, as claimed.

Because X is complete, the Cauchy sequence $p_1, p_2, \ldots$ converges to a limit $\ell \in X$. I claim that $\ell \in F_n$ for every n: to see this, note that the sequence $p_n, p_{n+1}, p_{n+2}, \ldots$ stays in F_n and still converges to ℓ; because F_n is closed, we have $\ell \in F_n$.

Thus $\ell \in F_1 \cap F_2 \cap \cdots$, so the intersection is not empty. $\square$

Exercise 5.7 asks you to explore how the preceding proposition can fail if the hypotheses are weakened.

Finally we turn to the proof of the Baire category theorem:

Proof of Theorem 5.4. We have a countable collection of dense, open subsets $U_1, U_2, \ldots \subset X$, and we want to prove that the intersection $U_1 \cap U_2 \cap \cdots$ is again dense: that is, for every $p \in X$ and every $r > 0$, the ball $B_r(p)$ meets $U_1 \cap U_2 \cap \cdots$.

Because U_1 is dense, we can choose a point $p_1 \in B_r(p) \cap U_1$. Because U_1 is open, there is a radius $r_1 > 0$ such that $B_{r_1}(p_1) \subset B_r(p) \cap U_1$. Moreover we can shrink r_1 if necessary to make $r_1 < 1$. Let $F_1 = \bar{B}_{r_1/2}(p_1)$, which is closed, contained in $B_r(p) \cap U_1$, and has $\operatorname{diam}(F_1) \leq r_1 < 1$.

Because U_2 is dense, we can choose a point $p_2 \in B_{r_1/2}(p_1) \cap U_2$. Because U_2 is open, there is a radius $r_2 > 0$ such that $B_{r_2}(p_2) \subset B_{r_1/2}(p_1) \cap U_2$. Moreover we can shrink r_2 if necessary to make $r_2 < \frac{1}{2}$. Let $F_2 = \bar{B}_{r_2/2}(p_2)$, which is closed, contained in $B_r(p) \cap U_1 \cap U_2$, and has $\operatorname{diam}(F_2) \leq r_2 < \frac{1}{2}$.

Continue in this way, getting a sequence of closed subsets $F_1 \supset F_2 \supset F_3 \supset \cdots$ with $F_n \subset B_r(p) \cap U_1 \cap U_2 \cap \cdots \cap U_n$ and $\operatorname{diam}(F_n) < \frac{1}{n}$. By Lemma 5.6, the intersection $F_1 \cap F_2 \cap \cdots$ is not empty, and by construction it is contained in $B_r(p) \cap U_1 \cap U_2 \cap \cdots$, so the latter is not empty either. $\square$

Exercises.

5.1. Use Proposition 2.6 to prove that Theorems 5.3 and 5.4 are equivalent.

5.2. Give examples to show that Theorem 5.4 can fail...

(a) if the subsets U_i are dense but not open.

(b) if the metric space X is not complete.

(c) if the collection of open, dense subsets is uncountable.

5.3. Let (X, d) be any metric space (possibly incomplete), and let $U, V \subset X$ be two open, dense subsets. Prove that $U \cap V$ is again dense.

5.4. (a) A point p in a metric space (X, d) is called *isolated* if there is some $r > 0$ such that $B_r(p) = \{p\}$. Use the Baire category theorem to prove that a complete metric space with no isolated points is uncountable.

(b) Give an example of a countable, complete metric space.

5.5. Prove that the p-adic metric on $\mathbb{Q}$, defined in Example 4.8, has no isolated points. Use Exercise 5.4(a) to conclude that this metric is incomplete.

5.6. In Proposition 5.6, prove moreover that the intersection $F_1 \cap F_2 \cap \cdots$ consists of exactly one point.

5.7. Give examples to show that Proposition 5.6 can fail...

(a) if the sets F_n are not closed.

(b) if the metric space X is not complete.

(c) if the diameters all finite, but do not go to zero.

(This is really tricky, because it can't happen in $\mathbb{R}^n$ with the Euclidean metric. But let $X = C([0,1])$ with the sup metric, and let F_n be the set of continuous functions $f\colon [0,1] \to [0,1]$ with $f(0) = 1$ and $f(x) = 0$ for $x \geq \frac{1}{n}$. We can see that $F_1 \supset F_2 \supset \cdots$, and the proof that each F_n is closed is similar to Exercise 3.4. Prove that $\operatorname{diam}(F_n) = 1$, but that $F_1 \cap F_2 \cap \cdots$ is empty.)

6 Image and Preimage

The goal of this section is to characterize continuity involving only open sets, with no direct reference to metrics. We begine with some set theory.

Definition 6.1. Let X and Y be sets, and let $f\colon X \to Y$.

(a) The *image* of a subset $A \subset X$ is

$$f(A) = \{f(a) : a \in A\},$$

or if you prefer,

$$f(A) = \{y \in Y : y = f(a) \text{ for some } a \in A\}.$$

It is a subset of Y.

