An Activity-Based Introduction to Topology

Steven Schlicker

Grand Valley State University

Contents

Part I

Sets and Functions

Section 1

Sets

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a set?
- What is a subset of a set?
- What is the union of two sets? How do we define the union of an arbitrary collection of sets?
- What is the intersection of two sets? How do we define the intersection of an arbitrary collection of sets?
- What is the complement of a set?
- What is the Cartesian product of sets?

Introduction

If you like geometry, you will probably like topology. Geometry is the study of objects with certain attributes (e.g., shape and size), while topology is more general than geometry. In topology, we aren't concerned about the attributes (shape and size) of an object, only about those characteristics that don't change when we transform the object in different ways (any way that doesn't involve tearing or poking holes the object). There are lots of really interesting theorems in topology - for example, the Hairy Ball Theorem which states that if you have a ball with hair all over it (think of a tribble from Star Trek - if that isn't too old of a reference), it is impossible to comb the hairs continuously and have all the hairs lay flat. Some hair must be sticking straight up!

Preview Activity 1.1.

(1) Take a pipe cleaner, a rubber band, or a pieces of string and make a square from it. You are allowed to deform the square by moving parts of the square without breaking it or lifting it off

the surface it is on. To which of the following shapes can you deform your square? Explain.

(a) a circle (b) the letter `S` (c) a five point star ☆ (d) the letter `D`

(2) Now take some play-doh (if you don't have any play-doh, just use your imagination). Use the play-doh (or your imagination) to determine which of the following shape can be deformed into others without breaking or making holes?

(a) a filled sphere (b) a doughnut (c) a bowl (d) a coffee mug with handle

(3) As the previous examples attempt to illustrate, at its most basic level topology deals with sets and how we can deform sets into other sets. So to start our study of topology, we begin with sets. Much of this material should be familiar, but some might be new. The first issue for us to settle on is as precise a definition of "set" as possible. Suppose we try to define a set to be a collection of elements. So, by definition, the elements are the objects contained in the set. We use the symbol $\in$ to denote that an object is an element of a set. So $\notin$ means an object is not in the set – if x is an object in a set S we write $x \in S$, and if x is not an object in a set S we write $x \notin S$. We write sets using set brackets. For example, the set $\{a, b, c\}$ is the set whose elements are the symbols a, b, and c. We can also include in the set notation conditions on elements of the set. For example, $\{x \in \mathbb{R} : x > 0\}$ is the set of positive real numbers. We typically use capitol letters to denote sets. Some familiar examples of sets are $\mathbb{R}$, the set of real numbers; $\mathbb{Q}$, the set of rational numbers; and $\mathbb{Z}$, the set of integers.

Consider the following set S, which is a set according to our definition:

$$S = \{A \text{ is a set} : A \notin A\}.$$

That is, S is the collection of sets that do not have themselves as elements.

Given any object x, either $x \in S$ or $x \notin S$.

(a) Is S an element of S? Explain.

(b) Is it the case that $S \notin S$? Explain.

(c) Based on your responses to parts i and ii, explain why our current definition of a set is not a good one.

(4) Assume that we have a working definition of a set. In this part of the activity we define a subset of a set. The notation we will use is $A \subset S$ if A is a subset of S that is not equal to S, and $A \subseteq S$ if A is a subset of S that could be the entire set S.

(a) How should we define a subset of a set? Give a specific example of a set and two examples of subsets of that set.

(b) If A is a set, is A a subset of A? Explain.

(c) What is the empty set $\emptyset$? If A is a set, is $\emptyset$ a subset of A? explain.

Activity Solution.

(1)

We can take the unit circle and fix the points where the circle intersects the coordinate axes. Push the remaining points on the circle toward the square in a radial manner. This deforms the circle onto the square.

(b) In order to make the letter `S`, we would have to cut the circle. Since this is not allowed, we cannot deform the circle to the letter `S`.

(c) Think of the circle as being inscribed inside the star. Then push out the portions of the circle that are not on the star toward the arms. This deforms the circle onto the star.

(d) Consider the unit circle as our circle. Project the left half of the circle onto the y-axis. This deforms the circle onto the letter `D`.

(2) Smash down the top half of the sphere onto the bottom half. This makes a bowl. To make a doughnut or a coffee mug with a handle, we would have to punch a hole in the sphere. So the sphere and the doughnut can be deformed onto one another, but neither can be deformed into a doughnut or a coffee mug. However, we can deform a coffee mug into a doughnut (see https://en.wikipedia.org/wiki/File:Mug_and_Torus_morph.gif for an animation. The process involves molding the mug part around the handle until the result is a doughnut. This process can be reversed, so a coffee mug and a doughnut can be deformed into each other.

(3) (a) If $S \in S$, then S does not contain itself as an element. But then S cannot be in S, so this is not possible.

(b) If $S \notin S$, then S must contain itself as an element. But this contradicts the fact that $S \notin S$.

(c) Our current definition of a set leads to a paradox – a set with an element that is neither in the set nor not in the set.

(4) (a) A subset U of a set S is a collection of elements of S. That is, U is a subset of S if $x \in S$ whenever $x \in U$. For example, if $S = \{1, 2, 3, 4\}$, then $\{1, 2\}$ and $\{1, 3, 4\}$ are subsets of S.

(b) Since every element of A is also an element of A, by our definition it is the case that A is a subset of A.

(c) The empty set is the set that contains no elements. Since $\emptyset$ contains no elements, by default every element of $\emptyset$ is an element of any other set. So $\emptyset$ isa subset of every set.

Unions, Intersections, and Complements of Sets

What we saw in our preview activity is what is called a paradox. Our original attempt to define a set led to an impossible situation since both $S \in S$ and $S \notin S$ lead to contradictions. This paradox is called *Russell's paradox* after Bertrand Russell, although it was apparently known before Russell. The moral of the story is that we need to be careful when making definitions. A set might seem like a simple object, and in our experience usually is, but formally defining a set can be problematic. As a result, we won't state a formal definition, but rather take a set to be a *well-defined* collection of objects. The objects are called the elements of the set. (In axiomatic set theory, a set is taken to be an undefined primitive – much as a point is undefined in Euclidean geometry.)

In order to effectively work with sets, we need to have an understanding what it means for two sets to be equal.

Activity 1.1.

(a) What should it mean for two sets to be equal? If A and B are sets, how do we prove that $A = B$?

(b) Let $A = \{x \in \mathbb{R} : x < 2\}$ and $B = \{x \in \mathbb{R} \mid x - 1 < 1\}$. Is $A = B$? If yes, prove your answer. If no, prove any containment that you can.

(c) Let $A = \{n \in \mathbb{Z} \mid 2 \text{ divides } n\}$ and $B = \{n \in \mathbb{Z} \mid 4 \text{ divides } (n-2)\}$. Is $A = B$? If yes, prove your answer. If no, prove any containment that you can.

(d) Let $A = \{n \in \mathbb{Z} \mid n \text{ is odd}\}$ and $B = \{n \in \mathbb{Z} \mid 4 \text{ divides } (n-1) \text{ or } 4 \text{ divides } (n-3)\}$. Is $A = B$? If yes, prove your answer. If no, prove any containment that you can.

Activity Solution.

(a) Two sets A and B are equal if $A \subseteq B$ and $B \subseteq A$. So to prove that two sets A and B are equal, we choose an arbitrary element a in A and show that $a \in B$, and then choose an arbitrary element $b \in B$ and show that $b \in A$.

(b) First we show that $A \subseteq B$. Let $a \in A$. Then $a < 2$. This implies that $a - 1 < 2 - 1 = 1$, so $a \in B$. Thus, $A \subseteq B$. Now assume that $b \in B$. Then $b - 1 < 1$. But this implies that $(b-1) + 1 < 1 + 1$ or $b < 2$. Thus $b \in A$ and $B \subseteq A$. The two containments demonstrate that $A = B$.

(c) We show that $B \subset A$, but $A \neq B$. Let $b \in B$. Then 4 divides $b - 2$. It follows that there exists an integer k such that $b - 2 = 4k$. So $b = 2 + 4k = 2(1 + 2k)$ and 2 divides b. We conclude that $b \in A$ and $B \subseteq A$. However, 2 divides 4, but 4 does not divide $4 - 2$. So A is not a subset of B.

(d) We will prove that $A = B$. Let $a \in A$. Then a is odd, so there exists an integer q such that $a = 2q + 1$. Now q is either even or odd, and we consider the cases. If q is even, then $q = 2p$ for some integer p. In this case $a = 2(2p) + 1 = 4p + 1$ and 4 divides $a - 1$. If q is odd, then $q = 2p + 1$ for some integer p. Then we have $a = 2(2p + 1) + 1 = 4p + 3$ and 4 divides $a - 3$. Thus, $a \in B$ and $A \subseteq B$. now let $b \in B$. Then 4 divides $b - 1$ or 4 divides $b - 3$. If 4 divides $b - 1$, then $b - 1 = 4k$ for some integer k. But then $b = 1 + 4k = 1 + 2(2k)$ for some integer k, which shows that b is odd. If 4 divides $b - 3$, then $b = 4k + 3$ for some integer k. It follows that $b = 3 + 4k = 1 + 2(1 + 2k)$ for some integer k, which again shows that b is odd. Thus $b \in A$ and $B \subseteq A$. The two containments demonstrate that $A = B$.

Once we have the notion of a set, we can build new sets from existing ones. For example, we define the union, intersection, and complement of a set as follows.

- The **union** of sets A and B is the set $A \cup B$ defined as

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

- The **intersection** of sets A and B is the set $A \cap B$ defined as

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

- Let A be a subset of a set U. The complement of A in U is the set

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

  The complement of a set A in a set U is also denoted by $C_U(A)$, $C(A)$ (if the set U is understood), A^c, or even $U - A$.

We can visualize these sets using Venn diagrams. A Venn diagram is a depiction of sets using geometric figures. For example, if U is a set containing all other sets of interest (we call U the *universal* set), we can represent U as a large container (say a rectangle) with subsets A and B as smaller containers (say circles), and shade the elements in a given set. The Venn diagrams in Figure 1.1 depict the sets A, B, $A \cup B$, $A \cap B$, A^c, and B^c.

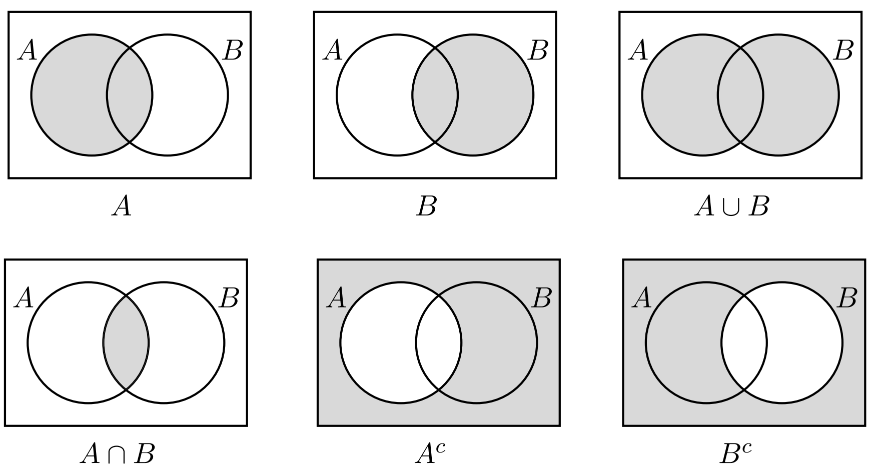


Figure 1.1: Vennn diagrams

Activity 1.2. In this activity we work with unions, intersections, and complements of sets. Let A and B be sets.

(a) If $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{2, 4, 6, 8, 10\}$, what are $A \cup B$ and $A \cap B$?
If $U = \{a, b, c, d, e, f, g\}$ and $A = \{c, e, g\}$, what is $U \setminus A$?

(b) Let A and B be subsets of a universal set U. There are connections between A, B, their complements, unions, and intersections.

i. Use Venn diagrams to draw $(A \cup B)^c$ and $(A \cap B)^c$.

ii. Find and prove a relationship between A^c, B^c and $(A \cup B)^c$.

iii. Find and prove a relationship between A^c, B^c and $(A \cap B)^c$.

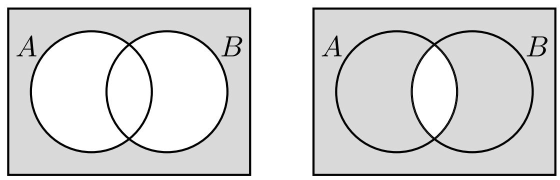
Activity Solution.

(a) In this example we have

$$A \cup B = \{1, 2, 3, 4, 5, 6, 8, 10\} \text{ and } A \cap B = \{2, 4, 6\}.$$

(b) In this example we have $U \setminus A = \{a, b, d, f\}$.

(c) Venn diagrams illustrating the sets $(A \cup B)^c$ and $(A \cap B)^c$ are shown below.



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

i. The Venn diagrams indicate that $(A \cup B)^c = A^c \cap B^c$. To prove this statement, start by letting $x \in (A \cup B)^c$. Then $x \notin (A \cup B)$. So $x \notin A$ and $x \notin B$. This means that $x \in A^c$ and $x \in B^c$, which implies that $x \in (A^c \cap B^c)$. We conclude that $(A \cup B)^c \subseteq (A^c \cap B^c)$.

Now suppose that $x \in (A^c \cap B^c)$. Then $x \in A^c$ and $x \in B^c$. So $x \notin A$ and $x \notin B$ and $x \notin (A \cup B)$. From this it follows that $x \in (A \cup B)^c$. We conclude that $(A^c \cap B^c) \subseteq (A \cup B)^c$.The two containments demonstrate that $(A \cup B)^c = A^c \cap B^c$.

ii. The Venn diagrams indicate that $(A \cap B)^c = A^c \cup B^c$. prove this statement, start by letting $x \in (A \cap B)^c$. Then $x \notin (A \cap B)$. So $x \notin A$ or $x \notin B$. This means that $x \in A^c$ or $x \in B^c$, which implies that $x \in (A^c \cup B^c)$. We conclude that $(A \cap B)^c \subseteq (A^c \cup B^c)$.

Now suppose that $x \in (A^c \cup B^c)$. Then $x \in A^c$ or $x \in B^c$. So $x \notin A$ or $x \notin B$, which means $x \notin (A \cap B)$. But this implies that $x \in (A \cap B)^c$. We conclude that $(A^c \cup B^c) \subseteq (A \cap B)^c$.The two containments demonstrate that $(A \cap B)^c = A^c \cup B^c$.

In Activity 1.2 we defined the union and intersection of two sets. There is no reason to restrict these definitions to only two sets, as the next activity illustrates.

Activity 1.3. Suppose $\{A_\alpha\}_{\alpha \in I}$ is a collection of sets indexed by some set I. For example, let $A_\alpha = \{1, 2, 3, \ldots \alpha\}$ in $\mathbb{Z}$, where α is any element of the indexing set $I = \mathbb{Z}^+$, the set of positive integers.

(a) What is A_5? What is A_8?

(b) How can we define the union of all of the sets A_α? In other words, how do we **define**

$$\bigcup_{\alpha \in I} A_\alpha?$$

In our particular example, what set is $\bigcup_{\alpha \in I} A_\alpha$?

(c) How can we define the intersection of all of the sets A_α? In other words, how do we **define**

$$\bigcap_{\alpha \in I} A_\alpha?$$

In our particular example, what set is $\displaystyle\bigcap_{\alpha \in I} A_\alpha$?

Activity Solution.

(a) In this case we have $A_5 = \{1, 2, 3, 4, 5\}$ and $A_8 = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

(b) We define $\bigcup_{\alpha \in I} A_\alpha$ as follows:

$$\bigcup_{\alpha \in I} A_\alpha = \{x \mid x \in A_\alpha \text{ for some } \alpha \in I\}.$$

In our example we have $\displaystyle\bigcup_{\alpha \in I} A_\alpha = \mathbb{Z}^+$.

(c) We define $\bigcap_{\alpha \in I} A_\alpha$ as follows:

$$\bigcap_{\alpha \in I} A_\alpha = \{x \mid x \in A_\alpha \text{ for all } \alpha \in I\}.$$

In our particular example, $\displaystyle\bigcap_{\alpha \in I} A_\alpha = \{1\}$.

The properties $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$ that we learned about in Activity 1.2 are called DeMorgan's Laws. These laws apply to any union or intersection of sets. The proofs are left to the reader.

Theorem 1.1 (DeMorgan's Laws)**.** *Let $\{A_\alpha\}$ is a collection of sets indexed by a set I in some universal set U. Then*

(1) $$\left(\bigcup_{\alpha \in I} A_\alpha\right)^c = \bigcap_{\alpha \in I} A_\alpha^c$$

(2) $$\left(\bigcap_{\alpha \in I} A_\alpha\right)^c = \bigcup_{\alpha \in I} A_\alpha^c$$

Activity 1.4. Verify DeMorgan's Laws in the specific case of $A_\alpha = \{1, 2, 3, \ldots \alpha\}$ in $U = \mathbb{Z}$, where α is any element of the indexing set $I = \mathbb{Z}^+$.

Activity Solution. Let $\mathbb{Z}^- = \{n \in \mathbb{Z} \mid n < 0\}$. First note that $A_\alpha^c = \mathbb{Z}^- \cup \{0\} \cup \{\alpha+1, \alpha+2, \ldots\}$. So $\bigcap_{\alpha \in I} A_\alpha^c = \mathbb{Z}^- \cup \{0\}$. Also, $\bigcup_{\alpha \in I} A_\alpha = \mathbb{Z}^+$, so $\left(\bigcup_{\alpha \in I} A_\alpha\right)^c = \mathbb{Z}^- \cup \{0\}$.

Also, $\bigcup_{\alpha \in I} A_\alpha^c = \mathbb{Z}$, and $\bigcap_{\alpha \in I} A_\alpha = \emptyset$ which implies that $\left(\bigcap_{\alpha \in I} A_\alpha\right)^c = \mathbb{Z}$.

Cartesian Products of Sets

The final operation on sets that we discuss is the *Cartesian product* (or *cross product*). This is an operation that we have seen before. When we draw the graph of a line $y = mx + b$ in the plane, we plot the points $(x, mx + b)$. These points are ordered pairs of real numbers. We can extend this idea to any sets.

Definition 1.2. Let A and B are sets. The **Cartesian product** of A and B is the set

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

In other words, the Cartesian product of A and B is the set of ordered pairs (a, b) with a coming from A and b coming from B. Note that the order is important.

Activity 1.5.

(a) List all of the elements in $\{\text{red}, \text{blue}\} \times \{r, s, t\}$.

(b) If A has m elements and B has n elements, how many elements does the set $A \times B$ have? Explain.

Activity Solution.

(a) The elements are

$$(red, r), (red, s), (red, t), (blue, r), (blue, s), (blue, t).$$

(b) Any element in $A \times B$ pairs an element of A with an element of B. We can choose n elements from A and m from B, so the total number of pairings is nm.

There is no reason to restrict ourselves to a Cartesian product of just two sets. This is an idea that we have encountered before. The Cartesian product $\mathbb{R} \times \mathbb{R}$ is the standard real plane that we denote as $\mathbb{R}^2$ and the Cartesian product $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ is the three-dimensional real space denoted as $\mathbb{R}^3$. If we have an indexed collection $\{X_i\}$ of sets, with i running through the set of positive integers, then we can define the Cartesian product of the sets X_i as the set of infinite sequences $(x_1, x_2, \ldots, x_n, \ldots)$, where $x_i \in X_i$ for each $i \in \mathbb{Z}^+$. We denote this cartesian product as

$$\Pi_{i \in \mathbb{Z}^+} X_i = \Pi_{i=1}^{\infty} X_i.$$

We will study sequences in more detail later in this course.