(b) The *inverse image* or *preimage* of a subset $B \subset Y$ is

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

or if you prefer,

$$f^{-1}(B) = \{x \in X : f(x) = b \text{ for some } b \in B\}.$$

It is a subset of X.

Example 6.2. Let $X = Y = \mathbb{R}$, and let $f\colon X \to Y$ be the map given by $f(x) = x^2$.

(a) If A is the set of even numbers

$$A = \{\dots, -6, -4, -2, 0, 2, 4, 6, \dots\},$$

then $f(A)$ is the set of squares of even numbers:

$$f(A) = \{0, 4, 16, 36, \dots\}.$$

(b) If A is the closed interval $[0, 2]$, then $f(A) = [0, 4]$.

(c) If A is the open interval $(-1, 2)$, then $f(A) = [0, 4)$.

(d) If B is the set of even numbers, then $f^{-1}(B)$ is the set of numbers whose squares are even:

$$f^{-1}(B) = \{0, \pm\sqrt{2}, \pm 2, \pm\sqrt{6}, \dots\}.$$

(e) If $B = [0, 2]$, then $f^{-1}(B)$ is the set of numbers whose squares lie in that interval: $f^{-1}(B) = [-\sqrt{2}, \sqrt{2}]$.

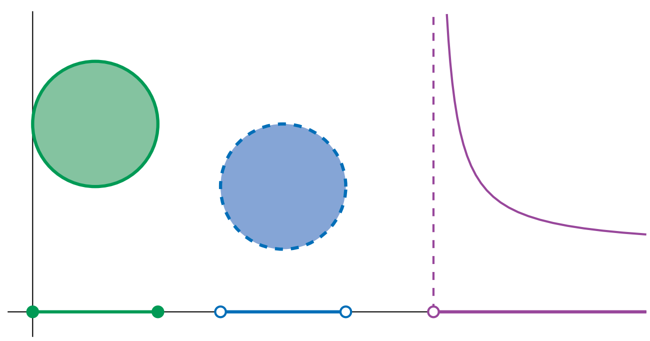
(f) If $B = (-1, 2)$, then $f^{-1}(B)$ is the set of numbers whose squares lie in that interval: $f^{-1}(B) = (-\sqrt{2}, \sqrt{2})$.

Warning 6.3. f^{-1} *is not a function.* Given a function $f\colon X \to Y$, we have defined the preimage $f^{-1}(B)$ of a subset $B \subset Y$, but you shouldn't think that f^{-1} is a function from Y to X: that is, you can't take a point $y \in Y$ and talk about $f^{-1}(y)$ as if it were a point of X. You can take $f^{-1}(\{y\})$, the preimage of a one-point set, but that might be one point, or many points, or the empty set. For example, if $f\colon \mathbb{R} \to \mathbb{R}$ is again given by $f(x) = x^2$, then $f^{-1}(\{0\}) = \{0\}$, but $f^{-1}(\{2\}) = \{-\sqrt{2}, \sqrt{2}\}$, which has two points, and $f^{-1}(\{-2\})$ is empty, because there is no $x \in \mathbb{R}$ with $f(x) = 2$.

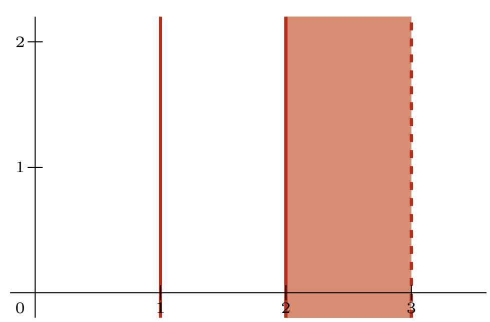
We *only* get an inverse function $f^{-1} : Y \to X$ when f is a bijection, and most of the functions we're interested in are not bijections.

Example 6.4. Let $X = \mathbb{R}^2$, let $Y = \mathbb{R}$, and let $f\colon X \to Y$ be the map given by $f(x, y) = x$, which projects down onto the x-axis.

(a) Here are some subsets $A \subset \mathbb{R}^2$ and their images $f(A) \subset \mathbb{R}$:



(b) Here is the preimage of $B = \{1\} \cup [2, 3) \subset \mathbb{R}$:



Exercises 6.1 and 6.2 ask you to explore how images and preimages interact with subsets, intersections, and unions. To give the flavor, let's prove now that the preimage of the complement is the complement of the preimage, but the image of the complement need not have much to do with the complement of the image.

Proposition 6.5. *If $f\colon X \to Y$ and $B \subset Y$, then $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$.*

Proof. For a point $p \in X$, we see that $p \in f^{-1}(Y \setminus B)$ if and only if $f(p) \in Y \setminus B$, which is true if and only if $f(p) \notin B$, which is true if and only if $p \notin f^{-1}(B)$, which is true if and only if $p \in X \setminus f^{-1}(B)$. $\square$

On the other hand, if $A \subset X$ then $f(X \setminus A)$ and $Y \setminus f(A)$ might be equal, or one might be contained in the other, or they might be incomparable. If we return to Example 6.2(c), with $X = Y = \mathbb{R}$ and $f(x) = x^2$ and $A = (-1, 2)$, then

$$f(X \setminus A) = f((-\infty, -1] \cup [2, \infty)) = [1, \infty),$$

while

$$Y \setminus f(A) = \mathbb{R} \setminus [0, 4) = (-\infty, 0) \cup [4, \infty).$$

Exercise 6.4 asks you to explore this further.

Now we come to the heart of the section:

Proposition 6.6. *Let (X, d_X) and (Y, d_Y) be metric spaces. A map $f\colon X \to Y$ is continuous if and only if for every open set $V \subset Y$, the preimage $f^{-1}(V)$ is open in X.*

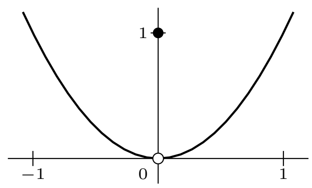
Proof. First suppose that f is continuous, and let $V \subset Y$ be open. We want to prove that $f^{-1}(V)$ is open in X, that is, for every $p \in f^{-1}(V)$ there is a $\delta > 0$ such that $B_\delta(p) \subset f^{-1}(V)$. Because $p \in f^{-1}(V)$, we have $f(p) \in V$. Because V is open, there is an $\epsilon > 0$ such that $B_\epsilon(f(p)) \subset V$. Because f is continuous, there is a $\delta > 0$ such that if $d_X(p, q) < \delta$ then $d_Y(f(p), f(q)) < \epsilon$. Thus if $q \in B_\delta(p)$, then $f(q) \in B_\epsilon(f(p))$, so $f(q) \in V$, or equivalently, $q \in f^{-1}(V)$. Thus we have proved that $B_\delta(p) \subset f^{-1}(V)$, which is what we wanted.

Conversely, suppose that f^{-1} of any open set is open. We want to prove that f is continuous, so let $p \in X$ and $\epsilon > 0$ be given. Take $V = B_\epsilon(f(p))$, which is an open subset of Y. By hypothesis, $f^{-1}(V)$ is open, and we have $p \in f^{-1}(V)$, so there is a $\delta > 0$ such that $B_\delta(p) \subset f^{-1}(V)$. Now if $d_X(p, q) < \delta$ then $q \in B_\delta(p)$, so $q \in f^{-1}(V)$, so $f(q) \in V = B_\epsilon(f(p))$, so $d_Y(f(p), f(q)) < \epsilon$, which is what we wanted. $\square$

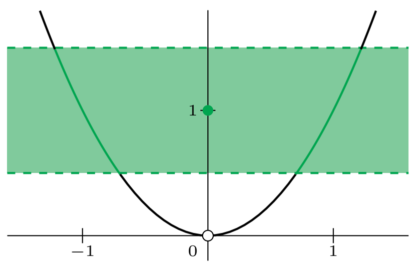
Example 6.7. Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^2 & \text{if } x \neq 0 \\ 1 & \text{if } x = 0, \end{cases}$$

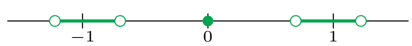
whose graph looks like this:



It is not continuous in the usual metric, so we should be able to find an open set $V \subset \mathbb{R}$ such that $f^{-1}(V)$ is not open. After a few tries, we might hit upon $V = (\frac{1}{2}, \frac{3}{2})$. To understand the preimage visually, take the horizontal strip $\frac{1}{2} < y < \frac{3}{2}$ and intersect it with the graph to get two arcs and an isolated point:



Then project those down onto the x-axis to get two open intervals and an isolated point:



So we see that $f^{-1}(V) = (-\sqrt{3/2}, -\sqrt{1/2}) \cup \{0\} \cup (\sqrt{3/2}, \sqrt{1/2})$, which is not open: the problem is at $x = 0$, which is exactly the point where f fails to be continuous.

Proposition 6.6 makes it very easy to prove that the composition of two continuous functions is continuous (in contrast to Exercise 1.7, which was not so easy). Suppose that $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous, and let $W \subset Z$ be open. Then $g^{-1}(W)$ is open in Y, so $f^{-1}(g^{-1}(W))$ is open in X. To conclude that $g \circ f$ is continuous, all that remains is to convince yourself that $f^{-1}(g^{-1}(W))$ is the same as $(g \circ f)^{-1}(W)$.

Using Proposition 6.5 and the fact that a set is closed if and only if its complement is open (Proposition 3.9), we can also characterize continuous maps in terms of closed sets:

Proposition 6.8. *Let (X, d_X) and (Y, d_Y) be metric spaces. A map $f\colon X \to Y$ is continuous if and only if for every closed set $G \subset Y$, the preimage $f^{-1}(G)$ is closed in X.*

Proof. First suppose that f is continuous, and let $G \subset Y$ be closed. Then $Y \setminus G$ is open by Proposition 3.9, so $f^{-1}(Y \setminus G)$ is open by Proposition 6.6, but this is the same as $X \setminus f^{-1}(G)$ by Proposition 6.5, so $f^{-1}(G)$ is closed.