To conclude this section we summarize some properties of sets. Many of these properties can be extended to arbitrary collections of sets. Statements and proofs are left to the reader.

Theorem. *Let A and B be subsets of a universal set U.*

Properties of the Empty Set.

i. $A \cap \emptyset = \emptyset$

ii. $A \cup \emptyset = A$

iii. $A - \emptyset = A$

iv. $\emptyset^c = U$

Properties of the Universal Set.

i. $A \cap U = A$
ii. $A \cup U = U$
iii. $A - U = \emptyset$
iv. $U^c = \emptyset$

Idempotent Laws.

i. $A \cap A = A$
ii. $A \cup A = A$

Commutative Laws.

i. $A \cap B = B \cap A$
ii. $A \cup B = B \cup A$

Associative Laws.

i. $(A \cap B) \cap C = A \cap (B \cap C)$
ii. $(A \cup B) \cup C = A \cup (B \cup C)$

Distributive Laws.

i. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
ii. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Basic Properties.

i. $(A^c)^c = A$
ii. $A - B = A \cap B^c$

Subsets and Complements.

$A \subseteq B$ *if and only if* $B^c \subseteq A^c$

Summary

Important ideas that we discussed in this section include the following.

- We can consider a set to be a well-defined collection of elements.
- A subset of a set is any collection of elements from that set. That is, a subset S of a set X is a set with the property that if $s \in S$, then $s \in X$.
- If X and Y are sets, then the union $X \cup Y$ is the set

$$X \cup Y = \{z \mid z \in X \text{ or } z \in Y\}.$$

  The union of an arbitrary collection $\{X_\alpha\}$ of sets for α in some indexing set I is the set

$$\bigcup_{\alpha \in I} X_\alpha = \{z \mid z \in X_\beta \text{ for some } \beta \in I\}.$$

- If X and Y are sets, then the intersection $X \cap Y$ is the set

$$X \cap Y = \{z \mid z \in X \text{ and } z \in Y\}.$$

  The intersection of an arbitrary collection $\{X_\alpha\}$ of sets for α in some indexing set I is the set

$$\bigcap_{\alpha \in I} X_\alpha = \{z \mid z \in X_\beta \text{ for all } \beta \in I\}.$$

- If X is a set and A is a subset of X, then the complement of A in X is the set

$$A^c = \{x \in X \mid x \notin A\}.$$

- If $\{X_i\}$ is a collection of sets with i in some indexing set I, where I is finite or I is the set of positive integers, the Cartesian product $\Pi_{i \in I} X_i$ of the sets X_i as the set of all ordered tuples of the form (x_i) where $i \in I$.

Section 2

Functions

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a function?
- What is the domain of a function?
- What is the difference between the range and codomain of a function?
- What does it mean for a function to be an injection? A surjection?
- When and how is the composite of two functions defined?
- When and how is the inverse of a function defined?
- What do we mean by the image and inverse image of a set under a function?
- What properties relate images and inverse images of sets and set unions?

Introduction

Many topological properties are defined using continuous functions. We will focus on continuity later – for now we review some important concepts related to functions. Much of this should be familiar, but some might be new.

First we present the basic definitions.

Definition 2.1. A **function** f from a nonempty set A to a set B is a collection of ordered pairs (a, b) so that

- for each $a \in A$ there is a pair (a, b) in f, and
- if (a, b) and (a, b') are in f, then $b = b'$.

We generally use an alternate notation for a function. If (a, b) is an element of a function f, we write

$$f(a) = b,$$

and in this way we think of f as a mapping from the set A to the set B. We indicate that f is a mapping from set A to set B with the notation

$$f : A \to B.$$

There is some familiar terminology and notation associated with functions. Let f be a function from a set A to a set B.

- The set A is called the **domain** of f, and we write $\text{dom}(f) = A$.
- The set B is called the **codomain** of f, and we write $\text{codom}(f) = B$.
- The subset $\{f(a) \mid a \in A\}$ of B is called the **range** of f, which we denote by $\text{range}(f)$. Note that the range of f could equivalently be defined as follows:

  $$\text{range}(f) = \{f\,(x) \mid x \in A\}.$$

- If $a \in A$, then $f(a)$ is the **image** of a under f.
- If $b \in B$ and $b = f(a)$ for some $a \in A$, then a is called a **pre-image** of b.

One-to-one functions (or injections) and onto functions (or surjections) are special types of functions.

Definition 2.2. Let f be a function from a set A to a set B.

(1) The function f is an **injection** if, whenever (a, b) and (a', b) are in f, then $a = a'$. Alternatively, using the function notation, f is an injection if $f(a) = f(a')$ implies $a = a'$.

(2) The function f is a surjection if, whenever $b \in B$, then there is an $a \in A$ so that (a, b) is in f. Alternatively, using the function notation, f is a surjection if for each $b \in B$ there exists an $a \in A$ so that $f(a) = b$.

(3) The function f is a **bijection** if f is both an injection and a surjection.

Preview Activity 2.1. We often define functions with rules, e.g., $f(x) = x^2$. (Note that f is the function and $f(x)$ is the image of x under f.) The functions in this activity will illustrate why the domain and the codomain are just as important as the rule defining the outputs when we are trying to determine if a given function is injective and/or surjective. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2 + 1$. Notice that

$$f(2) = 5 \text{ and } f(-2) = 5.$$

This observation is enough to prove that the function f is not an injection since we can see that there exist two different inputs that produce the same output.

Since $f(x) = x^2 + 1$, we know that $f(x) \geq 1$ for all $x \in \mathbb{R}$. This implies that the function f is not a surjection. For example, -2 is in the codomain of f and $f(x) \neq -2$ for all x in the domain of f.

(1) We can change the domain of a function so that the function is defined on a subset of the original domain. Such a function is called a restriction.

Definition 2.3. Let f be a function from a set A to a set B and let C be a subset of A. The **restriction** of f to C is the function $F : C \to B$ satisfying

$$F(c) = f(c) \text{ for all } c \in C.$$

A notation used for the restriction is also $F = f \mid_C$. We also call f an *extension* of F.

Let $h = f \mid_{\mathbb{R}^+}$, where $\mathbb{R}^+$ is the set of positive real numbers. So h has the same codomain as f, but a different domain.

(a) Show that h is an injection.

(b) Is h a surjection? Justify your conclusion.

(2) Let $T = \{y \in \mathbb{R} : y \geq 1\}$, and let $F : \mathbb{R} \to T$ be defined by $F(x) = f(x)$. Notice that the function F uses the same formula as the function f and has the same domain as f, but has a different codomain than f.

(a) Explain why F is not an injection.

(b) Is F a surjection? Justify your conclusion.

(3) Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \geq 0\}$. Define $g : \mathbb{R}^* \to T$ by $g(x) = x^2 + 1$.

(a) Prove or disprove: the function g is an injection.

(b) Prove or disprove: the function g is a surjection.

Activity Solution.

(1) Let $h = f \mid_{\mathbb{R}}^{+}$, where $\mathbb{R}^+$ is the set of positive real numbers.

(a) Suppose $h(x) = h(y)$ for some $x, y \in \mathbb{R}^+$. Then $x^2 + 1 = y^2 + 1$ or $x^2 = y^2$. Since x and y are positive, this implies that $x = y$ and h is an injection.

(b) The answer is no. Since $h(x) = x^2 + 1 \geq 1$, there is no input into h that produces the output

(2) Let $T = \{y \in \mathbb{R} : y \geq 1\}$, and let $F : \mathbb{R} \to T$ be defined by $F(x) = f(x)$.

(a) Note that $F(-1) = 2 = F(1)$. Since $-1 \neq 1$ it follows that F is not an injection.

(b) Yes, F is a surjection. to see why, let $y \in T$. Then $y \geq 1$ so $y - 1 \geq 0$. So $x = \sqrt{y-1}$ is a real number and $F(x) = (\sqrt{y-1})^2 + 1 = y$. Therefore, F is a surjection.

(3) Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \geq 0\}$. Define $g : \mathbb{R}^* \to T$ by $g(x) = x^2 + 1$.

(a) Suppose $g(x) = g(y)$ for some $x, y \in \mathbb{R}^*$. Then $x^2 + 1 = y^2 + 1$ or $x^2 = y^2$. Since x and y are nonnegative, this implies that $x = y$ and g is an injection.

(b) Let $y \in T$. Then $y \geq 1$ so $y - 1 \geq 0$. So $x = \sqrt{y-1}$ is a nonnegative real number and $g(x) = (\sqrt{y-1})^2 + 1 = y$. Therefore, g is a surjection.

In our preview activity, the same mathematical formula was used to determine the outputs for the functions. However:

- One of the functions was neither an injection nor a surjection.
- One of the functions was not an injection but was a surjection.
- One of the functions was an injection but was not a surjection.
- One of the functions was both an injection and a surjection.

This illustrates the important fact that whether a function is injective or surjective not only depends on the formula that defines the output of the function but also on the domain and codomain of the function.

Composites of Functions

If we have two functions f and g mapping $\mathbb{R}$ to $\mathbb{R}$, then we can create the sum and product of f and g utilizing the additive and multiplicative structure from $\mathbb{R}$. In topology, we generally don't care about any algebraic structure a set might have, so the only operations on functions that are of concern are composition and inverses. We start with composition.

The basic idea of function composition is that, when possible, the output of a function f is used as the input of a function g. The resulting function can be referred to as "f followed by g" and is called the composite of f with g. The notation we use is $g \circ f$ (note the order – f is applied first). For example, if

$$f(x) = 3x^2 + 2 \text{ and } g(x) = \sin(x),$$

both mapping $\mathbb{R}$ to $\mathbb{R}$, then we can compute $(g \circ f)(x)$ as follows:

$$(g \circ f)(x) = g(f(x)) = g(3x^2 + 2) = \sin\left(3x^2 + 2\right).$$

In this case, $f(x)$, the output of the function f, was used as the input for the function g. This idea motivates the formal definition of the composition of two functions.

Definition 2.4. Let A, B, and C be nonempty sets, and let $f : A \to B$ and $g : B \to C$ be functions. The **composite** of f and g is the function $g \circ f : A \to C$ defined by

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

for all $x \in A$

We often refer to the function $g \circ f$ as a composite function.

Activity 2.1. Let $A = \{1, 2, 3\}$, $B = \{a, b, c, d\}$, and $C = \{\alpha, \beta, \gamma\}$. Define $f : A \to B$, $g : A \to B$, and $h : B \to C$ by

$$f(1) = b,\ f(2) = c,\ f(3) = a,$$

$$g(1) = d,\ g(2) = c,\ g(3) = d, \text{ and}$$

$$h(a) = \gamma,\ h(b) = \alpha,\ h(c) = \beta,\ h(d) = \alpha.$$

(a) Find the images of the elements in A under the function $h \circ f$.

(b) Find the images of the elements in A under the function $h \circ g$.

(c) Is $h \circ f$ an injection? Is $h \circ f$ a surjection? Explain.

(d) Is $h \circ g$ an injection? Is $h \circ g$ a surjection? Explain.

Activity Solution.

(a) Applying the rules for f and h gives us

$$\begin{aligned}(h \circ f)(1) &= h(f(1)) = h(b) = \alpha \\ (h \circ f)(2) &= h(f(2)) = h(c) = \beta \\ (h \circ f)(3) &= h(f(3)) = h(a) = \gamma.\end{aligned}$$

(b) Applying the rules for g and h gives us

$$\begin{aligned}(h \circ g)(1) &= h(g(1)) = h(d) = \alpha \\ (h \circ g)(2) &= h(g(2)) = h(c) = \beta \\ (h \circ g)(3) &= h(g(3)) = h(d) = \alpha.\end{aligned}$$

(c) Since $(h \circ f)(x)$ is always different than $(h \circ f)(y)$ when $x \neq y$, we see that $h \circ f$ is an injection. We can also see by inspection that the images of the elements in A under $h \circ f$ produce all of the elements of C, so $h \circ f$ is a surjection.

(d) Since $(h \circ g)(1) = \alpha = (h \circ g)(3)$, we see that $h \circ g$ is not an injection. It is also the case there there is no input to $h \circ g$ that produces the output γ, so $h \circ g$ is not a surjection.

In Activity 2.1, we asked questions about whether certain composite functions were injections and/or surjections. In mathematics, it is typical to explore whether certain properties of an object transfer to related objects. In particular, we might want to know whether or not the composite of two injective functions is also an injection. (Of course, we could ask a similar question for surjections.) These questions are explored in the next activity.

Activity 2.2. Let the sets A, B, C, and D be as follows:

$$A = \{a, b, c\}, \quad B = \{p, q, r\}, \quad C = \{u, v, w, x\}, \quad \text{and} \quad D = \{u, v\}.$$

(a) Construct a function $f : A \to B$ that is an injection and a function $g : B \to C$ that is an injection. In this case, is the composite function $g \circ f : A \to C$ an injection? Explain.

(b) Construct a function $f : A \to B$ that is a surjection and a function $g : B \to D$ that is a surjection. In this case, is the composite function $g \circ f : A \to D$ a surjection? Explain.

(c) Construct a function $f : A \to B$ that is a bijection and a function $g : B \to A$ that is a bijection. In this case, is the composite function $g \circ f : A \to A$ a bijection? Explain.

Activity Solution.

(a) Define f and g by

$$f(a) = p,\ f(b) = q,\ f(c) = r \text{ and } g(p) = u,\ g(q) = v,\ g(r) = w.$$

So both f and g are injections. Notice that

$$(g \circ f)(a) = u,\ (g \circ f)(b) = v,\ \text{ and } (g \circ f)(c) = w,$$

so $g \circ f : A \to C$ is also an injection.

(b) Define f and g by

$$f(a) = p,\ f(b) = q,\ f(c) = r \text{ and } g(p) = u,\ g(q) = v,\ g(r) = u.$$

So both f and g are surjections. Notice that

$$(g \circ f)(a) = u,\ (g \circ f)(b) = v,\ \text{ and } (g \circ f)(c) = u,$$

so $g \circ f : A \to D$ is also a surjection.

(c) Define f and g by

$$f(a) = p,\ f(b) = q,\ f(c) = r \text{ and } g(p) = c,\ g(q) = b,\ g(r) = a.$$

So both f and g are bijections. Notice that

$$(g \circ f)(a) = c,\ (g \circ f)(b) = b,\ \text{ and } (g \circ f)(c) = a,$$

so $g \circ f : A \to A$ is also a bijection.

In Activity 2.2, we explored some properties of composite functions related to injections, surjections, and bijections. The following theorem summarizes the results that these explorations were intended to illustrate.

Theorem 2.5. *Let A, B, and C be nonempty sets, and assume that $f : A \to B$ and $g : B \to C$.*

(1) If f and g are both injections, then $(g \circ f) : A \to C$ is an injection.

(2) If f and g are both surjections, then $(g \circ f) : A \to C$ is a surjection.

(3) If f and g are both bijections, then $(g \circ f) : A \to C$ is a bijection.

Activity 2.3.

(a) Prove part (1) of Theorem 2.5.

(b) Prove part (2) of Theorem 2.5.

(c) Why is the proof of part (3) of Theorem 2.5 a direct consequence of parts (1) and (2)?

Activity Solution.

(a) Let A, B, and C be nonempty sets, and assume that $f : A \to B$ and $g : B \to C$ are both injections. We will prove that $g \circ f : A \to C$ is an injection.

Suppose $(g \circ f)(x) = (g \circ f)(y)$ for some x and y in A. Then $g(f(x)) = g(f(y))$. The fact that g is an injection means that $f(x) = f(y)$. Now the fact that f is an injection implies that $x = y$. Thus, $g \circ f$ is an injection.

(b) Let A, B, and C be nonempty sets, and assume that $f : A \to B$ and $g : B \to C$ are both surjections. We will prove that $g \circ f : A \to C$ is a surjection.

Let c be an arbitrary element of C. We will prove there exists an $a \in A$ such that $(g \circ f)(a) = c$. Since $g : B \to C$ is a surjection, it follows that there exists a $b \in B$ such that $g(b) = c$. Now $b \in B$ and $f : A \to B$ is a surjection. Hence, there exists an $a \in A$ such that $f(a) = b$. We now see that

$$\begin{aligned}(g \circ f)(a) &= g\left(f(a)\right) \\ &= g(b) \\ &= c.\end{aligned}$$

We have therefore shown that for every $c \in C$, there exists an $a \in A$ such that $(g \circ f)(a) = c$. This proves that $g \circ f$ is a surjection.

(c) Suppose that f and g are both bijections. The fact that f and g are both injections implies that $g \circ f$ is an injection by part (1). The fact that f and g are both surjections implies that $g \circ f$ is a surjection by part (2). So $g \circ f$ is a bijection.

Inverse Functions

Now that we have studied composite functions, we will move on to consider another important idea: the inverse of a function. In previous mathematics courses, you probably learned that the exponential function (with base e) and the natural logarithm functions are inverses of each other. You may have seen this relationship expressed as follows:

$$\text{For each } x \in \mathbb{R} \text{ with } x > 0 \text{ and for each } y \in \mathbb{R}, \\ y = \ln(x) \text{ if and only if } x = e^y.$$

Notice that x is the input and y is the output for the natural logarithm function if and only if y is the input and x is the output for the exponential function. In essence, the inverse function (in this case, the exponential function) reverses the action of the original function (in this case, the natural logarithm function). In terms of ordered pairs (input-output pairs), this means that if (x, y) is an ordered pair for a function, then (y, x) is an ordered pair for its inverse. The idea of reversing the roles of the first and second coordinates is the basis for our definition of the inverse of a function.

Definition 2.6. Let $f : A \to B$ be a function. The **inverse** of f, denoted by f^{-1}, is the set of ordered pairs

$$f^{-1} = \{(b, a) \in B \times A : (a, b) \in f\}.$$

Notice that this definition does not state that f^{-1} is a function. Rather, f^{-1} is simply a subset of $B \times A$. In Activity 2.4, we will explore the conditions under which the inverse of a function $f : A \to B$ is itself a function from B to A.

Activity 2.4. Let $A = \{a, b, c\}$, $B = \{a, b, c, d\}$, and $C = \{p, q, r\}$. Define

$f : A \to C$ by	$g : A \to C$ by	$h : B \to C$ by
$f(a) = r$	$g(a) = p$	$h(a) = p$
$f(b) = p$	$g(b) = q$	$h(b) = q$
$f(c) = q$	$g(c) = p$	$h(c) = r$
		$h(d) = q$

(a) Determine the inverse of each function as a set of ordered pairs.

(b) i. Is f^{-1} a function from C to A? Explain.
  ii. Is g^{-1} a function from C to A? Explain.
  iii. Is h^{-1} a function from C to B? Explain.

(c) Make a conjecture about what conditions on a function $F : S \to T$ will ensure that its inverse is a function from T to S.

Activity Solution.