Conversely, suppose that f^{-1} of every closed set is closed. Let $V \subset Y$ be open; then $Y \setminus V$ is closed, so $f^{-1}(Y \setminus V)$ is closed by hypothesis, but this is the same as $X \setminus f^{-1}(V)$ by Proposition 6.5, so $f^{-1}(V)$ is open. Thus f is continuous by Proposition 6.6. □

It is not true, however, that if $f\colon X \to Y$ is continuous and $U \subset X$ is open, then $f(U)$ is open in Y, as we see from Example 6.2(c). Nor is it true that if f is continuous and $F \subset X$ is closed, then $f(F)$ is closed in X, as we see from Example 6.4(a): the hyperbola on the right is closed, but its image in $\mathbb{R}$ is a half-open interval of the form (x, ∞).

Exercises.

6.1. (a) Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Find $f(A)$ for the following subsets $A \subset \mathbb{R}$: the intervals $[-1, 1]$, $[-1, 1)$, $(-1, 1)$, $[0, 1]$, $[0, 1)$, and $(0, 1)$, and the singletons $\{-1\}$, $\{0\}$, and $\{1\}$.

(b) Now let $f\colon X \to Y$ be arbitrary, and let $A, B \subset X$. Prove that if $A \subset B$ then $f(A) \subset f(B)$. Prove that $f(A \cup B) = f(A) \cup f(B)$. Prove that $f(A \cap B) \subset f(A) \cap f(B)$, but give an example where they are not equal.

6.2. (a) Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Find $f^{-1}(B)$ for the following subsets $B \subset \mathbb{R}$: the intervals $[-1, 1]$, $[-1, 1)$, $(-1, 1)$, $[0, 1]$, $[0, 1)$, and $(0, 1)$, and the singletons $\{-1\}$, $\{0\}$, and $\{1\}$.

(b) Now let $f\colon X \to Y$ be arbitrary, and let $A, B \subset Y$. Prove that if $A \subset B$ then $f^{-1}(A) \subset f^{-1}(B)$. Prove that $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$, and that $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

6.3. Let $f\colon X \to Y$, let $A \subset X$, and let $B \subset Y$.

(a) Prove that $A \subset f^{-1}(B)$ if and only if $f(A) \subset B$.

(b) Prove that $A \subset f^{-1}(f(A))$. Prove that they are equal if f is injective. Give an example to show that they need not be equal in general.

(c) Prove that $f(f^{-1}(B)) \subset B$. Prove that they are equal if f is surjective. Give an example to show that they need not be equal in general.

6.4. Let $f\colon X \to Y$, and let $A \subset X$. After Proposition 6.5 we saw an example where neither one of $f(X \setminus A)$ and $Y \setminus f(A)$ was contained in the other.

(a) Prove that if f is surjective then $Y \setminus f(A) \subset f(X \setminus A)$. Give an example where they are not equal.

(b) Prove that if f is injective then $f(X \setminus A) \subset Y \setminus f(A)$. Give an example where they are not equal.

6.5. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & \text{if } x \leq 0 \\ x+1 & \text{if } x > 0 \end{cases}$$

is discontinuous (with respect to the usual metric). Find an open set $V \subset \mathbb{R}$ such that $f^{-1}(V)$ is not open.

6.6. Use Proposition 6.6 to prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is continuous with respect to the usual metric, as follows.

(a) Let A be an open interval of the form $(a, \infty) \subset \mathbb{R}$. Find $f^{-1}(A)$ and observe that it is open.

(b) Let B be an the open interval of the form $(-\infty, b) \subset \mathbb{R}$. Find $f^{-1}(B)$ and observe that it is open.

(c) Let C be an open interval of the form $(a, b) \subset \mathbb{R}$. Prove from the previous two parts that $f^{-1}(C)$ is open.

(d) Let $V \subset \mathbb{R}$ be an artibrary open set. Prove that V is a union of open intervals. Conclude that $f^{-1}(V)$ is open.

(e) Look up a δ-ϵ proof that f is continuous and reproduce it here, or write one yourself. Which proof would you say is more straightforward?

6.7. Proposition 6.6 gave a statement involving only open sets, with no reference to metrics, that was equivalent to "$f\colon X \to Y$ is continuous."

   (a) Give a similar statement that is equivalent to "$f\colon X \to Y$ is continuous at a point $p \in X$."

   (b) Give a similar statement that is equivalent to "the sequence $p_1, p_2, p_3, \ldots \in X$ converges to a limit $\ell \in X$."

6.8. A map $f\colon X \to Y$ is called *open* if for every open set $U \subset X$, the image $f(U)$ is open in Y.

   (a) Give an example of a continuous map that is not open.

   (b) Give an example of an open map that is not continuous.

   (c) Give an example of a map that is both continuous and open.

   (d) Give an example of a map that is neither continuous nor open.

7 Topological Spaces

We have seen that which maps are continuous and which sequences converge depends only on which sets are open – that is, if two metrics on a set X give the same open sets, then they give the same continuous maps into X or out of X, and convergent sequences. So we introduce the following abstraction:

Definition 7.1. A *topology* on a set X is a collection subsets of X, called "open subsets," such that

(a) the empty set is open, and the whole set X is open,

(b) if U and V are open then $U \cap V$ are open, and

(c) if $\{U_i : i \in I\}$ is a collection of open sets, then $\bigcup_{i \in I} U_i$ is open.

A set equipped with a topology is called a *topological space*.

Example 7.2. If d is a metric on X, then the subsets $U \subset X$ that are open in the sense of §3 constitute a topology, by Exercise 3.5. But different metrics might give the same topology: for example, the Euclidean, taxicab, and square metrics give the same topology on $\mathbb{R}^n$ by Exercise 3.1. Or on $\mathbb{Z}$, the usual metric inherited from $\mathbb{R}$ gives the same topology as the discrete metric: every point is open, so every subset is open. On the other hand, the sup metric and the L^1 metric give different topologies on $C([0, 1])$, because we saw in Example 1.8 that they have different convergent sequences and continuous functions.

Example 7.3. Here are four more topologies on $\mathbb{R}$:

(a) The *finite complement topology*: a subset $U \subset \mathbb{R}$ is open if either the complement $\mathbb{R} \setminus U$ is finite, or $U = \varnothing$.

(b) The *particular point topology*: a subset $U \subset \mathbb{R}$ is open if either $0 \in U$, or $U = \varnothing$.

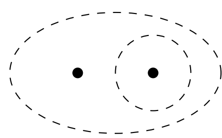
(c) The *lower semi-continuous topology*: the open sets are the intervals of the form (a, ∞) for some $a \in \mathbb{R}$, together with the empty set and the whole set $\mathbb{R}$. (The *upper semi-continuous topology* is similar but with $(-\infty, b)$.)

(d) The *lower limit topology*: a subset $U \subset \mathbb{R}$ is open if it can be written as a union of half-open intervals $[a, b)$. (The *upper limit topology* is similar but with $(a, b]$.)

Exercise 7.1 asks you to prove that these are topologies. Eventually we will see that none of them can come from a metric, but for now let us just see it for two of them: in a metric space, one-point sets are always closed, but in the particular point topology, $\{0\}$ is not closed (its complement is not open), and in the lower semi-continuous topology, no one-point sets are closed.

Example 7.4. On any set X we get two topologies for free: the *discrete topology*, in which every subset is open, and the *indiscrete topology*, in which the only open sets are $\varnothing$ and X. The discrete topology is the topology determined by the discrete metric (Exercise 1.4), but the indiscrete topology does not come from a metric if X has more than one point, again because points are not closed.

Example 7.5. *Sierpiński space* is the simplest topological space which is neither discrete nor indiscrete. It has two points, say $X = \{0, 1\}$, and the open sets are $\varnothing$, $\{1\}$, and X.



Sierpiński space is mostly used to construct counterexamples, but it does turn up in algebraic geometry as the Zariski topology on the set of prime ideals in a discrete valuation ring, such as the p-adic integers $\mathbb{Z}_p$ or the ring of formal power series $\mathbb{C}[[x]]$.

In moving from metric spaces to topological spaces, many theorems become definitions:

Definition 7.6. Let X be a topological space.

(a) If Y is another topological space, then a map $f\colon X \to Y$ is *continuous* if for every open subset $V \subset Y$, the preimage $f^{-1}(V)$ is open in X.

(b) A sequence of points $p_1, p_2, p_3, \ldots \in X$ *converges* to a limit $\ell \in X$ if for every open subset $U \subset X$ containing ℓ there is a natural number N such that for all $n \geq N$ we have $p_n \in U$.

(c) A subset $F \subset X$ is *closed* if $X \setminus F$ is open.

(d) For a subset $A \subset X$, the *interior* $\operatorname{int} A$ is the union of all open sets $U \subset A$, the *closure* $\bar{A}$ is the intersection of all closed sets $F \supset A$, and the *boundary* ∂A is the closure minus the interior.

So for example, in Sierpiński space, $\{0\}$ is closed, $\{1\}$ is open, and 0 is in the closure of $\{1\}$. We can still speak about subsets that are dense in X, or nowhere dense: so $\{1\}$ is dense in Sierpiński space, and $\{0\}$ is nowhere dense. Exercise 7.2 asks you to find the interiors and closures of some subsets of $\mathbb{R}$ in the four topologies of Example 7.3.