(a) We reverse each ordered pair to obtain the ordered pairs for the inverse. So

$$\begin{aligned} f^{-1} &= \{(r, a), (p, b), (q, c)\}, \\ g^{-1} &= \{(p, a), (q, b), (p, c)\}, \\ h^{-1} &= \{(p, a), (q, b), (r, c), (q, d)\}. \end{aligned}$$

(b) i. Since f^{-1} contains no ordered pairs with the same first coordinate, f^{-1} defines a function.
  ii. The fact that (p, a) and (p, c) are in g^{-1} means that g^{-1} is not a function.
  iii. Since (q, b) and (q, d) are in h^{-1}, h^{-1} is not a function.Explain.

(c) Suppose f is a function from A to B. For f^{-1} to define a function, there can be no ordered pairs of the form (a, b) and (c, b) in f. That is, if (x, y) and (w, y) are in f, then $x = w$. If f^{-1} is to be a function from B to A, then for each $b \in B$ there must exist a pair $(a, b) \in f$. In other words, for f^{-1} to be a function from B to A, then f has to be both an injection and a surjection.

The result of the Activity 2.4 should have been the following theorem.

Theorem 2.7. *Let A and B be nonempty sets, and let $f : A \to B$. The inverse of f is a function from B to A if and only if f is a bijection.*

The proof of Theorem 2.7 is outlined in the following activity.

Activity 2.5. Theorem 2.7 is a biconditional statement, so we need to prove both directions. Let A and B be nonempty sets, and let $f : A \to B$.

(a) Assume that f is a bijection. We will prove that f^{-1} is a function, that is that f^{-1} satisfies the conditions of Definition 2.1.

i. Let $b \in B$. What property does f have that ensures that $(b, a) \in f^{-1}$ for some $a \in A$? What conclusion can we draw about f^{-1}?

ii. Now let $b \in B$, $a_1, a_2 \in A$ and assume that

$$(b, a_1) \in f^{-1} \text{ and } (b, a_2) \in f^{-1}.$$

What does this tell us about elements that must be in f? What property of f ensures that $a_1 = a_2$? What conclusion can we draw about f^{-1}?

(b) Now assume that f^{-1} is a function from B to A. We will prove that f is a bijection.

i. What does it take to prove that f is an injection? Use the fact that f^{-1} is a function to prove that f is an injection.

ii. What does it take to prove that f is a surjection? Use the fact that f^{-1} is a function to prove that f is a surjection.

Activity Solution.

(a) i. Since the f is a surjection, there exists an $a \in A$ such that $f(a) = b$. This implies that $(a, b) \in f$ and hence that $(b, a) \in f^{-1}$. Thus, each element of B is the first coordinate of an ordered pair in f^{-1}.

ii. We must now prove that each element of B is the first coordinate of exactly one ordered pair in f^{-1}. So let $b \in B$, $a_1, a_2 \in A$ and assume that

$$(b, a_1) \in f^{-1} \text{ and } (b, a_2) \in f^{-1}.$$

This means that $(a_1, b) \in f$ and $(a_2, b) \in f$. Since f is a bijection, f is by definition an injection, and we can conclude that $a_1 = a_2$. This proves that b is the first element of only one ordered pair in f^{-1}. Consequently, we have proved that f^{-1} satisfies the conditions of Definition 2.1 and hence f^{-1} is a function from B to A.

(b) i. To prove that f is an injection, we will assume that $a_1, a_2 \in A$ and that $f(a_1) = f(a_2)$, and we must show that $a_1 = a_2$. If we let $b = f(a_1) = f(a_2)$, we can conclude that

$$(a_1, b) \in f \text{ and } (a_2, b) \in f.$$

But this means that

$$(b, a_1) \in f^{-1} \text{ and } (b, a_2) \in f^{-1}.$$

Since we have assumed that f^{-1} is a function, we can conclude that $a_1 = a_2$. Hence, f is an injection.

ii. Now to prove that f is a surjection, we will choose an arbitrary $b \in B$ and show that there exists an $a \in A$ such that $f(a) = b$. Since f^{-1} is a function, b must be the first coordinate of some ordered pair in f^{-1}. Consequently, there exists an $a \in A$ such that

$$(b, a) \in f^{-1}.$$

Now this implies that $(a, b) \in f$, and so $f(a) = b$. This proves that f is a surjection. Since we have also proved that f is an injection, we can conclude that f is a bijection, as desired.

In the situation where $f : A \to B$ is a bijection and f^{-1} is a function from B to A, we can write $f^{-1} : B \to A$. In this case, we frequently say that f is an **invertible function**, and we usually do not use the ordered pair representation for either f or f^{-1}. Instead of writing $(a, b) \in f$, we write $f(a) = b$, and instead of writing $(b, a) \in f^{-1}$, we write $f^{-1}(b) = a$. Using the fact that $(a, b) \in f$ if and only if $(b, a) \in f^{-1}$, we can now write $f(a) = b$ if and only if $f^{-1}(b) = a$. Theorem 2.8 formalizes this observation.

Theorem 2.8. *Let A and B be nonempty sets, and let $f : A \to B$ be a bijection. Then $f^{-1} : B \to A$ is a function, and for every $a \in A$ and $b \in B$,*

$$f(a) = b \text{ if and only if } f^{-1}(b) = a.$$

The next two results are two important theorems about inverse functions. The first can be considered to be a corollary of Theorem 2.8. The proofs are left to the reader.

Corollary 2.9. *Let A and B be nonempty sets, and let $f : A \to B$ be a bijection. Then*

(1) For every x in A, $\left(f^{-1} \circ f\right)(x) = x$.

(2) For every y in B, $\left(f \circ f^{-1}\right)(y) = y$.

The next question to address is what we can say about a composition of bijections. In particular, if $f : A \to B$ and $g : B \to C$ are both bijections, then $f^{-1} : B \to A$ and $g^{-1} : C \to B$ are both functions. Must it be the case that $g \circ f$ is invertible and, if so, what is $(g \circ f)^{-1}$?

Activity 2.6. Let $f : A \to B$ and $g : B \to C$ both be bijections.

(a) Why do we know that $g \circ f$ is invertible?

(b) Now we determine the inverse of $g \circ f$. We might be tempted to think that $(g \circ f)^{-1}$ is $g^{-1} \circ f^{-1}$, but this composite is not defined because g^{-1} maps B to C and f^{-1} maps B to A. However, $f^{1} \circ g^{-1}$ is defined. To prove that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$, we need to prove that two functions are equal. How do we prove that two functions are equal?

(c) Suppose $c \in C$.

i. What tells us that there is a $b \in B$ so that $g(b) = c$?

ii. What tells us that there is an $a \in A$ so that $f(a) = b$?

iii. What element is $(g \circ f)^{-1}(c)$? Why?

iv. What element is $f^{-1}(b)$? Why? What element is $g^{-1}(c)$? Why?

v. What element is $(f^{-1} \circ g^{-1})(c)$? Why? What can we conclude about $(g \circ f)^{-1}$ and $f^{-1} \circ g^{-1}$? Explain.

Activity Solution.

(a) Theorem 2.5 tells us that $g \circ f : A \to C$ is a bijection, and hence invertible. Note that $(g \circ f)^{-1} : C \to A$.

(b) To prove that $g \circ f)^{-1} = (f^{-1} \circ g^{-1})$, we need to show that

$$(g \circ f)^{-1}(c) = (f^{-1} \circ g^{-1})(c)$$

for every $c \in C$.

i. Let $c \in C$. Since the function g is a surjection, there exists a $b \in B$ such that $g(b) = c$.

ii. Since f is a surjection, there exists an $a \in A$ such that $f(a) = b$.

iii. We then have $(g \circ f)(a) = g(f(a)) = g(b) = c$, so $a = (g \circ f)^{-1}(c)$.

iv. By definition, $f^{-1}(b) = a$ and $g^{-1}(c) = b$.

v. It follows that

$$\left(f^{-1} \circ g^{-1}\right)(c) = f^{-1}\left(g^{-1}(c)\right) = f^{-1}b) = a.$$

vi. Since

$$(g \circ f)^{-1}(c) = a = \left(f^{-1} \circ g^{-1}\right)(c),$$

we conclude that $(g \circ f)^{-1}(c) = \left(f^{-1} \circ g^{-1}\right)(c)$ for every $c \in C$. Consequently, $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

The result of Activity 2.6 is contained in the next theorem.

Theorem 2.10. *Let $f : A \to B$ and $g : B \to C$ be bijections. Then $g \circ f$ is a bijection and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Functions and Sets

We conclude this section with a connection between subsets and functions. A bit of notation first. If f is a function from a set X to a set Y, and if B is a subset of Y, we define $f^{-1}(B)$ as

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

So $f^{-1}(B)$ is the set of the elements in X that map to B. When we work with continuous functions in later sections, we will need to understand how a function behaves with respect to subsets. One result is in the following lemma.

Lemma 2.11. *Let $f : X \to Y$ be a function and let $\{A_\alpha\}$ be a collection of subsets of X for α in some indexing set I, and $\{B_\beta\}$ be a collection of subsets of Y for β in some indexing set J. Then*

(1) $f\left(\bigcup_{\alpha\in I} A_\alpha\right) = \bigcup_{\alpha\in I} f(A_\alpha)$ *and*

(2) $f^{-1}\left(\bigcup_{\beta\in J} B_\beta\right) = \bigcup_{\beta\in J} f^{-1}(B_\beta)$.

Proof. Let $f : X \to Y$ be a function and let $\{A_\alpha\}$ be a collection of subsets of X for α in some indexing set I. To prove part 1, we demonstrate the containment in both directions.

Let $b \in f\left(\bigcup_{\alpha\in I} A_\alpha\right)$. Then $b = f(a)$ for some $a \in \bigcup_{\alpha\in I} A_\alpha$. It follows that $a \in A_\rho$ for some $\rho \in I$. Thus, $b \in f(A_\rho) \subseteq \bigcup_{\alpha\in I} f(A_\alpha)$. We conclude that $f\left(\bigcup_{\alpha\in I} A_\alpha\right) \subseteq \bigcup_{\alpha\in I} f(A_\alpha)$.

Now let $b \in \bigcup_{\alpha\in I} f(A_\alpha)$. Then $b \in f(A_\rho)$ for some $\rho \in I$. Since $A_\rho \subseteq \bigcup_{\alpha\in I} A_\alpha$, it follows that $b \in f\left(\bigcup_{\alpha\in I} A_\alpha\right)$. Thus, $\bigcup_{\alpha\in I} f(A_\alpha) \subseteq f\left(\bigcup_{\alpha\in I} A_\alpha\right)$. The two containments prove part 1.

For part 2, we again demonstrate the containments in both directions. Let $a \in f^{-1}\left(\bigcup_{\beta\in J} B_\beta\right)$. Then $f(a) \in \bigcup_{\beta\in J} B_\beta$. So there exists $\mu \in J$ such that $f(a) \in B_\mu$. This implies that $a \in f^{-1}(B_\mu) \subseteq \bigcup_{\beta\in J} f^{-1}(B_\beta)$. We conclude that $f^{-1}\left(\bigcup_{\beta\in J} B_\beta\right) \subseteq \bigcup_{\beta\in J} f^{-1}(B_\beta)$.

For the reverse containment, let $a \in \bigcup_{\beta\in J} f^{-1}(B_\beta)$. Then $a \in f^{-1}(B_\mu)$ for some $\mu \in J$. Thus, $f(a) \in B_\mu \subseteq \bigcup_{\beta\in J} B_\beta$. So $a \in f^{-1}\left(\bigcup_{\beta\in J} B_\beta\right)$. Thus, $\bigcup_{\beta\in J} f^{-1}(B_\beta) \subseteq f^{-1}\left(\bigcup_{\beta\in J} B_\beta\right)$. The two containments verify part 2. ∎

At this point it is reasonable to ask if Lemma 2.11 would still hold if we replace unions with intersections. We leave that question to the reader.

Another result is contained in the next activity.

Activity 2.7. Let X, Y, and Z be sets, and let $f : X \to Y$ and $g : Y \to Z$ be functions. Let C be a subset of Z. There is a relationship between $(g \circ f)^{-1}(C)$ and $f^{-1}(g^{-1}(C))$. Find and prove this relationship.

Activity Solution. We will show that $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$. Let $x \in (g \circ f)^{-1}(C)$. Then $g \circ f)(x) \in C$. So $g(f(x)) \in C$ and $f(x) \in g^{-1}(C)$. From this it follows that $x \in f^{-1}(g^{-1}(C))$. So $(g \circ f)^{-1}(C) \subseteq f^{-1}(g^{-1}(C))$.

Now assume that $x \in f^{-1}(g^{-1}(C))$. Then $f(x) \in g^{-1}(C)$. We then have $g(f(x)) \in C$ or $(g \circ f)(x) \in C$. Thus, $x \in (g \circ f)^{-1}(C)$ and $f^{-1}(g^{-1}(C)) \subseteq (g \circ f)^{-1}(C)$. The two containments demonstrate that $f^{-1}(g^{-1}(C)) = (g \circ f)^{-1}(C)$.

The Cardinality of a Set

How big is a set? When a set is finite, we can count the number of elements in the set and answer the question directly. When a set is infinite, the question is a little more complicated. For example, how big is $\mathbb{Z}$? How big is $\mathbb{Q}$? Since $\mathbb{Z}$ is a subset of $\mathbb{Q}$, we might think that $\mathbb{Q}$ contains more elements than $\mathbb{Z}$. But $\mathbb{Z}$ is infinite and how many more elements can we have than infinity? We won't answer that question in this section, but it is an interesting one to consider.

If two finite sets have the same number of elements, then it should seem natural to say that the sets are of the same size. How do we extend this to infinite sets? If two finite sets have the same number of elements, then we can pair each element in one set with exactly one element in the other. This is exactly what a bijection does. So a set with n elements can be paired with the set $\{1, 2, \ldots, n\}$, where n is a positive integer. This is how we can define a finite set.

Definition 2.12. A set A is a **finite** set if $A = \emptyset$ or there is a bijection f mapping A to the set $\{1, 2, 3, \ldots, n\}$ for some positive integer n.

In the case that $A = \emptyset$, we say that A has *cardinality* 0, and if there is a bijection from A to the set $\{1, 2, \ldots, n\}$, we say that A has cardinality n. If there is no positive integer n such that there is a bijection from set A to $\{1, 2, \ldots, n\}$ we say that A is an *infinite* se and say that A has infinite cardinality. We use the word *cardinality* instead of number of elements because we can't actually count the number of elements in an infinite set. We denote the cardinality of the set (the number of elements in the set) A by $|A|$. It is left to the homework to show that if A and B are sets with $|A| = n$ and $|B| = m$, then $n = m$ if and only if there is a bijection $f : A \to B$. This tells us that cardinality is well defined. Since composites of bijections are bijections with inverses that are bijections, if there is a bijection from set A to $\{1, 2, \ldots, n\}$ and a bijection from a set B to $\{1, 2, \ldots, n\}$ for some positive integer n, then there is a bijection between A and B. Using this idea, we say that two sets (either finite or infinite) have the same cardinality if there is a bijection between the sets. We will discuss cardinality in more detail a bit later.

Summary

Important ideas that we discussed in this section include the following.

- A function f from a nonempty set A to a set B is a collection of ordered pairs (a, b) so that for each $a \in A$ there is a pair (a, b) in f, and if (a, b) and (a, b') are in f, then $b = b'$. If f is a function we use the notation $f(a) = b$ to indicate that $(a, b) \in f$.
- If f is a function from A to B, the set A is the domain of the function.
- If f is a function from A to B, the set B is the codomain of the function. The set
$$\{f(a) \mid a \in A\}$$
is the range of the function. So the range of a function is a subset of the codomain.
- A function f from a set A to a set B is an injection if, whenever $f(a) = f(a')$ for $a, a' \in A$, then $a = a'$. The function f is a surjection if, whenever $b \in B$, then there is an $a \in A$ so that $f(a) = b$.
- If f is a function from a set A to a set B and if g is a function from B to a set C, then the composite $g \circ f$ is a function from A to C defined by $(g \circ f)(a) = g(f(a))$ for every $a \in A$.
- A function f from a set A to a set B is a bijection if f is both a surjection and injection. When f is a bijection from A to B, then f has an inverse f^{-1} defined by $f^{-1}(b) = a$ when $f(a) = b$.
- If f is a function from a set A to a set B, and if C is a subset of A, then image of C under f is the set
$$f(C) = \{f(c) \mid c \in C\},$$
and if D is a subset of Y, the inverse image of D is the set
$$f^{-1}(D) = \{a \in A \mid f(a) \in D\}.$$

- Important properties that relate images and inverse images of sets and set unions are the following. If f is a function from a set X to a set Y, and if $\{A_\alpha\}$ is a collection of subsets of X for α in some indexing set I, and $\{B_\beta\}$ be a collection of subsets of Y for β in some indexing set J, then

  i. $f\left(\bigcup_{\alpha\in I} A_\alpha\right) = \bigcup_{\alpha\in I} f(A_\alpha)$ and
  ii. $f^{-1}\left(\bigcup_{\beta\in J} B_\beta\right) = \bigcup_{\beta\in J} f^{-1}(B_\beta)$.

Part II

Metric Spaces

Section 3

Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a metric and what is a metric space?
- How are the Euclidean, taxicab, and max metric different and how are they similar?

Introduction

Metric spaces are particular examples of topological spaces. A metric space is a space that has a metric defined on it. A metric is a function that measures the distance between points in a metric space.

We are familiar with one special metric, the Euclidean metric d_E in $\mathbb{R}^2$ where

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

Using this metric, the distance between two points (x_1, x_2) and (y_1, y_2) is the length of the segment connecting the points, while the unit circle (the set of points a distance 1 from the origin) looks like what we think of as a circe as illustrated in Figure 3.1.

As we will see, there are many other metrics that can be defined on $\mathbb{R}^n$, or on other sets.

Preview Activity 3.1. Consider the function d_T that assigns to each pair of points in $\mathbb{R}^2$ the real number

$$d_T((x_1, x_2), (y_1, y_2)) = |x_1 - y_1| + |x_2 - y_2|.$$

Any distance function should satisfy certain properties: the distance between two points should never be negative, the distance from point A to point B should be the same as the distance from point B to point A, the shortest distance between two points A and B should never be more than the distance from A to some point C plus the distance from C to B, and the distance between

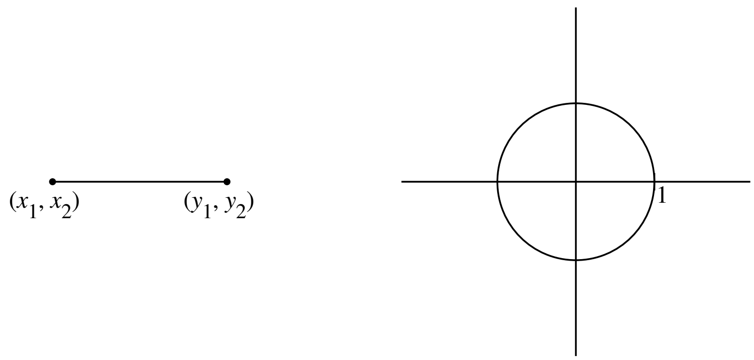


Figure 3.1: The Euclidean distance between (x_1, x_2) and (y_1, y_2) and the Euclidean unit circle in $\mathbb{R}^2$.

points should only be zero if the points are the same. In this activity, we determine if d_T has these properties. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in $\mathbb{R}^2$.

(1) Prove or disprove: $d_T(x, y) \geq 0$.

(2) Prove or disprove: $d_T(x, y) = d_T(y, x)$.

(3) Prove or disprove: $d_T(x, y) = 0$ if and only if $x = y$.