We cannot, however, define a *Cauchy* sequence in a topological space. To see this, let us revisit Example 4.5, with the metric on $\mathbb{R}$ given by

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

This metric determines the same open sets as the usual metric: essentially this follows from the fact that tan and arctan are continuous functions. But we have seen that the sequence $1, 2, 3, \ldots$ is Cauchy in this metric and not in the usual metric, so Cauchy-ness depends on more than just which sets are open. Similarly, the hyperbolic metrics of Example 4.7 determine the same open sets as the Euclidean metric, but they are complete while the Euclidean metric is not.

Proposition 7.7. *Let X and Y be topological spaces. A map $f\colon X \to Y$ is continuous if and only if for every closed set $G \subset Y$, the preimage $f^{-1}(G)$ is closed in X.*

Proof. The same as the proof of Proposition 6.8; just replace Proposition 6.6 (continuous iff the preimage of every open set is open) and Proposition 3.9 (closed iff the complement is open) with Definition 7.6(a) and (b). $\square$

Exercises.

7.1. Prove that each of the four topologies on $\mathbb{R}$ given in Example 7.3 is a topology, that is, it satisfies the three conditions in Definition 7.1.

7.2. Find the interiors, closures, and boundaries of the following subsets $A \subset \mathbb{R}$ in the topologies from Example 7.3:

(a) $\mathbb{Z} \subset \mathbb{R}$ in the finite complement topology.

(b) $\{0\} \subset \mathbb{R}$ and $\{1\} \subset \mathbb{R}$ in the particular point topology.

(c) $(0, 1) \subset \mathbb{R}$ in the lower semi-continuous topology.

(d) $(0, 1) \subset \mathbb{R}$ in the lower limit topology.

7.3. Let (X, d) be a metric space. A function $f \colon X \to \mathbb{R}$ is called *lower semi-continuous* at a point $p \in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d(p, q) < \delta$ implies $f(q) > f(p) - \epsilon$. The idea is that in the limit, f can only jump down. You can guess what *upper semi-continuous* means.

(a) Let $X = \mathbb{R}$ with the usual metric, and consider the floor function $\lfloor x \rfloor$, which returns the greatest integer $\leq x$, and the ceiling function $\lceil x \rceil$, which returns the least integer $\geq x$. Which one is lower semi-continuous, and which one is upper semi-continuous?

(You don't have to prove it.)

(b) Prove that $f \colon X \to \mathbb{R}$ is lower semi-continuous (at every point) if and only if it is continuous as a map of topological spaces when the codomain $\mathbb{R}$ is given the lower semi-continuous topology from Example 7.3(c).

(c) Perhaps the most important lower semi-continuous function is the rank of a matrix. Identify the space of $m \times n$ matrices with $\mathbb{R}^{mn}$ in the usual topology, and define the rank of a matrix M as the dimension of the column space (which is the same as the dimension of the row space). This is not continuous in the usual sense: for example, the matrices $\begin{pmatrix} \frac{1}{n} & 0 \\ 1 & 1 \end{pmatrix}$ have rank 2 for all n, but the limit as $n \to \infty$ has rank 1. Prove however that the rank is a lower semi-continuous function, using part (b) and the fact that a matrix has rank $\geq k$ if and only the determinant of some $k \times k$ submatrix is not zero.

7.4. Let (X, d) be a metric space. A function $f \colon \mathbb{R} \to X$ is called *continuous from the right* at a point $p \in \mathbb{R}$ if $\lim_{q \to p^+} f(q) = f(p)$: that is, for every $\epsilon > 0$ there is a $\delta > 0$ such that $p \leq q < p + \delta$ implies $d(f(p), f(q)) < \epsilon$. You can guess what *continuous from the left* means.

(a) Again take $X = \mathbb{R}$ in the usual metric, and consider the floor and ceiling functions. Which one is continuous from the right, and which one is continuous from the left?

(You don't have to prove it.)

(b) Prove that a map $f \colon \mathbb{R} \to X$ is continuous from the right (at every point) if and only if it is continuous as a map of topological spaces when the domain $\mathbb{R}$ is given the lower limit topology from Example 7.3(d).

7.5. The *countable complement* topology on $\mathbb{R}$ is like the finite complement topology, but we say that a subset $U \subset \mathbb{R}$ is open if either the complement $\mathbb{R} \setminus U$ is countable, or $U = \varnothing$.

(a) Prove that in this topology, a sequence $p_1, p_2, p_3, \ldots \in \mathbb{R}$ converges to a limit $\ell \in \mathbb{R}$ if and only if it is eventually constant, meaning that there is an N such that for all $n \geq N$ we have $p_n = \ell$.

(b) Give an example of a subset $A \subset \mathbb{R}$ that is not closed in this topology, but nonetheless for every sequence $p_1, p_2, \ldots \in A$ converging to a limit $\ell \in \mathbb{R}$, we have $\ell \in A$. (Thus Proposition 2.10 does not hold in a general topological space.)

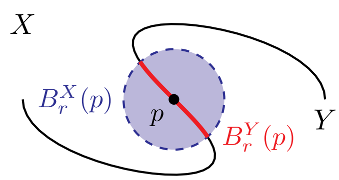
8 The Subspace Topology

Let $Y \subset X$. We saw in Example 1.7 that a metric d on X induces a metric on Y, just by declaring the distance between two points in Y to be the same as it was in the ambient space X. Now we want to explore how open subsets of Y are related to open subsets of X, and this will show us what to do with subspaces of topological spaces in general.

To get the feel of things, let $X = \mathbb{R}^2$ with the Euclidean metric, let Y be a curve, and let $p \in Y$. We can consider the ball of radius r around p in either X or Y; let us introduce some notation to distinguish the two:

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

Observe that $B_r^Y(p) = B_r^X(p) \cap Y$. We know that $B_r^Y(p)$ is open as a subset of Y, but we see that it is not open as a subset of X:



Proposition 8.1. *Let (X,d) be a metric space, let $Y \subset X$. Then a subset $V \subset Y$ is open in the induced metric if and only if there is an open set $U \subset X$ such $V = U \cap Y$.*

Proof. First suppose that $V = U \cap Y$ for some open set $U \subset X$. Let $p \in V$. Because U is open in X, there is a radius $r > 0$ such that $B_r^X(p) \subset U$. Intersecting with Y, we get $B_r^X(p) \cap Y \subset U \cap Y$, that is, $B_r^Y(p) \subset V$. Thus V is open in Y.

Conversely, suppose that V is open in Y. For each $p \in V$, choose a radius $r_p > 0$ such that $B_{r_p}^Y(p) \subset V$. Let

$$U = \bigcup_{p \in V} B_{r_p}^X(p),$$

which is a union of open balls in X, hence is open in X. Intersecting with Y, we get

$$U \cap Y = \bigcup_{p \in V} (B_{r_p}^X(p) \cap Y) = \bigcup_{p \in V} B_{r_p}^Y(p).$$

The latter union is contained in V by construction, but on the other hand it contains every point of V, so it is equal to V. $\square$

Example 8.2. Exercise 3.2 asked you to prove that in $\mathbb{Q}$ with the metric induced from the usual one on $\mathbb{R}$, the subsets

$$\{x \in \mathbb{Q} : x^2 < 1\} \qquad \text{and} \qquad \{x \in \mathbb{Q} : x^2 < 2\}$$

are both open. The first can be written as $(-1, 1) \cap \mathbb{Q}$, and the second as $(-\sqrt{2}, \sqrt{2}) \cap \mathbb{Q}$.

When we write $V = U \cap Y$, the choice of open set U in the ambient space is far from unique: for example, we could also have written the first set above as

$$\big((-1, \tfrac{1}{e}) \cup (\tfrac{1}{e}, 1)\big) \cap \mathbb{Q}.$$

From Proposition 8.1 we see that if two metrics give the same topology on X, then the induced metrics on Y give the same topology on Y. This leads us to the following definition for topological spaces:

Definition 8.3. Let X be a topological space, and let $Y \subset X$. The *subspace topology* on Y consists of all subsets that can be written as $U \cap Y$ for some open set $U \subset X$.

Proposition 8.4. *The subspace topology is a topology.*

Proof. Let us work through the three axioms in Definition 7.1:

(a) We know that $\varnothing$ and X are open in X, and we can write $\varnothing = \varnothing \cap Y$ and $Y = X \cap Y$, so $\varnothing$ and Y are open in Y.

(b) Let V_1 and V_2 be open subsets of Y. Write $V_1 = U_1 \cap Y$ and $V_2 = U_2 \cap Y$ for some open subsets U_1 and U_2 of X. Then $U_1 \cap U_2$ is open in X, and

$$V_1 \cap V_2 = (U_1 \cap Y) \cap (U_2 \cap Y) = (U_1 \cap U_2) \cap Y,$$

so $V_1 \cap V_2$ is open in Y.

(c) Let $\{V_i : i \in I\}$ be a collection of open subsets of Y. For each $i \in I$, write $V_i = U_i \cap Y$ for some open set $U_i \subset X$. Then $\bigcup_{i \in I} U_i$ is open in X, and

$$\bigcup_{i \in I} V_i = \bigcup_{i \in I} (U_i \cap Y) = \left(\bigcup_{i \in I} U_i \right) \cap Y,$$

so $\bigcup_{i \in I} V_i$ is open in Y. $\square$

Proposition 8.5. *Let X be a topological space, and let $Y \subset X$. Then a subset $G \subset Y$ is closed in the subspace topology if and only if there is a closed subset $F \subset X$ such that $G = F \cap Y$.*

Proof. The key set-theoretic observation is that for any subset $A \subset X$,

$$(X \setminus A) \cap Y = Y \setminus (A \cap Y).$$

You may want to draw a Venn diagram to convince yourself of this.