(4) Let $z = (z_1, z_2)$ in $\mathbb{R}^2$. Read the proof of Lemma 3.1 (below) and then use Lemma 3.1 to show that
$$d_T(x, y) \leq d_T(x, z) + d_T(z, y).$$
(Do you have any questions about the proof of the lemma?)

Lemma 3.1. *Let a and b be real numbers. Then*
$$|a + b| \leq |a| + |b|.$$

Proof. Let a and b be real numbers. To prove the lemma we consider cases.

Case 1: $a \geq 0$ and $b \geq 0$. In this case $a + b$ is nonnegative and so $|a| = a$, $|b| = b$, and $|a + b| = a + b$. Then
$$|a + b| = a + b = |a| + |b|.$$

Case 2: $a \leq 0$ and $b \leq 0$. In this case $a = -a'$ and $b = -b'$ where a' and b' are nonnegative. It follows from Case 1 that
$$|a + b| = | - (a' + b')| = |a' + b'| = a' + b' = |a'| + |b'| = | - a'| + | - b'| = |a| + |b|.$$

Case 3: One of a or b is positive and the other negative. Without loss of generality we assume $a > 0$ and $b < 0$. Again we consider cases. Note that $b < 0$ implies $a + b < a$.

- Suppose $b \geq -2a$. Then $a + b \geq -a$ and so $-a \leq a + b < a$. It follows that
$$|a + b| \leq a = |a| < |a| + |b|.$$

- The last case is when $b < -2a$. In this case $-b > 2a$ and so

$$|b| = -b > 2a = 2|a| > |a|.$$

Then $a + b < a = |a| < |b|$. Finally, $a > 0$ implies $a + b > b = -|b|$. So

$$-|b| < a + b < |b|$$

and

$$|a + b| \leq |b| < |a| + |b|.$$

This proves our lemma for every possible pair a, b. ■

(5) A picture to illustrate the distance d_T between (points x_1, x_2) and (y_1, y_2) is shown in Figure 3.2. The metric d_T is sometimes called the *taxicab metric* because the distance between points x and y can be thought of as obtained by driving around a city block rather than going directly from point x to point y. Draw a picture of the unit circle using the Taxicab metric. Explain your reasoning.

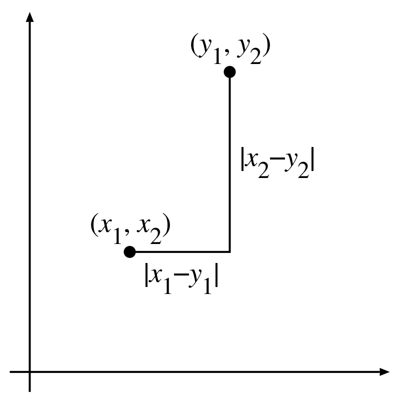


Figure 3.2: The taxicab distance between (x_1, x_2) and (y_1, y_2) in $\mathbb{R}^2$.

Activity Solution.

(1) Since $|a| \geq 0$ for any $a \in \mathbb{R}$, we have that $|x_1 - y_1| \geq 0$ and $|x_2 - y_2| \geq 0$. Thus,

$$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| \geq 0.$$

(2) Since $|-1| = 1$, we have that

$$\begin{aligned} d_T(x, y) &= |x_1 - y_1| + |x_2 - y_2| \\ &= |(-1)(y_1 - x_1)| + |(-1)(y_2 - x_2)| \\ &= |(-1)||y_1 - x_1| + |(-1)||y_2 - x_2| \\ &= |y_1 - x_1| + |y_2 - x_2| \\ &= d_T(y, x). \end{aligned}$$

(3) Suppose
$$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| = 0.$$
Since $|a| \geq 0$ for every real number a, it follows that
$$|x_1 - y_1| = 0 = |x_2 - y_2|.$$
Therefore, $x_1 = y_1$ and $x_2 = y_2$, which makes $x = y$.

Now suppose $x = y$. Then $x_1 = y_1$ and $x_2 = y_2$ and
$$|x_1 - y_1| = 0 = |x_2 - y_2|.$$
it follows that
$$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| = 0 + 0 = 0.$$

(4) First note that
$$\begin{aligned} d(x, z) + d(z, y) &= (|x_1 - z_1| + |x_2 - z_2|) + (|z_1 - y_1| + |z_2 - y_2|) \\ &= (|x_1 - z_1| + |z_1 - y_1|) + (|x_2 - z_2| + |z_2 - y_2|) . \end{aligned}$$
Applying Lemma 3.1 give us
$$\begin{aligned} d(x, z) + d(z, y) &= (|x_1 - z_1| + |z_1 - y_1|) + (|x_2 - z_2| + |z_2 - y_2|) \\ &\geq (|(x_1 - z_1) + (z_1 - y_1)|) + (|(x_2 - z_2) + (z_2 - y_2)|) \\ &= |x_1 - y_1| + |x_2 - y_2| \\ &= d(x, y). \end{aligned}$$

(5) A point (x_1, x_2) will be a distance 1 from the origin if $|x_1| + |x_2| = 1$. This will only happen if
$$|x_2| = 1 - |x_1| \text{ and } -1 \leq x_1, x_2 \leq 1.$$
In other words, the point (x_1, x_2) must lie on one of the lines
$$\begin{aligned} &x_2 = 1 - x_1 \text{ if } x_1, x_2 \geq 0 \\ &x_2 = 1 + x_1 \text{ if } x_1 < 0 \text{ and } x_2 \geq 0 \\ &x_2 = -1 + x_1 \text{ if } x_1 \geq 0 \text{ and } x_2 < 0 \\ &x_2 = -1 - x_1 \text{ if } x_1, x_2 \leq 0. \end{aligned}$$
So the unit circle using the metric d_T in $\mathbb{R}^2$ looks as depicted in Figure 3.3.

Metric Spaces

For most of our mathematical careers our mathematics has taken place in $\mathbb{R}^2$, where we measure the distance between points (x_1, x_2) and (y_1, y_2) with the standard Euclidean distance d_E. In our preview activity we saw that the function d_T satisfies many of the same properties as d_E. These properties allow us to use d_E or d_T as distance functions. We call any distance function a *metric*, and any space on which a metric is defined is called a *metric space*.

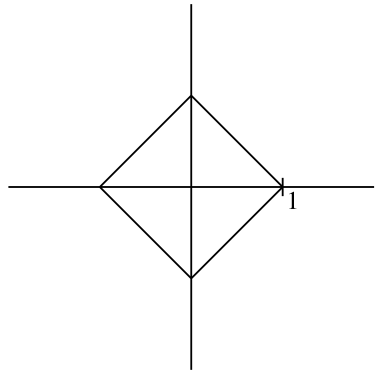


Figure 3.3: The unit circle in $\mathbb{R}^2$ using the taxicab metric.

Definition 3.2. A **metric** on a space X is a function $d : X \times X \to \mathbb{R}^+ \cup \{0\}$ that satisfies the properties:

(1) $d(x, y) \geq 0$ for all $x, y \in X$,

(2) $d(x, y) = 0$ if and only if $x = y$ in X,

(3) $d(x, y) = d(y, x)$ for all $x, y \in X$, and

(4) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

Properties 1 and 2 of a metric say that a metric is *positive definite*, while property 3 states that a metric is *symmetric*. Property 4 of the definition is usually the most difficult property to verify for a metric and is called the *triangle inequality*.

Definition 3.3. A **metric space** is a pair (X, d), where d is a metric on the space X.

When the metric is clear from the context, we just refer to X as the metric space.

Activity 3.1. For each of the following, determine if (X, d) is a metric space. If (X, d) is a metric space, explain why. If (X, d) is not a metric space, determine which properties of a metric d satisfies and which it does not. If (X, d) is a metric space, give a geometric description of the unit circle (the set of all points in X a distance 1 from the zero element) in the space.

(a) $X = \mathbb{R}$, $d(x, y) = \max\{|x|, |y|\}$.

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

(c) $X = \mathbb{R}^2$, $d((x_1, x_2), (y_1, y_2)) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$

(d) $X = C[0, 1]$, the set of all continuous functions on the interval $[0, 1]$,

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

Activity Solution.

(a) This function d is not a metric on $\mathbb{R}$. It is the case by definition that $d(x, y) \geq 0$ for all $x, y \in \mathbb{R}*$. If $x \neq y$, then at least one of x, y is not 0 and $d(x, y) > 0$. Also, $d(1, 1) = 1$, not 0, so d does not satisfy the second property of a metric. This d is symmetric, and also satisfies the triangle inequality. To see why the triangle inequality is satisfied, let $x, y, z \in \mathbb{R}^*$. We look at two cases.

- If $|x| \geq |z|$, then
$$d(x, z) = |x| \text{ and } d(x, y) + d(y, z) \geq |x| + d(y, z) \geq |x| = d(x, z).$$
- If $|x| < |z|$, then
$$d(x, z) = |z| \text{ and } d(x, y) + d(y, z) \geq d(x, y) + |z| \geq |z| = d(x, z).$$

In either case,
$$d(x, z) \leq d(x, y) + d(y, z)$$
and d satisfies the triangle inequality.

(b) This d does define a metric on X. By definition, $d(x, y) \geq 0$ for all $x, y \in X$. If $x \neq y$, then $y \neq x$, so d is symmetric. By definition, $d(x, y) = 0$ if and only if $x = y$. Let $x, y, z \in X$. Then $d(x, y) + d(x, z) = 0$ if and only if $x = y = z$. In this case
$$0 = d(x, z) = 0 + 0 = d(x, y) = d(y, z).$$
If x, y, and z are not all equal, then
$$d(x, y) + d(y, z) \geq 1 \geq d(x, z).$$
So d satisfies the triangle inequality and d is a metric. Note that this argument did not depend on the elements of X, so this d defines a metric on any set. It is interesting to notice that the unit circle in this metric is all of $\mathbb{R}$ except for the origin.

(c) Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Now $|a| \geq 0$ for any real number a, so $d(x, y) \geq 0$. That d is symmetric follows from the fact that $| - a| = |a|$ for any real number a:
$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\} = \max\{|y_1 - x_1|, |y_2 - x_2|\} = d(y, x).$$
If $x = y$, then
$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\} = \max\{0, 0\} = 0.$$
If $d(x, y) = 0$, then
$$0 = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$
But if $x_1 \neq y_1$ or $x_2 \neq y_2$, then $|x_1 - y_1| > 0$ or $|x_2 - y_2| > 0$, forcing $d(x, y) > 0$. So if $d(x, y) = 0$, then $x = y$.

Now we consider the triangle inequality. Let $x = (x_1, x_2)$, $y = (y_1, y_2)$, $z = (z_1, z_2)$ in $\mathbb{R}^2$. It is the case that either $|x_1 - y_1| \geq |x_2 - y_2|$ or $|x_1 - y_1| \leq |x_2 - y_2|$. Without loss of generality, assume $|x_1 - y_1| \geq |x_2 - y_2|$. We know that $d(x, z) = \max\{|x_1 - z_1|, |x_2 - z_2|\}$. Assume that $d(x, z) = |x_1 - z_1|$. We consider cases on $d(y, z)$.

- Suppose $d(y, z) = |y_1 - z_1| = |z_1 - y_1|$. Then

$$d(x, z)+d(z, y) = |x_1-z_1|+|z_1-y_1| \geq |(x_1-z_1)+(z_1-y_1)| = |x_1-y_1| = d(x, y).$$

- Suppose $d(y, z) = |y_2 - z_2|$. Then $|y_2 - z_2| \geq |y_1 - z_1|$. So

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

The case where $d(x, z) = |x_2 - z_2|$ is similar. Thus, d is a metric on X. This metric is called the *max* metric. In this case, the unit circle is the set of points (x_1, x_2) such that $\max\{|x_1|, |x_2|\} = 1$. This is the set of points (x_1, x_2) where $x_1 = 1$ and $-1 \leq x_2 \leq 1$ or $x_2 = 1$ and $-1 \leq x_1 \leq 1$. In other words, the unit circle in this space is the square centered at the origin of side length 1.

(d) We show that d is a metric. Let f, g be in $C[0, 1]$. If we integrate a nonnegative function over any interval, the resulting area is never negative. So $d(f, g) \geq 0$. Since $|a-b| = |b-a|$ for any real numbers a and b, it follows that d is symmetric. From calculus we know that

$$d(f, f) = \int_0^1 |f(x) - f(x)|\, dx = \int_0^1 0\, dx = 0.$$

Now suppose that $d(f, g) = 0$. To show that $f = g$, we proceed by contradiction and assume that $f \neq g$. So there exists $c \in [0, 1]$ such that $f(c) \neq g(x)$. Since f and g are continuous, there must be some small interval I in $[0, 1]$ containing c such that $f(x) \neq g(x)$ on I. Then $f(x) - g(x) > 0$ on I and

$$\int_I |f(x) - g(x)|\, dx > 0.$$

Therefore,

$$d(f, g) = \int_0^1 |f(x) - g(x)|\, dx = \int_{x \notin I} |f(x) - g(x)|\, dx + \int_{x \in I} |f(x) - g(x)|\, dx > 0.$$

So $d(f, g) = 0$ implies that $f = g$.

Finally, we consider the triangle inequality. Let $h \in C[0, 1]$. Then

$$\begin{aligned} \int_0^1 |f(x) - g(x)|\, dx &= \int_0^1 |(f(x) - h(x)) + (h(x) - g(x)|\, dx \\ &\leq \int_0^1 |f(x) - h(x)| + |h(x) - g(x)|\, dx \\ &= \int_0^1 |f(x) - h(x)|\, dx + \int_0^1 |h(x) - g(x)|\, dx \\ &= d(f, h) + d(h, g). \end{aligned}$$

The unit circle is difficult to visualize here, but it is the set of functions f that bound one unit of area between the graph of f and the x-axis on $[0, 1]$.

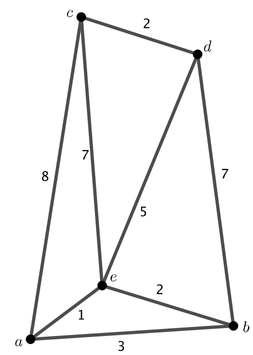


Figure 3.4: A graph to define a metric.

It should be noted that not all metric spaces are infinite. We could make a finite metric space by taking any finite subset S of a metric space (X, d) and use a metric the restriction of d to S. Another way to construct a finite metric space is to start with a finite set of points and make a graph with the points as vertices. Construct edges so that the graph is connected (that is, there is a path from any one vertex to any other) and give weights to the edges as illustrated in Figure 3.4. We then define a metric d on S by letting $d(x, y)$ be the length of a shortest path between vertices x and y in the graph. For example, $d(b, c) = d(b, e) + d(e, c) = 9$ in this example.

The Euclidean Metric on $\mathbb{R}^n$

The metric space that is most familiar to us is the metric space $(\mathbb{R}^2, d_E)$, where

$$d_E((x_1, x_2), (y_1, y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

The metric d_E called the *standard* or *Euclidean* metric on $\mathbb{R}^2$.

We can generalize this Euclidean metric from $\mathbb{R}^2$ to any dimensional real space. Let n be a positive integer and let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. We define $d_E : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$d_E(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

In the next activity we will show that d_E satisfies the first three properties of a metric.

Activity 3.2. Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$.

(a) Show that $d_E(x, y) \geq 0$.

(b) Show that $d_E(x, y) = d_E(y, x)$.

(c) Show that if $x = y$, then $d_E(x, y) = 0$.

(d) Show that if $d_E(x, y) = 0$, then $x = y$.

Activity Solution.

(a) Since $(x_i - y_i)^2 \geq 0$ for each i, we have

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \geq 0.$$

(b) Since $(x_i - y_i)^2 = (y_i - x_i)^2$ for each i, we have

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

(c) If $x = y$, then $x_i = y_i$ and $x_i - y_i = 0$ for each i . Then

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{\sum_{i=1}^{n} 0} = 0.$$

(d) Suppose $d_E(x, y) = 0$. If $x_k \neq y_k$ for some k, then $(x_k - y_k)^2 > 0$. This makes

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{\sum_{i \neq k}(x_i - y_i)^2 + (x_k - y_k)^2} \geq \sqrt{0 + (x_k - y_k)^2} > 0.$$

So we must have $x_i = y_i$ for every i and $x = y$.

Proving that the triangle inequality is satisfied is often the most difficult part of proving that a function is a metric. We will work through this proof with the help of the Cauchy-Schwarz Inequality.

Lemma 3.4 (Cauchy-Schwarz Inequality)**.** *Let n be a positive integer and $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. Then*

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sqrt{\sum_{i=1}^{n} x_i^2}\right)\left(\sqrt{\sum_{i=1}^{n} y_i^2}\right). \tag{3.1}$$

Activity 3.3. Before we prove the Cauchy-Schwarz Inequality, let us analyze it in two specific situations.

(a) Let $x = (1, 4)$ and $y = (3, 2)$ in $\mathbb{R}^2$. Verify the Cauchy-Schwarz Inequality in this case.

(b) Let $x = (1, 2, -3)$ and $y = (-4, 0, -1)$ in $\mathbb{R}^3$. Verify the Cauchy-Schwarz Inequality in this case.

Activity Solution.

(a) Here we have

$$\sum_{i=1}^{2} x_i y_i = (1)(3) + (4)(2) = 11$$

and

$$\left(\sqrt{\sum_{i=1}^{2} x_i^2}\right)\left(\sqrt{\sum_{i=1}^{2} y_i^2}\right) = \sqrt{1+9}\sqrt{16+4} = \sqrt{200} \approx 14.1.$$

(b) Here we have

$$\sum_{i=1}^{3} x_i y_i = (1)(-4) + (2)(0) + (-3)(-1) = -1$$

and

$$\left(\sqrt{\sum_{i=1}^{3} x_i^2}\right)\left(\sqrt{\sum_{i=1}^{3} y_i^2}\right) = \sqrt{1+4+9}\sqrt{16+0+1} = \sqrt{238} \approx 15.4.$$

Now we prove the Cauchy-Schwarz Inequality.

Proof of the Cauchy-Schwarz Inequality. Let n be a positive integer and $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. To verify (3.1) it suffices to show that

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

This is difficult to do directly, but there is a nice trick one can use. Consider the expression

$$\sum (x_i - \lambda y_i)^2.$$

(All of our sums are understood to be from 1 to n, so we will omit the limits on the sums for the remainder of the proof.) Now

$$\begin{aligned} 0 &\le \sum (x_i - \lambda y_i)^2 \\ &= \sum \left(x_i^2 - 2\lambda x_i y_i + \lambda^2 y_i^2\right) \\ &= \left(\sum y_i^2\right)\lambda^2 - 2\left(\sum x_i y_i\right)\lambda + \left(\sum x_i^2\right). \end{aligned} \tag{3.2}$$

To interpret this last expression more clearly, let $a = \sum y_i^2$, $b = -2\sum x_i y_i$ and $c = \sum x_i^2$. The inequality defined by (3.2) can then be written in the form

$$p(\lambda) = a\lambda^2 + b\lambda + c \ge 0.$$

So we have a quadratic $p(\lambda)$ that is never negative. This implies that the quadratic $p(\lambda)$ can have at most one real zero. The quadratic formula gives the roots of $p(\lambda)$ as

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

If $b^2 - 4ac > 0$, then $p(\lambda)$ has two real roots. Therefore, in order for $p(\lambda)$ to have at most one real zero we must have

$$0 \ge b^2 - 4ac = 4\left(\sum x_i y_i\right)^2 - 4\left(\sum y_i^2\right)\left(\sum x_i^2\right)$$

or
$$\left(\sum y_i^2\right)\left(\sum x_i^2\right) \geq \left(\sum x_i y_i\right)^2.$$
This establishes the Cauchy-Schwarz Inequality. ■

One consequence of the Cauchy-Schwarz Inequality that we will need to show that d_E is a metric is the following.