If $G = Y \cap F$ for some closed set $F \subset X$, then taking complements we get

$$Y \setminus G = Y \setminus (Y \cap F) = (X \setminus F) \cap Y.$$

Now $X \setminus F$ is open in X, so $Y \setminus G$ is open in Y by definition of the subspace topology, so G is closed in Y.

Conversely, If G is closed in Y, then $Y \setminus G$ is open in Y, so there is an open set $U \subset X$ such that $Y \setminus G = U \cap Y$. Taking complements, we get

$$G = Y \setminus (U \cap Y) = (X \setminus U) \cap Y.$$

So we can take $F = X \setminus U$, which is closed in X. $\square$

Example 8.6. Exercise 3.2(b) also asked you to prove that

$$\{x \in \mathbb{Q} : x^2 < 2\}$$

is closed in $\mathbb{Q}$ with the metric induced from the usual one on $\mathbb{R}$. We see that it can be written as $[-\sqrt{2}, \sqrt{2}] \cap \mathbb{Q}$.

Example 8.7. Let $X = \mathbb{R}$ with the usual topology, and let Y be the half-open interval $(0, 1]$. Then $(\frac{1}{2}, 1]$ is not open in X, but it is open in Y because we can write it as $(\frac{1}{2}, \frac{3}{2}) \cap Y$. Similarly, $(0, \frac{1}{2}]$ is not closed in X, but it is closed in Y because we can write it as $[0, \frac{1}{2}] \cap Y$. Exercise 8.1 asks you to explore everything that can happen along these lines.

Exercises.

8.1. Let X be a topological space. Let $Y \subset X$, and give Y the subspace topology. Let $A \subset Y$.

   (a) Prove that if A is closed in Y and Y is closed in X, then A is closed in X.

(b) Give two examples to show that if A is closed in Y and Y is *not* closed in X, then A may or may not be closed in X.

(c) Prove that if A is open in Y and Y is open in X, then A is open in X.

(d) Give two examples to show that if A is open in Y and Y is *not* open in X, then A may or may not be open in X.

(e) Let $A \subset Y$, let $\mathrm{cl}_X(A)$ denote the closure of A in X, and let $\mathrm{cl}_Y(A)$ denote the closure of A in Y. Prove that

$$\mathrm{cl}_Y(A) = \mathrm{cl}_X(A) \cap Y.$$

(f) Let $A \subset Y$, let $\mathrm{int}_X(A)$ denote the interior of A as a subset of X, and let $\mathrm{int}_Y(A)$ denote the interior of A as a subset of Y. Prove that

$$\mathrm{int}_X(A) \subset \mathrm{int}_Y(A).$$

(g) Give an example where the inclusion in part (f) is strict.

8.2. (a) Let X be a topological space, and suppose that we can write $X = F_1 \cup \ldots \cup F_n$, where each F_i is closed. Let Y be another topological space, let $f\colon X \to Y$, and let $f_i\colon F_i \to Y$ be the restriction of f to F_i: that is, for $x \in F_i$ we set $f_i(x) = f(x)$.

Prove that f is continuous if and only if f_i is continuous for all i.

Hint: Use the fact that a map is continuous if and only if the preimage of every closed set is closed.

(b) This is usually applied to show that a piecewise function is continuous. Consider the function $f\colon [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 1/3, \\ 3x - 1 & \text{if } 1/3 \leq x \leq 2/3, \\ 1 & \text{if } x \geq 2/3. \end{cases}$$

If we wanted to apply part (a) to show that f is continuous, which should sets should we take for the F_i?

(c) Give an example of how part (a) can fail if we allow countably many closed sets F_i.

8.3. *The universal property of the subspace topology.* Let X be a topological space, let $Y \subset X$, and give Y the subspace topology. If Z is another topological space and $f\colon Z \to Y$, prove that f is continuous as a map to Y if and only if f is continuous as a map to X.

(Here's an equivalent formulation, if you like it better: let $i\colon Y \to X$ be the inclusion map given by $i(y) = y$, and prove that $f\colon Z \to Y$ is continuous if and only if $i \circ f$ is continuous.)

References

[1] Stephen Abbott. *Understanding analysis.* Undergraduate Texts in Mathematics. Springer, New York, second edition, 2015.

[2] René Baire. Sur les fonctions de variables réelles. *Annali di Matematica, Serie III*, 3:1–123, 1899.

[3] Douglas Dunham. M. C. Escher's use of the Poincaré models of hyperbolic geometry. In *Mathematics and modern art*, volume 18 of *Springer Proc. Math.*, pages 69–77. Springer, Heidelberg, 2012. Also available at `math-art.eu/Documents/pdfs/Dunham.pdf`.

[4] Felix Hausdorff. *Grundzüge der Mengenlehre.* Verlag von Veit & Comp, Leipzig, 1914.

[5] Michael Spivak. *Calculus.* Publish or Perish, Inc., Houston, fourth edition, 2008.

[6] Heinrich Tietze. Beiträge zur allgemeinen Topologie. I. *Math. Ann.*, 88(3-4):290–312, 1923.