Corollary 3.5. *Let n be a positive integer and $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. Then*
$$\sqrt{\sum_{i=1}^{n}(x_i + y_i)^2} \leq \sqrt{\sum_{i=1}^{n} x_i^2} + \sqrt{\sum_{i=1}^{n} y_i^2}.$$

Activity 3.4. Before we prove the corollary, let us analyze it in two specific situations.

(a) Let $x = (1, 4)$ and $y = (3, 2)$ in $\mathbb{R}^2$. Verify Corollary 3.5 in this case.

(b) Let $x = (1, 2, -3)$ and $y = (-4, 0, -1)$ in $\mathbb{R}^3$. Verify Corollary 3.5 in this case.

Activity Solution.

(a) Here we have
$$\sqrt{\sum_{i=1}^{2}(x_i + y_i)^2} = \sqrt{4^2 + 6^2} = \sqrt{52} \approx 7.2$$
and
$$\sqrt{\sum_{i=1}^{2} x_i^2} + \sqrt{\sum_{i=1}^{2} y_i^2} = \sqrt{1 + 16} + \sqrt{9 + 4} = \sqrt{17} + \sqrt{13} \approx 7.7.$$

(b) Here we have
$$\sqrt{\sum_{i=1}^{3}(x_i + y_i)^2} = \sqrt{(-3)^1 + 2^2 + (-4)^2} = \sqrt{29} \approx 5.4$$
and
$$\sqrt{\sum_{i=1}^{3} x_i^2} + \sqrt{\sum_{i=1}^{3} y_i^2} = \sqrt{1 + 4 + 9} + \sqrt{16 + 0 + 1} = \sqrt{14} + \sqrt{17} \approx 7.8.$$
Let $x = (1, 2, -3)$ and $y = (-4, 0, -1)$ in $\mathbb{R}^3$. Verify Corollary 3.5 in this case.

Now we prove Corollary 3.5.

Proof of Corollary 3.5. Let n be a positive integer and $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. Now
$$\begin{aligned}
\sum (x_i + y_i)^2 &= \sum \left(x_i^2 + 2x_i y_i + y_i^2\right) \\
&= \sum x_i^2 + 2\sum x_i y_i + \sum y_i^2 \\
&\leq \sum x_i^2 + 2\left(\sqrt{\sum x_i^2}\right)\left(\sqrt{\sum y_i^2}\right) + \sum y_i^2 \\
&= \left(\sqrt{\sum x_i^2} + \sqrt{\sum y_i^2}\right)^2.
\end{aligned}$$

Taking the square roots of both sides yields the desired inequality. ■

We can now complete the proof that d_E is a metric.

Activity 3.5. Let n be a positive integer and $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$, and $z = (z_1, z_2, \ldots, z_n)$ be in $\mathbb{R}^n$. Use Corollary 3.5 to show that

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

Activity Solution. Now we can apply Corollary 3.5 to verify that d_E satisfies the triangle inequality. Let n be a positive integer and $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. Then

$$\begin{aligned} d_E(x, z) + d_E(z, y) &= \sqrt{\sum (x_i - z_i)^2} + \sqrt{\sum (z_i - y_i)^2} \\ &\geq \sqrt{\sum \left[(x_i - z_i) + (z_i - y_i)\right]^2} \\ &= \sqrt{\sum (x_i - y_i)^2} \\ &= d_E(x, y). \end{aligned}$$

This concludes our proof that the Euclidean metric is in fact a metric.

We have seen several metrics in this section, some of which are given special names. Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$

- The Euclidean metric d_E, where

$$d_E(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

- The Taxicab metric d_T, where

$$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n| = \sum_{i=1}^{n} \{|x_i - y_i|\}.$$

- The max metric d_M, where

$$d_M(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|, |x_3 - y_3|, \ldots, |x_n - y_n|\} = \max_{1 \leq i \leq n} \{|x_i - y_i|\}.$$

We have only shown that d_T and d_M are metrics on $\mathbb{R}^2$, but similar arguments apply in $\mathbb{R}^n$. Proofs are left to the reader. In addition, the discrete metric

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

makes any set X into a metric space.

Summary

Important ideas that we discussed in this section include the following.

- A metric on a space X is a function that measures distance between elements in the space. More formally, a metric on a space X is a function $d : X \times X \to \mathbb{R} + \cup\{0\}$ such that

  (1) $d(x, y) \geq 0$ for all $x, y \in X$,

  (2) $d(x, y) = 0$ if and only if $x = y$ in X,

  (3) $d(x, y) = d(y, x)$ for all $x, y \in X$, and

  (4) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

  A metric space is any space combined with a metric defined on that space.

- The Euclidean, taxicab, and max metric are all metrics on $\mathbb{R}^n$, so they all provide ways to measure distances between points in $\mathbb{R}^n$. These metric are different in how they define the distances.

  The Euclidean metric is the standard metric that we have used through our mathematical careers. For elements $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, the Euclidean metric d_E is defined as

  $$d_E(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

  With this metric, the unit circle in $\mathbb{R}^2$ (the set of points a distance 1 from the origin) is the standard unit circle we know from Euclidean geometry.

  The taxicab metric d_T is defined as

  $$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n| = \sum_{i=1}^{n} |x_i - y_i|.$$

  The unit circle in $\mathbb{R}^2$ using the taxicab metric is the square with vertices $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$ when viewed in Euclidean geometry.

  The max metric d_M is defined by

  $$d_M(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|, |x_3 - y_3|, \ldots, |x_n - y_n|\} = \max_{1 \leq i \leq n} \{|x_i - y_i|\}.$$

  Under the max metric, the unit circle in $\mathbb{R}^2$ is the square with vertices $(1, 1)$, $(-1, 1)$, $(-1, -1)$, and $(1, -1)$ when viewed in Euclidean geometry.

Section 4

The Greatest Lower Bound

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a lower bound and a greatest lower bound of a subset of $\mathbb{R}$?
- What is an upper bound and a least upper bound of a subset of $\mathbb{R}$?
- How is a greatest lower bound used to define the distance from a point to a set? Why is it necessary to use a greatest lower bound?

Introduction

The real numbers have a special property that allows us to, among other things, define the distance between a point and a set in a metric space. It also allows us to define distances between subsets of certain types of metric spaces, which creates a whole new metric space whose elements are the subsets of the metric space. We will examine that property of the real numbers in this activity.

We begin by considering the problem of defining the distance between a real number and an interval in $\mathbb{R}$ with the Euclidean metric d_E defined by

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

Let $x = 1$ and let A be the closed interval $[-1, 0]$. It is natural to suggest that the distance between the point x and the set A, denoted $d_E(x, A)$, should be the distance from the point x to the point in A closest to x. So in this case we would say

$$d(x, A) = d(x, [-1, 0]) = d_E(x, 0) = 1.$$

This might lead us to suggest that the distance from a point x to a set A, denoted by $d(x, A)$ is the minimum distance from the point to any point in the set, or $d(x, A) = \min\{d_E(x, a) \mid a \in A\}$.

What if we changed the set A to be the open interval $(-1, 0)$? What then should $d(x, A)$ be, or should this distance even exist? If we think of the distance between a point and a set as measuring how far we have to travel from the point until we reach the set, then in the case of $x = 1$ and $A = (-1, 0)$, as soon as we travel a distance more than 1 from x in the direction of A, we reach the set A. So we might intuitively say that $d_E(x, (-1, 0)) = 1$ as well. But we cannot define this distance as a distance from x to a point in A since $0 \notin A$. We need a different way to formulate the notion of a distance from a point to a set.

In a case like this, with $x = 1$ and $A = (-1, 0)$, we can examine the set $T = \{d_E(x, a) \mid a \in A\}$ and notice some facts about this set. For example, the set T is a subset of the nonnegative real numbers. Also, in this example there are no numbers in T that are smaller than 1. Because of this property, we will call the number 1 a *lower bound* for T. More generally,

Definition 4.1. Let S be a nonempty subset of $\mathbb{R}$. A **lower bound** for S is a real number m such that $m \leq s$ for all $s \in S$.

So our set $T = \{d_E(1, a) \mid a \in (-1, 0)\}$ is bounded below by 1. The set T is also bounded below by 0.5 and 0. In fact, any number less than 1 is a lower bound for T. The critical idea, though, is that no number larger than 1 is a lower bound for T. Because of this we call 1 a *greatest lower bound* of T. More generally,

Definition 4.2. Let S be a nonempty subset of $\mathbb{R}$ that is bounded below. A **greatest lower bound** for S is a real number m such that

(1) m is a lower bound for S and

(2) if k is a lower bound for S, then $m \geq k$.

A greatest lower bound is also called an *infimum*. We might now use this idea of a greatest lower bound to define the distance between 1 and $A = (-1, 0)$ as the greatest lower bound of the set $\{d_E(1, a) \mid a \in (-1, 0)\}$. However, there are questions we need to address before we can do so. One question is whether or not every subset of $\mathbb{R}$ that is bounded below has an infimum. The answer to this question is yes, and we will take this result as an axiom of the real number system (often called the *completeness axiom*).

Preview Activity 4.1.

(1) Does every subset of $\mathbb{R}$ have a lower bound? Explain. (When a subset of $\mathbb{R}$ has a lower bound we say that the set is *bounded below*.)

(2) Which of the following subsets S of $\mathbb{R}$ are bounded below? If the set is bounded below, what is its infimum? Assume the Euclidean metric throughout.

i. $S = \{x \mid 3x^2 - 12x + 3 < 0\}$
ii. $S = \{3x^3 - 1 \mid x \in \mathbb{R}\}$
iii. $S = \{2^r + 3^s \mid r, s \in \mathbb{Z}^+\}$

(3) How would you define a least upper bound of a subset S of $\mathbb{R}$?

Activity Solution.

(1) The answer is no. The set $\mathbb{Z}$ has no lower bound, and neither does the interval $(-\infty, 0]$.

(2) Which of the following subsets S of $\mathbb{R}$ are bounded below? If the set is bounded below, what is its infimum? Assume the Euclidean metric throughout.

i. Since $p(x) = 3x^2 - 12x + 3$ is a parabola that opens up, so will only be negative on a bounded set. The intercepts of p are found to be $x = 2 \pm \sqrt{3}$ via the quadratic formula. Since $p(0) < 0$, $p(x)$ will be negative on the interval $(2-\sqrt{3}, 2+\sqrt{3})$. So S is bounded below by its infimum $2 - \sqrt{3}$.

ii. We know that $\lim_{x \to -\infty} 3x^3 - 1 = -\infty$, so the set S is not bounded below.

iii. Since $2^r + 3^s > 0$ for any real numbers r and s, the set S is bounded below by 0. Both 2^r and 3^s increase as r, s increase, so the smallest value of $2^r + 3^s$ occurs when r and s are as small as they can be. This happens when $r = s = 1$. So the infimum of S is $2^1 + 3^1 = 5$.

(3) A least upper bound should be defined in a way similar to the infimum:

Definition 4.3. Let S be a nonempty subset of $\mathbb{R}$ that is bounded above. A **least upper bound** for S is a real number M such that

(a) M is an upper bound for S (that is, $M \geq s$ for all $s \in S$) and

(b) if k is an upper bound for s, then $M \leq k$.

The Distance from a Point to a Set

Metrics are used to establish separation between objects. Topological spaces can be placed into different categories based on how well certain types of sets can be separated. We have defined metrics as measuring distances between points in a metric space, and in this activity we extend that idea to measure the distance between a point and a subset in a metric space. However, there are two questions we need to address before we can do so. The first we mentioned in our preview activity. We will assume the *completeness axiom* of the reals, that is that any subset of $\mathbb{R}$ that is bounded below always has a greatest lower bound. The second question is whether or not a greatest lower bound is unique.

Activity 4.1. Let S be a subset of $\mathbb{R}$ that is bounded below, and assume that S has a greatest lower bound. In this activity we will show that the infimum of S is unique.

(a) What method can we use to prove that there is only one greatest lower bound for S?

(b) Suppose m and m' are both greatest lower bounds for S. Why are m and m' both lower bounds for S?

(c) What two things does the second property of a greatest lower bound tell us about the relationship between m and m'?

(d) Why must the greatest lower bound of S be unique?

Activity Solution.

(a) We assume that there are two greatest lower bounds for S and show that they are equal.

(b) Suppose m and m' are both greatest lower bounds for S. By definition, every greatest lower bound is also a lower bound.

(c) Since m is a lower bound for S and m' is a greatest lower bound for S, it follows that $m \leq m'$. Similarly, m' is a lower bound for S and m is a greatest lower bound for S so $m' \leq m$.

(d) The two inequalities $m \leq m'$ and $m' \leq m$ show that $m = m'$ and so there is only one greatest lower bound of S.

With the existence and uniqueness of greatest lower bounds considered, we can now say that any nonempty subset S of $\mathbb{R}$ that is bounded below has a unique greatest lower bound. We use the notation glb(S) (or $\inf(S)$ for infimum of S) for the greatest lower bound of S. There is also a *least upper bound* (lub, or *supremum* (sup)) of a subset of $\mathbb{R}$ that is bounded above.

Now we can formally define the distance between a point and a subset in a metric space.

Definition 4.4. Let (X, d) be a metric space, let $x \in X$, and let A be a non-empty subset of X. The **distance from x to A** is

$$\inf\{d(x, a) \mid a \in A\}.$$

We denote the distance from x to A by $d(x, A)$. When calculating these distances, it must be understood what the underlying metric is.

Activity 4.2. There are a couple of facts about the distance between a point and a set that we examine in this activity. Let (X, d) be a metric space, let $x \in X$, and let A be a non-empty subset of X

(a) Why must $d(x, A)$ exist?

(b) If $d(x, A) = 0$, must $x \in A$?

Activity Solution.

(a) Since $d(x, y) \geq 0$, the set $\{d(x, a) \mid a \in A\}$ is bounded below by 0. Since A is not empty, the set $\{d(x, a) \mid a \in A\}$ is also not empty. So $\inf\{d(x, a) \mid a \in A\}$ exists.

(b) The answer is no. Let $A = (0, 1)$ in $\mathbb{R}$ using the Euclidean metric. Then $d(1, A0 = 0$, but $1 \notin A$.

Summary

Important ideas that we discussed in this section include the following.

- A lower bound for a nonempty subset S of $\mathbb{R}$ that is bounded below is a real number m such that $m \leq s$ for all $s \in S$. A greatest lower bound for a nonempty subset S of $\mathbb{R}$ that is bounded below is a real number m such that

   i. m is a lower bound for S and
   ii. if k is a lower bound for S, then $m \geq k$.

- A lower bound for a nonempty subset S of $\mathbb{R}$ that is bounded below is a real number m such that $m \leq s$ for all $s \in S$. A greatest lower bound for a nonempty subset S of $\mathbb{R}$ that is bounded below is a real number m such that

   i. m is a lower bound for S and
   ii. if k is a lower bound for S, then $m \geq k$.

- An upper bound for a nonempty subset S of $\mathbb{R}$ that is bounded above is a real number M such that $M \geq s$ for all $s \in S$. A least upper bound for a nonempty subset S of $\mathbb{R}$ that is bounded above is a real number M such that

   i. M is an upper bound for S and
   ii. if k is an upper bound for s, then $M \leq k$.

- The distance from a point x to a set A in a metric space (X, d) is $d(x, A) = \inf\{d(x, a) \mid a \in A\}$. There may be no point $a \in A$ such that $d(x, A) = d(x, a)$, so it is necessary to use an infimum to define this distance.

Section 5

Continuous Functions in Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What does it mean for a function between metric spaces to be continuous at a point?
- What does it mean for a function between metric spaces to be continuous?

Introduction

In calculus we defined a function $f : \mathbb{R} \to \mathbb{R}$ to be continuous at a point a if

$$\lim_{x \to a} f(x) = f(a).$$

This involved providing some explanation about what it means for a function f to have a limit at a point. We waved our hands a bit, providing a superficial, intuitive "definition" that a function f has a limit L at $x = a$ if we can make all of the value of $f(x)$ as close to L as we want by choosing x as close to (but not equal to) a as we need. To extend this informal notion of limit to continuity at a point we would say that a function f is continuous at a point a if if we can make all of the value of $f(x)$ as close to $f(a)$ as we want by choosing x as close to a as we need (now x can equal a).

In order to define continuity in a more general context (in topological spaces) we will need to have a rigorous definition of continuity to work with. We will begin by discussing continuous functions from $\mathbb{R}$ to $\mathbb{R}$, and build from that to continuous functions in metric spaces. These ideas will allow us to ultimately define continuous functions in topological spaces.

We begin by working with continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Our goal is to make more rigorous our informal definition of continuity at a point. To do so will require us to formally defining what we mean by

- making the values of $f(x)$ "as close to $f(a)$ as we want", and

- choosing x "as close to a as we need".

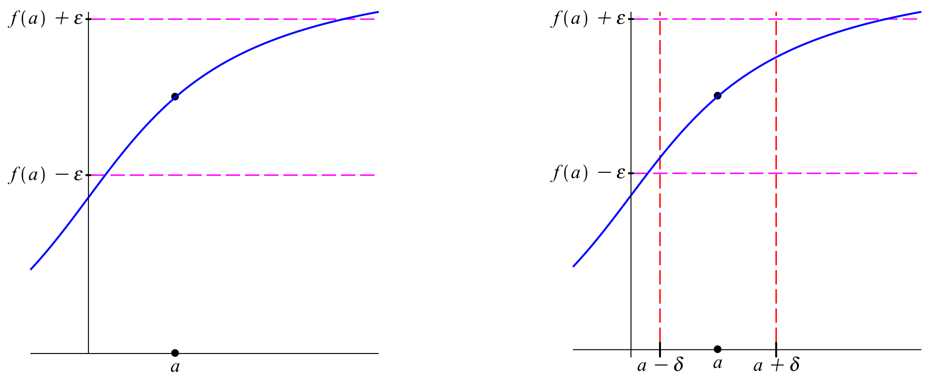


Figure 5.1: Demonstrating the definition of continuity at a point.

Let's deal with the first statement, making the values of $f(x)$ "as close to $f(a)$ as we want". What this means is that if we set any tolerance, say 0.0001, then we can make the values of $f(x)$ within 0.0001 of $f(a)$. Since the absolute value $|f(x) - f(a)|$ measures how close $f(x)$ is to $f(a)$, we can rewrite the statement that the values of $f(x)$ are within 0.0001 of $f(a)$ as $|f(x) - f(a)| < 0.0001$. Of course, 0.0001 may not be as close as we want to $f(a)$, so we need a way to indicate that we can make the values of $f(x)$ arbitrarily close to $f(a)$ – within any tolerance at all. We do this by making the tolerance a parameter, ϵ. Then our job is to make the values of $f(x)$ within ϵ of $f(a)$ regardless of the size of ϵ. We write this as

$$|f(x) - f(a)| < \epsilon.$$

We can picture this as shown at left in Figure 5.1. Here we want to make the values of $f(x)$ lie within an ϵ band of $f(a)$ above and below $f(a)$. That is, we want to be able to make the values of $f(x)$ lie between $f(a) - \epsilon$ and $f(a) + \epsilon$.

Now we have to address the question of how we "make" the values of $f(x)$ to be within ϵ of $f(a)$. Since the values $f(x)$ are the dependent values, dependent on x, we "make" the values of $f(x)$ have the property we want by choosing the inputs x appropriately. In order for f to be continuous at $x = a$, we must be able to find x values close enough to a to force $|f(x) - f(a)| < \epsilon$. Pictorially, we can see how this might happen in the image at right in Figure 5.1. We need to be able to find an interval around $x = a$ so that the graph of $f(x)$ lies in the ϵ band around $f(a)$ for values of x in that interval. In other words, we need to be able to find some positive number δ so that if x is in the interval $(a - \delta, a + \delta)$, then the graph of $f(x)$ lies in the ϵ band around $y = f(a)$. More formally, if we are given any positive tolerance ϵ, we must be able to find a positive number δ so that if $|x - a| < \delta$ (that is, x is in the interval $(a - \delta, a + \delta)$), then $|f(x) - f(a)| < \epsilon$ (or $f(x)$ lies in the ϵ band around $y = f(a)$).

This gives us a rigorous definition of what it means for a function $f : \mathbb{R} \to \mathbb{R}$ to be continuous at a point.

Definition 5.1. A function $f : \mathbb{R} \to \mathbb{R}$ is **continuous at a point** a if, given any $\epsilon > 0$, there exists a $\delta > 0$ so that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

Note that the value of δ can depend on the value of a and on ϵ, but not on values of x.

Preview Activity 5.1. The GeoGebra file at `https://www.geogebra.org/m/rym36sqs` will allow us to play around with this definition. Use this GeoGebra applet for the first two problems in this activity.

(1) Enter $f(x) = x\sin(x)$ as your function. (You can change the viewing window coordinates, the base point a, and the function using the input boxes at the left on the screen.) Determine a value of δ so that $|f(x) - f(1)| < 0.5$ whenever $|x - 1| < \delta$. Explain your method.

(2) Now find a value of δ so that $|f(x) - f(2.5)| < 0.25$ whenever $|x - 2.5| < \delta$. Explain your method.

(3)

(a) What is the negation of the definition of continuity at a point? In other words, what do we need to do to show that a function $f : \mathbb{R} \to \mathbb{R}$ is not continuous at a point $x = a$?

(b) Use the negation of the definition to explain why the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} -1 & \text{if } x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

is not continuous at $x = 1$.

Activity Solution.

(1) We look for a value of δ so that the graph of f on the interval $(1 - \delta, 1 + \delta)$ is contained entirely within the band $(f(1) - 0.5, f(1) + 0.5)$. A value of δ that works in this case is $\delta = 0.3$, as illustrated at left in Figure 5.2.

(2) We look for a value of δ so that the graph of f on the interval $(2.5 - \delta, 2.5 + \delta)$ is contained entirely within the band $(f(2.5) - 0.25, f(2.5) + 0.25)$. A value of δ that works in this case is $\delta = 0.1$, as illustrated at right in Figure 5.2.

(3)

(a) We negate a universal quantifier with an existential quantifier and an existential quantifier with a universal quantifier. So a function $f : \mathbb{R} \to \mathbb{R}$ is not continuous at $x = a$ if there exists an $\epsilon > 0$ so that for all $\delta > 0$, the fact that $|x - a| < \delta$ does not imply $|f(x) - f(a)| < \epsilon$.

(b) The graph of f is shown in Figure 5.3. Let $\epsilon = 0.5$. Given $\delta > 0$, the point $x = 1 - \frac{\delta}{2}$ is in the interval $(1 - \delta, 1 + \delta)$, but

$$|f(x) - f(1)| = |f\left(1 - \frac{\delta}{2}\right) - f(1)| = |-1 - 1| = 2.$$

So $|f(x) - f(1)| > \epsilon$. This shows that f is not continuous at $x = 1$.

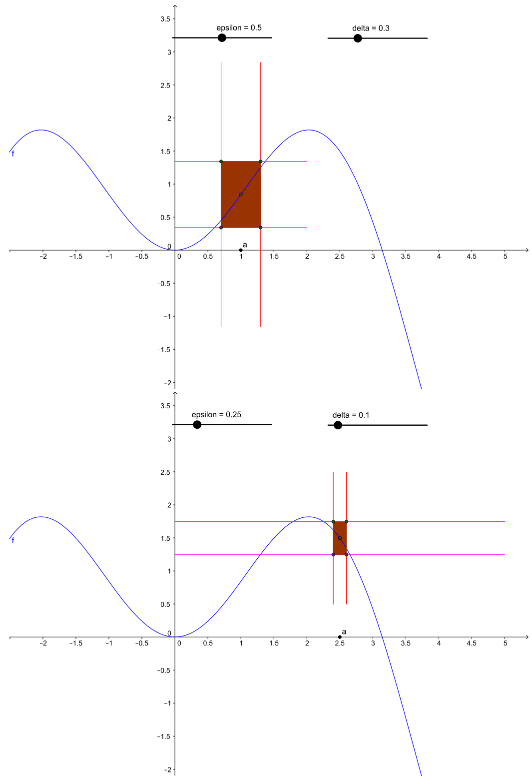


Figure 5.2: Demonstrating the definition of continuity at a point.

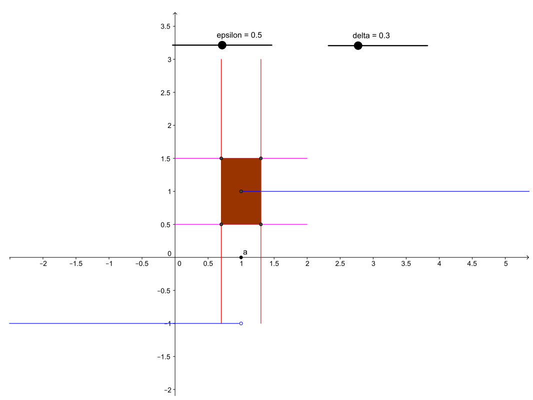


Figure 5.3: The function f is not continuous at $x = 1$.

Continuous Functions Between Metric Spaces

In our preview activity we saw how to formally define what it means for a function $f : \mathbb{R} \to \mathbb{R}$ to be continuous at a point.

Note that Definition 5.1 depends only on being able to measure how close points are to each other. Since that is precisely what a metric does, we can extend this notion of continuity to define continuity for functions between metric spaces. Continuity is an important idea in topology, and we will work with this idea extensively throughout the semester.

If we let $d_E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by $d_R(x, y) = |x - y|$, then we have seen that d_E is a metric on $\mathbb{R}$ (note that d_E is the Euclidean metric on $\mathbb{R}$). Using this metric we can reformulate what it means for a function $f : \mathbb{R} \to \mathbb{R}$ to be continuous at a point.

Definition 5.2 (Alternate Definition)**.** A function $f : \mathbb{R} \to \mathbb{R}$ is **continuous at a point** a if, given any $\epsilon > 0$, there exists a $\delta > 0$ so that $d_E(x, a) < \delta$ implies $d_E(f(x), f(a)) < \epsilon$.

This alternate definition depends on the metric d_E. We could easily replace the metric d with any other metric we choose. This allows us to define continuity at a point for functions between metric spaces.

Definition 5.3. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is **continuous** at $a \in X$ if, given any $\epsilon > 0$, there exists a $\delta > 0$ so that $d_X(x, a) < \delta$ implies $d_Y(f(x), f(a)) < \epsilon$.

Once we have defined continuity at a point, we can define continuous functions.

Definition 5.4. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is **continuous** if f is continuous at every point in X.

Example 5.5. In general, to prove that a function $f : X \to Y$ is continuous, where (X, d_X) and (Y, d_Y) are metric spaces, we begin by choosing an arbitrary element a in X. Then we let ϵ be a number greater than 0 and show that there is a $\delta > 0$ so that $d_Y(f(x), f(a)) < \epsilon$ whenever $d_X(x, a) < \delta$. The δ we need cannot depend on x (since x isn't known), but can depend on the value of a that we choose. As an example, let $X = \mathbb{R}$ and let d_X be defined as

$$d_X(x, y) = \min\{|x - y|, 1\}.$$

The proof that d_X is a metric is left to the reader. Consider $f : X \to Y$ defined by $f(x) = x^2$, where $(Y, d_Y) = (\mathbb{R}, d_E)$. To show that f is continuous, we let $a \in \mathbb{R}$ and let $\epsilon > 0$.

> **Scratch work.** What happens next is not part of the proof, but shows how we go about finding a δ we need. We are looking for $\delta > 0$ such that $d_X(x, a) < \delta$ implies that $d_E(f(x), f(a)) < \epsilon$. That is, we want to make
>
> $$d_E(f(x), f(a)) = \sqrt{(f(x) - f(a))^2} = |f(x) - f(a)| = |x^2 - a^2| < \epsilon$$
>
> whenever
>
> $$d_X(x, a) = \min\{|x - a|, 1\} < \delta.$$
>
> Now $|x^2 - a^2| = |(x - a)(x + a)| = |x - a|\,|x + a|$. If $d_X(x, a) < \delta$, then $\min\{|x - a|, 1\} < \delta$. If we choose $\delta < 1$, then $d_X(x, a) < \delta < 1$ implies that $|x - a| < 1$ and so $d_X(x, a) = |x - a|$. Now
>
> $$|x + a| = |(x - a) + 2a| \leq |x - a| + 2|a| < 1 + 2|a|.$$
>
> It follows that
>
> $$|x - a|\,|x + a| < \delta(1 + 2|a|).$$
>
> To make this product less that ϵ, we can choose δ such that $\delta(1 + 2|a|) < \epsilon$ or $\delta < \frac{\epsilon}{1+2|a|}$.

> Now we ignore this paragraph and present the proof, which is essentially reversing the steps we just made. If the steps can't be reversed, then we have to rethink our argument. The next step in the proof might seem like magic to the uninitiated reader, but we have seen behind the curtain so it isn't a mystery to us.

Let δ be a positive number less than $\min\left\{1, \frac{\epsilon}{1+2|a|}\right\}$. Then

$$d_X(x, a) = \min\{|x - a|, 1\} < \delta$$

implies that $d_X(x, a) < \delta < 1$ and so $d_X(x, a) = |x - a| < \delta < 1$. Then

$$|x + a| = |(x - a) + 2a| \leq |x - a| + 2|a| < 1 + 2|a|.$$

It follows that

$$\begin{aligned}
d_E(f(x), f(a)) &= \sqrt{(f(x) - f(a))^2} \\
&= |f(x) - f(a)| \\
&= |x^2 - a^2| \\
&= |(x - a)(x + a)| \\
&= |x - a|\,|x + a| \\
&< \delta(1 + 2|a|) \\
&< \left(\frac{\epsilon}{1 + 2|a|}\right)(1 + 2|a|) \\
&= \epsilon.
\end{aligned}$$

We conclude that f is continuous at every point in X and so f is a continuous function.

Not all functions are continuous.

Example 5.6. Let $X = Y = \mathbb{R}$ and define $f : X \to Y$ by $f(x) = x$. Let d_X be the Euclidean metric and d_Y the discrete metric. (Recall that $d_Y(x, y) = 1$ whenever $x \neq y$.) Let $a \in X$ and let $0 < \epsilon < 1$.

Let $\delta > 0$, and let $x = a + \frac{\delta}{2}$. Then $x \neq a$ and $d_X(x, a) < \delta$. However,

$$d_Y(f(x), f(a)) = d_Y(x, a) = 1 > \epsilon.$$

So if $0 < \epsilon < 1$, there is no $\delta > 0$ such that $d_X(x, a) < \delta$ implies that $d_Y(f(x), f(a)) < \epsilon$. We conclude that f is continuous at no point in X.

Certain functions are always continuous, as the next activity shows.

Activity 5.1.

(a) Let (X, d_X) and (Y, d_Y) be metric spaces, and let $b \in Y$. Define $f : X \to Y$ by $f(x) = b$ for every $x \in X$. Show that f is a continuous function.

(b) Let (X, d) be a metric space. Define the function $i_X : X \to X$ by $i_X(x) = x$ for every $x \in X$. Show that i_X is a continuous function. (The function i_X is called the *identity function* on X.)

Activity Solution.

(a) Let $a \in X$ and let ϵ be greater than 0. Choose δ to be any positive number. If $d_X(x, a) < \delta$, then $d_Y(f(x), f(a)) = d_Y(b, b) = 0 < \epsilon$. Therefore, f is continuous at every point in X and so is a continuous function.

(b) Let $a \in X$ and let ϵ be greater than 0. Choose $\delta = \epsilon$. If $d_X(x, a) < \delta$, then $d_X(f(x), f(a)) = d_X(x, a) < \delta = \epsilon$. Therefore, f is continuous at every point in X and so is a continuous function.

More complicated examples are in the next activity.

Activity 5.2. Let $X = (\mathbb{R}^2, d_T)$ and $Y = (\mathbb{R}^2, d_M)$, where

$$d_T((x_1, x_2), (y_1, y_2)) = |x_1 - y_1| + |x_2 - y_2|$$

is the taxicab metric and

$$d_M((x_1, x_2), (y_1, y_2)) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

is the max metric. Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f((a, b)) = (a + b, b)$.

(a) Is f a continuous function from X to Y? Justify your answer.

(b) Is f a continuous function from Y to X? Justify your answer.

Activity Solution.

(a) Let $a = (a_1, a_2) \in \mathbb{R}^2$ and let $\epsilon > 0$. Choose any $0 < \delta < \epsilon$. Let $x = (x_1, x_2)$ in $\mathbb{R}^2$ and suppose $d_T(x, a) < \delta$. Then

$$d_T(x, a) = |x_1 - a_1\ | + |x_2 - a_2| < \delta = \epsilon.$$

Then

$$\begin{aligned} d_M(f(x), f(a)) &= d_M((x_1 + x_2, x_1), (a_1 + a_2, a_2)) \\ &= \max\{|(x_1 + x_2) - (a_1 + a_2)|, |x_2 - a_2|\} \\ &= \max\{|(x_1 - a_1) + (x_2 - a_2)|, |x_2 - a_2|\} \\ &\leq \max\{|x_1 - a_1| + |x_2 - a_2|, |x_2 - a_2|\} \\ &= |x_1 - a_1 + |x_2 - a_2| < \epsilon. \end{aligned}$$

Thus, the function f is continuous at every point in $\mathbb{R}^2$ and is therefore a continuous function from X to Y.

(b) Let $a = (a_1, a_2) \in \mathbb{R}^2$ and let $\epsilon > 0$. Choose any $0 < \delta < \frac{\epsilon}{3}$. Let $x = (x_1, x_2)$ in $\mathbb{R}^2$ and suppose $d_M(x, a) < \delta$. Then

$$d_M(x, a) = \max\{|x_1 - a_1\ |, |x_2 - a_2|\} < \delta < \frac{\epsilon}{3}.$$

So $|x_1 - a_1| < \frac{\epsilon}{3}$ and $|x_2 - a_2| < \frac{\epsilon}{3}$. Then

$$\begin{aligned} d_T(f(x), f(a)) &= d_T((x_1 + x_2, x_1), (a_1 + a_2, a_2)) \\ &= |(x_1 + x_2) - (a_1 + a_2)| + |x_2 - a_2| \\ &= |(x_1 - a_1) + (x_2 - a_2)| + |x_2 - a_2|\} \\ &\leq (|x_1 - a_1| + |x_2 - a_2|) + |x_2 - a_2| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon. \end{aligned}$$

Thus, the function f is continuous at every point in $\mathbb{R}^2$ and is therefore a continuous function from Y to X.

Composites of Continuous Functions

Let (X, d_X), (Y, d_Y), and (Z, d_Z) be metric spaces, and suppose $f : X \to Y$ and $g : Y \to Z$ are continuous functions. It seems natural to ask if $g \circ f : X \to Z$ is a continuous function.

Activity 5.3. Let (X, d_X), (Y, d_Y), and (Z, d_Z) be metric spaces, and suppose $f : X \to Y$ and $g : Y \to Z$ are continuous functions. We will prove that $g \circ f$ is a continuous function.

(a) What do we have to do to show that $g \circ f$ is a continuous function? What are the first two steps in our proof?

(b) Let $a \in X$ and let $b = f(a)$. Suppose $\epsilon > 0$ is given. Explain why there must exist a $\delta_1 > 0$ so that $d_Y(y, b) < \delta_1$ implies $d_Z(g(y), g(b)) < \epsilon$.

(c) Now explain why there exists a $\delta_2 > 0$ so that $d_X(x, a) < \delta_2$ implies that $d_Y(f(x), f(a)) < \delta_1$.

(d) Prove that $g \circ f : X \to Z$ is a continuous function.

Activity Solution.

(a) We have to start with a point $a \in X$ and an ϵ greater than 0. We have to find a positive number δ such that $d_X(x, a) < \delta$ implies that $d_Z((g \circ f)(x), (g \circ f)(a)) < \epsilon$.

(b) Since g is continuous at b, there exists a $\delta_1 > 0$ so that $d_Y(y, b) < \delta_1$ implies $d_Z(g(y), g(b)) < \epsilon$.

(c) Since f is continuous at a, there exists a $\delta_2 > 0$ so that $d_X(x, a) < \delta_2$ implies that $d_Y(f(x), f(a)) < \delta_1$ (here we are using δ_1 as our ϵ).

(d) Let $\delta = \delta_2$ and suppose that $d_X(x, a) < \delta$. Then $d_Y(f(x), f(a)) < \delta_1$. But then $d_Z(g(f(x)), g(f(a)) < \epsilon$. This shows that $g \circ f : X \to Z$ is a continuous function.

Continuity is an important concept in topology. We have seen how to define continuity in metric spaces, and we will soon expand on this idea to define continuity without reference to metrics at all. This will allow us to later define continuous functions between arbitrary topological spaces.

Summary

Important ideas that we discussed in this section include the following.

- Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous at $a \in X$ if, given any $\epsilon > 0$, there exists a $\delta > 0$ so that $d_X(x, a) < \delta$ implies $d_Y(f(x), f(a)) < \epsilon$.

- Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous if f is continuous at every point in X.

Section 6

Open Balls and Neighborhoods in Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is an open ball in a metric space? Give one important property of open balls.
- What is a neighborhood of a point in a metric space?
- How can we use open balls or neighborhoods to determine the continuity of a function at a point?

Introduction

Open sets are vitally important in topology. In fact, we will see later that every topological space is completely defined by its open sets. In this section we introduce the idea of open balls and neighborhoods in metric spaces and discover a few of their properties.

Recall that the continuity of a function f from a metric space (X, d_X) to a metric space (Y, d_Y) at a point a is defined in terms of sets of points $x \in X$ such that $d_X(x, a) < \delta$ and $y \in Y$ such that $d_Y(y, f(a)) < \epsilon$ for positive real numbers δ and ϵ. In $\mathbb{R}$ with the Euclidean metric d_E, for real numbers x and a the set of x values satisfying $d_E(x, a) < \delta$ is the set of x values so that $|x - a| < \delta$. We often write this set in interval notation as $(a - \delta, a + \delta)$ and call $(a - \delta, a + \delta)$ an open interval. An informal reason that we call such an interval open (as opposed to the intervals $[a - \delta, a + \delta)$, $(a - \delta, a + \delta]$, or $[a - \delta, a + \delta]$) is that the open interval does not contain either of its endpoints. A more substantial reason to call such an interval open is that if x' is any element in $(a - \delta, a + \delta)$, then we can find another open interval around x' that is completely contained in the interval $(a - \delta, a + \delta)$. So you could naively think of an open interval as one in which there is enough room in the interval for any point in the interval to wiggle around a bit and stay within the interval.

Since the open interval $(a-\delta, a+\delta)$ can be described completely by the Euclidean metric as the set of x values so that $d_E(x,a) < \delta$, there is no reason why we can't extend this notation of open interval to any metric space. We must note, though, that $\mathbb{R}$ is one-dimensional while most metric spaces are not, so the term "interval" will no longer be appropriate. We replace the concept of interval with that of an open ball.

Definition 6.1. Let (X, d_X) be a metric space, and let $a \in X$. For $\delta > 0$, the **open ball** $B(a,\delta)$ **of radius** δ **around** a is the set

$$B(a,\delta) = \{x \in X \mid d_X(x,a) < \delta\}.$$

Preview Activity 6.1. Describe and draw a picture of the indicated open ball in each of the following metric spaces.

(1) The open ball $B(2,1)$ in the metric space $(\mathbb{R}, d_E)$ with the Euclidean metric

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

(2) The open ball $B((3,2),1)$ in the metric space $(\mathbb{R}^2, d_E)$ with the Euclidean metric

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

(3) The open ball $B((3,2),1)$ in the metric space $(\mathbb{R}^2, d_M)$ with the max metric

$$d_M((x_1,x_2),(y_1,y_2)) = \max\{|x_1-y_1|, |x_2-y_2|\}.$$

(4) The open ball $B((3,2),1)$ in the metric space $(\mathbb{R}^2, d_T)$ with the taxicab metric

$$d_T((x_1,x_2),(y_1,y_2)) = |x_1-y_1| + |x_2-y_2|.$$

(5) The open ball $B((3,2),1)$ in the metric space $(\mathbb{R}^2, d)$ with the discrete metric

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

What is the difference between $B((3,2),1)$ and $B((3,2),r)$ in this metric space if $r > 1$?

*

Neighborhoods

We are familiar with the idea of open intervals in $\mathbb{R}$. We next introduce the idea of an open neighborhood of a point and characterize continuity in terms of neighborhoods.

The open ball $B(a,\delta)$ in a metric space (X,d) is also called the *δ-neighborhood* around a. A neighborhood of a point can be thought of as any set that envelops that point. Neighborhoods can be larger than just open balls.

Definition 6.2. Let (X, d_X) be a metric space, and let $a \in X$. A subset N of X is a **neighborhood** of a if there exists a $\delta > 0$ such that $B(a, \delta) \subseteq N$.

In particular, the open ball $B(a, \delta)$ is a neighborhood of a. In fact, we can say something more about open balls.

Activity 6.1. Let (X, d) be a metric space, let $a \in X$, and let $\delta > 0$. In this activity we ask the question, is $B(a, \delta)$ a neighborhood of each of its points?

(a) Let $b \in B(a, \delta)$. What do we have to do to show that $B(a, \delta)$ is a neighborhood of b?

(b) Use Figure 6.1 to help show that $B(a, \delta)$ is a neighborhood of b.

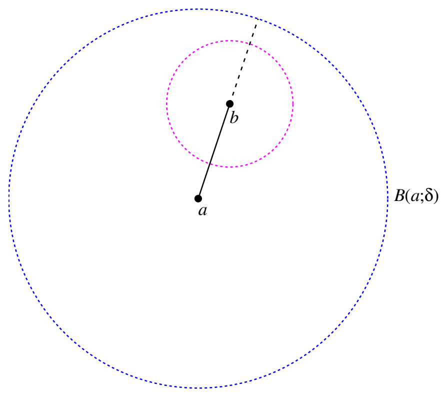


Figure 6.1: $B(a, \delta)$ as a neighborhood of b.

(c) Is the converse true? That is, if a set is a neighborhood of each of its points, is the set an open ball? No proof is necessary, but a convincing argument is in order.

*

Continuity and Neighborhoods

We can define continuity now in terms of neighborhoods instead of using metrics. The advantage here is that this idea does not explicitly depend on the existence of a metric, so we will be able to adopt this concept of continuity for arbitrary topological spaces.

Recall that a function f from a metric space (X, d_X) to a metric space (Y, d_Y) is continuous at $a \in X$ if, for any $\epsilon > 0$ there exists $\delta > 0$ so that $d_X(x, a) < \delta$ implies $d_Y(f(x), f(a)) < \epsilon$. We can interpret this definition of continuity to say that for every $\epsilon > 0$, the inverse image under f of the open ball $B(f(a), \epsilon)$ contains the open ball $B(a, \delta)$ for some $\delta > 0$.

It is a reasonable question to ask at this point if it must be the case that the sets $f^{-1}(B(f(a), \epsilon))$ and $B(a, \delta)$ are always equal.

Activity 6.2. Let f be a function a metric space (X, d_X) to a metric space (Y, d_Y) that is continuous at $a \in X$. Using the notation from the paragraph above, in this activity we determine if $f^{-1}(B(f(a), \epsilon))$ must equal $B(a, \delta)$.

Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = x^2,$$

where we use the Euclidean metric d_E throughout. Assume that f is a continuous function. Then f is continuous at $x = 2$.

(a) What is $B(f(2), 1)$?

(b) What is $f^{-1}(B(f(2), 1))$?

(c) Is $f^{-1}(B(f(2), 1))$ an open ball centered at 2? Explain.

$*$

The conclusion to be drawn from Activity 6.2 is that if f is continuous, we can only conclude that the inverse image of $B(f(a), \epsilon)$ *contains* an open ball centered at a. By definition of continuity, if for every $\epsilon > 0$ there exists a $\delta > 0$ so that the open ball $f^{-1}(B(f(a), \epsilon))$ contains $B(a, \delta)$, then f is continuous at a. We summarize this in the next theorem.

Theorem 6.3. *Let f be a function a metric space (X, d_X) to a metric space (Y, d_Y), and let $a \in X$. Then f is continuous at $a \in X$ if and only if, given any $\epsilon > 0$ there exists $\delta > 0$ so that*

$$B(a, \delta) \subseteq f^{-1}(B(f(a), \epsilon)).$$

We can extend this idea of continuity to describe continuity in terms of neighborhoods.

Theorem 6.4. *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a function. Then f is continuous at $a \in X$ if and only if the inverse image of every neighborhood of $f(a)$ is a neighborhood of a.*

Proof. Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a function. To prove this biconditional statement we need to prove both implications. First assume that f is continuous at some point $a \in X$. We will show that for any neighborhood N of $f(a)$ in Y, its inverse image $f^{-1}(N)$ in X is a neighborhood of a in X. Let N be a neighborhood of $f(a)$ in Y. To demonstrate that $f^{-1}(N)$ is a neighborhood of a in X, we need to find an open ball around a that is contained in $f^{-1}(N)$. Since N is a neighborhood of $f(a)$, by definition there exists $\epsilon > 0$ so that $B(f(a), \epsilon) \subseteq N$. Since f is continuous at a, there exists $\delta > 0$ such that $B(a, \delta) \subseteq f^{-1}(B(f(a), \epsilon))$. So if $x \in B(a, \delta)$, then $f(x) \in B(f(a), \epsilon) \subseteq N$. So $B(a, \delta) \subseteq f^{-1}(N)$, and $f^{-1}(N)$ is a neighborhood of a in X.

The proof of the reverse implication is left for the next activity. ■

Activity 6.3. Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a function. Let $a \in X$. In this activity we prove that if the inverse image of every neighborhood of $f(a)$ is a neighborhood of a, then f is continuous at a.

(a) What does Theorem 6.3 tell us that we need to do to show that f is continuous at a?

(b) Suppose ϵ is greater than 0, why is $B(f(a), \epsilon)$ a neighborhood of $f(a)$ in Y?

(c) What does our hypothesis tell us about $f^{-1}(B(f(a), \epsilon))$?

(d) What can we conclude from part (c)?

(e) How do (a)-(d) show that f is continuous at a?

*

We conclude this section with some important facts about neighborhoods. Assume that (X, d) is a metric space and $a \in X$.

- There is a neighborhood that contains a.
- If N is a neighborhood of a and $N \subseteq N'$, then N' is a neighborhood of a.
- If M and N are neighborhoods of a, then so is $M \cap N$.

The proofs are straightforward and left to the reader.

Summary

Important ideas that we discussed in this section include the following.

- If (X, d) is a metric space and $a \in X$, then an open ball centered at a is a set of the form

$$B(a, \delta) = \{x \in X \mid d(x, a) < \delta\}$$

  for some positive number δ. An important property of open balls is that every open ball is a neighborhood of each of its points.
- A subset N of a metric space (X, d) is s neighborhood of a point $a \in N$ if there is a positive real number δ such that $B(a, \delta) \subseteq N$.
- A function f from a metric space (X, d_X) to a metric space (Y, d_Y) is continuous at $a \in X$ if $f^{-1}(N)$ is a neighborhood of a in X for any neighborhood N of $f(a)$ in Y.

Section 7

Open Sets in Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is an open set in a metric space?
- What is an interior point of a subset of a metric space? How are interior points related to open sets?
- What is the interior of a set? How is the interior of a set related to open sets?
- How can we use open sets to determine the continuity of a function?
- What important properties do open sets have in relation to unions and intersections?

Introduction

Consider the interval (a, b) in $\mathbb{R}$ using the Euclidean metric. If $m = \frac{a+b}{2}$, then $(a, b) = B\left(m, \frac{b-a}{2}\right)$, so every open interval is an open ball. As an open ball, an open interval (a, b) is a neighborhood of each of its points. This is the foundation for the definition of an open set in a metric space.

Recall that we defined a subset N of X to be neighborhood of point a in a metric space (X, d) if N contains an open ball $B(a, \epsilon)$ for some $\epsilon > 0$. We saw that every open ball is a neighborhood of each of its points, and we will now extend that idea to define an *open set* in a metric space.

Definition 7.1. A subset O of a metric space X is an **open set** if O is a neighborhood of each of its points.

So, by definition, any open ball is an open set. Also by definition, open sets are neighborhoods of each of their points. Open sets are different than non-open sets. For example, $(0, 1)$ is an open set in $\mathbb{R}$ using the Euclidean metric, but $[0, 1)$ is not. The reason $[0, 1)$ is not an open set is that there is no open ball centered at 0 that is entirely contained in $[0, 1)$. So 0 has a different property than the other points in $[0, 1)$. The set $[0, 1)$ is a neighborhood of each of the points in $(0, 1)$, but is not a neighborhood of 0. We can think of the points in $(0, 1)$ as being in the interior of the set $[0, 1)$. This leads to the next definition.

Definition 7.2. Let A be a subset of a metric space X. A point $a \in A$ is an **interior point** of A if A is a neighborhood of a.

As we will soon see, open sets can be characterized in terms of interior points.

Preview Activity 7.1.

(1) Determine if the set A is an open set in the metric space (X, d). Explain your reasoning.

  (a) $X = \mathbb{R}$, $d = d_E$, the Euclidean metric, $A = [0, 0.5)$.

  (b) $X = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$, $d = d_E$, the Euclidean metric, $A = [0, 0.5)$. Assume that the Euclidean metric is a metric on X.

  (c) $X = \{a, b, c, d\}$, d is the discrete metric defined by

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

  and $A = \{a, b\}$.

(2)

  (a) What are the interior points of the following sets in $(\mathbb{R}, d_E)$? Explain.

$$(0, 1) \quad (0, 1] \quad [0, 1) \quad [0, 1].$$

  (b) Let $A = \{0, 1, 2\}$ in $(\mathbb{R}, d_E)$. What are the interior points of A? Explain.

  (c) Let $\mathbb{Q}$ be the set of rational numbers in $(\mathbb{R}, d_E)$. What are the interior points of $\mathbb{Q}$? Explain.

Activity Solution.

(1) Determine if the set A is an open set in the metric space (X, d). Explain your reasoning.

  (a) The set A is not an open set in X. Note that $0 \in A$, but the open ball of radius ϵ in $\mathbb{R}$ centered at 0 has the form $(-\epsilon, \epsilon)$. So not open ball in $\mathbb{R}$ centered at 0 is entirely contained in A. Thus, A is not a neighborhood of 0 and A is not an open set.

  (b) The set A is an open set in X. Note that every open ball in X must be contained in X. So the open ball of radius 0.25 in X centered at 0 has the form $[0, 0.25)$, which is entirely contained in A. If a is any point of A other than 0, then the open ball centered at a of radius $\min\{a, 0.5 - a\}$ is contained in A. So A is a neighborhood of each of its points and A is an open set.

  (c) Notice that if $x \in X$, then $B(x, 0.5) = \{x\}$. So A contains the open balls $B(a, 0.5)$ and $B(b, 0.5)$. Thus, A is a neighborhood of each of its points and A is an open set.

(2)

(a) If $0 < a < 1$, then each of these sets contains the ball $B(a, \min\{a, 1-a\})$. So every point in $(0, 1)$ is an interior point of each set. Note that any open ball $B(0, r)$ contains $-\frac{r}{2}$ and any open ball $B(1, r)$ contains the point $1 + \frac{r}{2}$. These points are not in any of the sets, so 0 and 1 are not interior points of any of these sets.

(b) Let $r > 0$. Since $B(0, r)$ contains $\min\left\{0.25, \frac{r}{2}\right\}$, $B(1, r)$ contains $\min\left\{1.25, 1 + \frac{r}{2}\right\}$, and $B(2, r)$ contains $\min\left\{2.25, 2 + \frac{r}{2}\right\}$, we see that no point in A is an interior point of A.

(c) Let $q \in \mathbb{Q}$, and let $r > 0$. If r is irrational, the ball $B(q, r)$ contains the irrational number $q + \frac{r}{2}$. If r is rational, then the ball $B(q, r)$ contains the irrational number $q + \frac{r}{2\pi}$. In either case, the ball $B(q, r)$ contains a number that is not in $\mathbb{Q}$. Thus, no point of $\mathbb{Q}$ is an interior point of $\mathbb{Q}$.

Open Sets

Open sets are vitally important in topology. In fact, we will see later that every topological space is completely defined by its open sets. Recall that an open ball is an open set. There are other subsets that every metric space contains, and we might ask if they are open or not.

Activity 7.1. Let X be a metric space.

(a) Is $\emptyset$ an open set in X? Explain.

(b) Is X an open set in X? Explain.

Activity Solution.

(a) Since $\emptyset$ contains no points, it follows that $\emptyset$ is a neighborhood of each of its points. So $\emptyset$ is an open set.

(b) Let $a \in X$. For any $\epsilon > 0$, the ball $B(a, \epsilon)$ is a subset of X. Thus, X is a neighborhood of each of its points and X is an open set in X.

We have defined open balls, and open balls are the canonical examples of open sets. In fact, as the following theorem shows, the open balls determine the open sets.

Theorem 7.3. *Let X be a metric space. A subset O of X is open if and only if O is a union of open balls.*

Proof. Let X be a metric space and O a subset of X. To prove this biconditional statement we first assume that O is an open set and demonstrate that O is a union of open balls. Let $a \in O$. Since O is open, there exists $\epsilon_a > 0$ so that $B(a, \epsilon_a) \subseteq O$. We will show that

$$O = \bigcup_{a \in O} B(a, \epsilon_a).$$

By definition, $B(a, \epsilon_a) \subseteq O$ for every $a \in O$, so $\bigcup_{a \in O} B(a, \epsilon_a) \subseteq O$. For the reverse containment, let $x \in O$. Then $x \in B(x, \epsilon_x)$ and so $x \in \bigcup_{a \in O} B(a, \epsilon_a)$. Thus, $O \subseteq \bigcup_{a \in O} B(a, \epsilon_a)$. We conclude that O is a union of open balls if O is an open set.

The proof of the converse is left for the following activity. ■

Activity 7.2. Let X be a metric space. To prove the remaining implication of Theorem 7.3, assume that a subset O of X is a union of open balls.

(a) What do we need to show to prove that O is an open set?

(b) Let $x \in O$. Why is there an open ball B in O that contains x?

(c) Complete the proof to show that O is an open set.

Activity Solution.

(a) We need to show that O is a neighborhood of each of its points.

(b) Let $x \in O$. Since O is a union of open balls, there must be an open ball B in O that contains x.

(c) Since every ball is a neighborhood of each of its points, there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq B \subseteq O$. Thus, O is a neighborhood of each of its points and is therefore an open set.

Theorem 7.3 tells us that every open set is made up of open balls, so the open balls generate all open sets much like a basis of a vector space in linear algebra generates all of the elements of the vector space. For this reason we call the set of open balls in a metric space a *basis* for the open sets of the metric space. We will discuss this idea in more detail in a subsequent section.

Unions and Intersections of Open Sets

Once we have defined open sets we might wonder about what happens if we take a union or intersection of open sets.

Activity 7.3.

(a) Let $A = (-2, 1)$ and $B = (-1, 2)$ in $(\mathbb{R}, d_E)$.

  i. Is $A \cup B$ open? Explain.
  ii. Is $A \cap B$ open? Explain.

(b) Let $X = \mathbb{R}$ with the Euclidean metric. Let $A_n = \left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right)$ for each $n \in \mathbb{Z}^+$.

  i. What is $\bigcup_{n \geq 1} A_n$? A proof is not necessary.
  ii. Is $\bigcup_{n \geq 1} A_n$ open in $\mathbb{R}$? Explain.
  iii. What is $\bigcap_{n \geq 1} A_n$? A proof is not necessary.
  iv. Is $\bigcap_{n \geq 1} A_n$ open in $\mathbb{R}$? Explain.

Activity Solution.

(a) Let $A = (-2, 1)$ and $B = (-1, 2)$ in $(\mathbb{R}, d_E)$.

i. Let $a \in A \cup B$. Then $a \in (-2, 1)$ or $a \in (-1, 2)$. These open intervals are open balls, and so are neighborhoods of each of their points. Thus, $A \cup B$ is a neighborhood of each of its points.

ii. In this case we have $A \cap B = (-1, 1)$. Since $(-1, 1)$ is an open ball, $(-1, 1)$ is a neighborhood of each of its points. Thus, $A \cap B$ is a neighborhood of each of its points and $A \cap B$ is an open set.

(b) Let $X = \mathbb{R}$ with the Euclidean metric. Let $A_n = \left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right)$ for each $n \in \mathbb{Z}^+$.

i. When $n = 1$ the interval is $(0, 2)$. All intervals for $n > 1$ are contained in $(0, 2)$, so $\bigcup_{n \geq 1} A_n = (0, 2)$.

ii. Since $\bigcup_{n \geq 1} A_n$ is an open ball, it is also an open set $\mathbb{R}$?

iii. Each interval $\left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right)$ contains 1. However, we can make n larger enough so that any integer not equal to 1 is not in $\left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right)$. So $\bigcap_{n \geq 1} A_n = \{1\}$.

iv. There is no open ball of positive radius contained in $\bigcap_{n \geq 1} A_n$, so $\bigcap_{n \geq 1} A_n$ is not open in $\mathbb{R}$.

Activity 7.3 demonstrates that an arbitrary intersection of open sets is not necessarily open. However, there are some things we can say about unions and intersections of open sets.

Theorem 7.4. *Let X be a metric space.*

(1) Any union of open sets in X is an open set in X.

(2) Any finite intersection of open sets in X is an open set in X.

Proof. Let X be a metric space. To prove part 1, assume that $\{O_\alpha\}$ is a collection of open sets in X for α in some indexing set I and let $O = \bigcup_{\alpha \in I} O_\alpha$. By Theorem 7.3, we know that O_α is a union of open balls for each $\alpha \in I$. Combining all of these open balls together shows that O is a union of open balls and is therefore an open set by Theorem 7.3.

For part 2, assume that O_1, O_2, …, O_n are open sets in X for some $n \in \mathbb{Z}^+$. To show that $O = \bigcap_{k=1}^{n} O_k$ is an open set, we will show that O is a neighborhood of each of its points. Let $x \in O$. Then $x \in O_k$ for each $1 \leq k \leq n$. Let k be between 1 and n. Since O_k is open, we know that O_k is a neighborhood of each of its points. So there exists $\epsilon_k > 0$ such that $B(x, \epsilon_k) \subseteq O_k$. Since there are only finitely many values of k, let $\epsilon = \min\{\epsilon_k \mid 1 \leq k \leq n\}$. Then $B(x, \epsilon) \subseteq B(x, \epsilon_k)$ for each k and so $B(x, \epsilon) \subseteq \bigcap_{k=1}^{n} O_k = O$. Therefore, O is a neighborhood of each of its points and O is an open set. ■

Continuity and Open Sets

Recall that we showed that a function f from a metric space (X, d_X) to a metric space (Y, d_Y) is continuous if and only if $f^{-1}(N)$ is a neighborhood of $a \in X$ whenever N is a neighborhood of

$f(a)$ in Y. We can now provide another characterization of continuous functions in terms of open sets. This is the characterization that will serve as our definition of continuity in topological spaces.

Theorem 7.5. *Let f be a function from a metric space (X, d_X) to a metric space (Y, d_Y). Then f is continuous if and only if $f^{-1}(O)$ is an open set in X whenever O is an open set in Y.*

Proof. Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a function. To prove this biconditional statement we need to prove both implications. First assume that f is a continuous function. We must show that $f^{-1}(O)$ is an open set in X for every open set O in Y. So let O be an open set in Y. To demonstrate that $f^{-1}(O)$ is open in X, we will show that $f^{-1}(O)$ is a neighborhood of each of its points. Let $a \in f^{-1}(O)$. Then $f(a) \in O$. Now O is an open set, so there is an open ball $B(f(a), \epsilon)$ around $f(a)$ that is entirely contained in O. Since $B(f(a), \epsilon)$ is a neighborhood of $f(a)$, we know that $f^{-1}(B(f(a), \epsilon))$ is a neighborhood of a. Thus, there exists $\delta > 0$ so that $B(a, \delta) \subseteq f^{-1}(B(f(a), \epsilon))$. Now $f(B(a, \delta)) \subseteq B(f(a), \epsilon) \subseteq O$, and so $B(a, \delta) \subseteq f^{-1}(O)$. We conclude that $f^{-1}(O)$ is a neighborhood of each of its points and is therefore an open set in X.

The proof of the reverse implication is left for the next activity. ■

Activity 7.4. Let f be a function from a metric space (X, d_X) to a metric space (Y, d_Y).

(a) What assumption do we make to prove the remaining implication of Theorem 7.5? What do we need to demonstrate to prove the conclusion?

(b) Let $a \in X$, and let N be a neighborhood of $f(a)$ in Y. Why does there exist an $\epsilon > 0$ so that $B(f(a), \epsilon) \subseteq N$.

(c) What does our hypothesis tell us about $f^{-1}(B(f(a), \epsilon))$ in X?

(d) Why is $f^{-1}(N)$ a neighborhood of a? How does this show that f is a continuous function?

Activity Solution.

(a) We assume that $f^{-1}(O)$ is an open set in X whenever O is an open set in Y. We need to show that for any $a \in X$, $f^{-1}(N)$ is a neighborhood of a whenever N is a neighborhood of $f(a)$.

(b) By definition of a neighborhood, there exists an $\epsilon > 0$ so that $B(f(a), \epsilon) \subseteq N$.

(c) Now $B(f(a), \epsilon)$ is an open set in Y, so $f^{-1}(B(f(a), \epsilon))$ is an open set in X by hypothesis.

(d) Every open set is a neighborhood of each of its points, and $a \in f^{-1}(B(f(a), \epsilon))$, so there exists $\delta > 0$ such that $B(a, \delta) \subseteq f^{-1}(B(f(a), \epsilon))$. Now $f(B(a, \delta)) \subseteq B(f(a), \epsilon) \subseteq N$, so $B(a, \delta) \subseteq f^{-1}(N)$. We conclude that $f^{-1}(N)$ is a neighborhood of a and so f is a continuous function.

The Interior of a Set

Open sets can be characterized in terms of their interior points. By definition, every open set is a neighborhood of each of its points, so every point of an open set O is an interior point of O.

Conversely, if every point of a set O is an interior point, then O is a neighborhood of each of its points and is open. This argument is summarized in the next theorem.

Theorem 7.6. *Let X be a metric space. A subset O of X is open if and only if every point of O is an interior point of O.*

The collection of interior points in a set form a subset of that set, called the *interior* of the set.

Definition 7.7. The **interior** of a subset A of a metric space X is the set

$$\text{Int}(A) = \{a \in A \mid a \text{ is an interior point of } A\}.$$

Activity 7.5. Determine Int(A) for each of the sets A.

(a) $A = (0, 1]$ in $(\mathbb{R}, d_E)$

(b) $A = [0, 1]$ in $(\mathbb{R}, d_E)$

(c) $A = \{-2\} \cup [0, 5] \cup \{7, 8, 9\}$ in $(\mathbb{R}, d_E)$

Activity Solution.

(a) The interior of A is the set $(0, 1)$.

(b) The interior of A is the set $(0, 1)$.

(c) There are no open ball containing -2, 7, 8, or 9 that is entirely contained in A. SotThe interior of A is the set $(0, 5)$.

One might expect that the interior of a set is an open set. This is true, but we can say even more.

Theorem 7.8. *Let (X, d) be a metric space, and let A be a subset of X. Then interior of A is the largest open subset of X contained in A.*

Proof. Let (X, d) be a metric space, and let A be a subset of X. We need to prove that Int(A) is an open set in X, and that Int(A) is the largest open subset of X contained in A. First we demonstrate that Int(A) is an open set. Let $a \in$ Int(A). Then a is an interior point of A, so A is a neighborhood of a. This implies that there exists an $\epsilon > 0$ so that $B(a, \epsilon) \subseteq A$. But $B(a, \epsilon)$ is a neighborhood of each of its points, so every point in $B(a, \epsilon)$ is an interior point of A. It follows that $B(a, \epsilon) \subseteq$ Int(A). Thus, Int(A) is a neighborhood of each of its points and, consequently, Int(A) is an open set.

The proof that Int(A) is the largest open subset of X contained in A is left for the next activity. ■

Activity 7.6. Let (X, d) be a metric space, and let A be a subset of X.

(a) What will we have to show to prove that Int(A) is the largest open subset of X contained in A?

(b) Suppose that O is an open subset of X that is contained in A, and let $x \in O$. What does the fact that O is open tell us?

(c) Complete the proof that $O \subseteq \text{Int}(A)$.

Activity Solution.

(a) We need to prove that any open subset of X that is contained in A is a subset of $\text{Int}(A)$.

(b) The fact that O is an open set tells us that there exists an open ball B centered at x that is contained in O.

(c) Then $B \subseteq A$ and A is a neighborhood of x. So $x \in \text{Int}(A)$. Therefore, $O \subseteq \text{Int}(A)$ and $\text{Int}(A)$ is the largest open subset of X contained in A.

One consequence of Theorem 11.11 is the following.

Corollary 7.9. *A subset O of a metric space X is open if and only if $O = \text{Int}(O)$.*

The proof is left to the reader.

Summary

Important ideas that we discussed in this section include the following.

- A subset O of a metric space (X, d) is an open set if O is a neighborhood of each of its points. Alternatively, O is open if O is a union of open balls.
- A point a in a subset A of a metric space (X, d) is an interior point of A is A is a neighborhood of a. A set O is open if every point of O is an interior point of O.
- The interior of a set is the set of all interior points of the set. The interior of a set A in a metric space X is the largest open subset of X contained in A. A set is open if and only if the set is equal to its interior.
- A function f from a metric space X to a metric space Y is continuous if $f^{-1}(O)$ is open in X whenever O is open in Y.
- Any union of open sets is open, while any finite intersection of open sets is open.

Section 8

Sequences in Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a sequence in a metric space?
- What does it mean for a sequence to have a limit in a metric space?
- How can we use sequences to determine the continuity of a function at a point?

Introduction

We were introduced to sequences in calculus, and we can extend the notion of the limit of a sequence to metric spaces. This will provide an alternate way to understand continuity, and also provide context for other definitions in our metric spaces.

Recall from calculus that a sequence of real numbers is a list of numbers in a specified order. We write a sequence a_1, a_2, $\ldots$, a_n, $\ldots$ as $(a_n)_{n\in\mathbb{Z}^+}$ or just (a_n). If we think of each a_n as the output of a function, we can give a more formal definition of a sequence as a function $f : \mathbb{Z}^+ \to \mathbb{R}$, where $a_n = f(n)$ for each n.

A sequence (a_n) of real numbers converges to a number L if we can make all of the numbers in the sequence as close to L as we like by choosing n to be large enough. Once again, this is an informal description that we need to make more rigorous. As we saw with continuous functions, we can make more rigorous the idea of "closeness" by introducing a symbol for a number that can be arbitrarily small. So we can say that the numbers a_n can get as close to a number L as we want if we can make $|a_n - L| < \epsilon$ for any positive number ϵ. The idea of choosing n large enough is just finding a large enough fixed integer N so that $|a_n - L| < \epsilon$ whenever $n \geq N$. This leads to the definition.

Definition 8.1. A sequence (a_n) of real numbers has a **limit** L if, given any $\epsilon > 0$ there exists a positive integer N such that

$$|a_n - L| < \epsilon \text{ whenever } n \geq N.$$

When a sequence (a_n) has a limit L, we write

$$\lim_{n\to\infty} a_n = L,$$

or just $\lim a_n = L$ (since we assume the limit for a sequence occurs as n goes to infinity) and we say that the sequence (a_n) *converges* to L.

Example 8.2. We can draw a graph of a sequence (a_n) of real numbers as the set of points (n, a_n). In this way we can visualize a sequence and its limit. By definition, L is a limit of the sequence (a_n) if, given any $\epsilon > 0$, we can go far enough out in the sequence so that the numbers in the sequence all lie in the horizontal band between $y = L - \epsilon$ and $L + \epsilon$ as illustrated in Figure 8.1 for the sequence $\left(\frac{n}{1+n}\right)$. To verify that the limit of the sequence $\left(\frac{n}{1+n}\right)$ is 1, we start with $\epsilon > 0$.

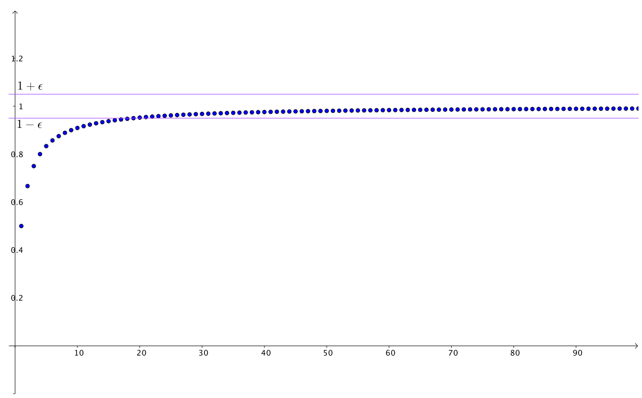


Figure 8.1: The limit of the sequence $\left(\frac{n}{1+n}\right)$.

Scratch work. Now we need to find N so that $n \geq N$ implies $\left|\frac{n}{1+n} - 1\right| < \epsilon$. Just as with our continuity example, this work is not part of the proof, but shows how we go about finding the N we need. To make $\left|\frac{n}{1+n} - 1\right| < \epsilon$ we need

$$\begin{aligned}
\left|\frac{n}{1+n} - 1\right| &< \epsilon \\
\left|\frac{n}{1+n} - \frac{1+n}{1+n}\right| &< \epsilon \\
\left|\frac{-1}{1+n}\right| &< \epsilon \\
1 + n &> \frac{1}{\epsilon} \\
n &> \frac{1}{\epsilon} - 1.
\end{aligned}$$

Now we ignore this paragraph and resume the proof.

Let $N > \frac{1}{\epsilon} - 1$. Then for $n \geq N$ we have

$$\begin{aligned} n > N > \frac{1}{\epsilon} - 1 \\ 1 + n > \frac{1}{\epsilon} \\ |-1|\left|\frac{1}{1+n}\right| < \epsilon \\ \left|\frac{-1}{1+n}\right| < \epsilon \\ \left|\frac{n}{1+n} - 1\right| < \epsilon \\ \left|\frac{n}{1+n} - \frac{1+n}{1+n}\right| < \epsilon. \end{aligned}$$

So the sequence $\left(\frac{n}{1+n}\right)$ has a limit of 1.

Definition 8.1 only applies to sequences of real numbers. Ultimately, we want to phrase the definition in a way that allows us to define limits of sequences in metric spaces and topological spaces. So we have to reformulate the definition in such a way that it does not depend on distances.

Recall that $|x - y|$ defined a metric d_E on $\mathbb{R}$, that is

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

So we can rephrase the definition of a limit of a sequence of real numbers as follows.

Definition 8.3 (Alternate Definition)**.** A sequence (a_n) of real numbers has a **limit** L if, given any $\epsilon > 0$ there exists a positive integer N such that

$$d_E(a_n, L) < \epsilon \text{ whenever } n \geq N.$$

Once we have described a limit of a sequence in terms of a metric, then we can extend the idea into any metric space.

Definition 8.4. A **sequence** in a metric space (X, d) is a function $f : \mathbb{Z}^+ \to X$.

If f is a sequence in X, we write the sequence defined by f as $(f(n))$, where $n \in \mathbb{Z}^+$. We also use the notation (a_n), when $a_n = f(n)$. As long as X has a metric defined on it, we can then describe the limit of a sequence.

Definition 8.5. Let (X, d) be a metric space. A sequence (a_n) in X has a **limit** $L \in X$ if, given any $\epsilon > 0$ there exists a positive integer N such that

$$d(a_n, L) < \epsilon \text{ whenever } n \geq N.$$

In other words, a sequence (a_n) in a metric space (X, d) has a limit $L \in X$ if $\lim d(a_n, L) = 0$ – or that the sequence $d(a_n, L)$ of real numbers has a limit of 0. Just as with sequences of real numbers, when a sequence (a_n) has a limit L, we say that the sequence (a_n) *converges* to L, or that L is a limit of the sequence (a_n).

Preview Activity 8.1.

(1) Explain why the sequence $\left(\frac{1}{n}\right)$ converges to 0 in $\mathbb{R}$ using the Euclidean metric d_E, where

$$d_E(a, y) = |x - y|.$$

(2) Consider the sequence $(a_n) = \left(\left(\frac{1}{n}, \frac{1}{n+1}\right)\right)$ in $(\mathbb{R}^2, d_T)$, where d_T is the taxicab metric

$$d_T((x_1, x_2), (y_1, y_1)) = |x_1 - y_1| + |x_2 - y_2|.$$

Does the sequence (a_n) converge? If so, find its limit and prove that your candidate is the limit. If not, explain why.

(3) Let $(b_n) = ((2n, n^2))$ in the metric space $(\mathbb{R}^2, d)$, where d is the discrete metric defined by

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

Does the sequence (b_n) converge? If so, find its limit and prove that your candidate is the limit. If not, explain why.

Activity Solution.

(1) Let $a_n = \frac{1}{n}$ for each positive integer n. Let $\epsilon > 0$. We will show that there is a positive integer N such that $n \geq N$ implies $d_E(a_n, 0) < \epsilon$. Since the set of positive integers is unbounded above, there exists a positive integer N so that $\frac{1}{N} < \epsilon$. Then, if $n \geq N$, it follows that

$$d_E(a_n, 0) = |a_n - 0| = |\frac{1}{n}| \leq \frac{1}{N} < \epsilon.$$

We conclude that $lim \frac{1}{n} = 0$.

(2) We will show that $\lim a_n = (0, 0)$. Let $\epsilon > 0$. Choose $N \in \mathbb{Z}^+$ so that $N > \frac{2}{\epsilon}$. This makes $\frac{1}{N} < \frac{\epsilon}{2}$. Note that $\frac{1}{N+1} < \frac{1}{N} < \frac{\epsilon}{2}$ as well. Now let $n \geq N$. Then

$$\begin{aligned} d_T(a_n, (0, 0)) &= \left|\frac{1}{n} - 0\right| + \left|\frac{1}{n+1} - 0\right| \\ &= \frac{1}{n} + \frac{1}{n+1} \\ &\leq \frac{1}{N} + \frac{1}{N+1} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

(3) We claim that the sequence (b_n) does not have a limit. To demonstrate this, we proceed by contradiction and assume that the sequence (b_n) has a limit b. So for any $\epsilon > 0$ there exists an integer $N > 0$ so that $n \geq N$ implies $d(b_n, b) < \epsilon$. Let $\epsilon = \frac{1}{2}$. Now $(2n, n^2) \neq (2m, m^2)$ if $m \neq n$ in $\mathbb{Z}$, so all of the entries of the sequence (b_n) are distinct. Now either $b = b_n$ for some n or $b \neq b_n$ for all n. In either case there is a large enough N so that $b \neq b_n$ for all $n \geq N$. Then $n \geq N$ implies

$$d(b_n, b) = 1 > \epsilon.$$

So there cannot be an integer $N > 0$ so that $d(b_n, b) < \epsilon$ for $n \geq N$, no matter what b is, so the sequence (b_n) does not have a limit.

Sequences and Continuity in Metric Spaces

There are different characterizations of continuity. We have already see the ϵ-δ definition and a characterization in terms of neighborhoods. In this section we investigate sequences and limits of sequences in metric spaces, and then provide a characterization of continuous functions in terms of sequences.

Activity 8.1. A reasonable question to ask is if a limit of a sequence is unique. We will answer that question in this activity. Let (X, d) be a metric space and (a_n) a sequence in X. Assume the sequence (a_n) has a limit in X. To show that a limit of the sequence (a_n) is unique, we need to show that if $\lim a_n = a$ and $\lim a_n = a'$ for some $a, a' \in X$, then $a = a'$.

Suppose $\lim a_n = a$ and $\lim a_n = a'$ for some $a, a' \in X$. Without much to go on it might appear that proving $a = a'$ is a difficult task. However, if $d(a, a') < \epsilon$ for any $\epsilon > 0$, then it will have to be the case that $a = a'$. So let $\epsilon > 0$.

(a) Why must there exist a positive integer N so that $d(a_n, a) < \frac{\epsilon}{2}$ for all $n \geq N$?

(b) Why must there exist a positive integer N' so that $d(a_n, a') < \frac{\epsilon}{2}$ for all $n \geq N'$?

(c) Now let $m = \max\{N, N'\}$. What can we say about $d(a_m, a)$ and $d(a_m, a')$? Why?

(d) Use the triangle inequality to conclude that $d(a, a') < \epsilon$. What else can we conclude?

Activity Solution.

(a) Since a is a limit of (a_n), the definition of a limit of a sequence states that, given $\frac{\epsilon}{2}$ (using $\frac{\epsilon}{2}$ in place of ϵ), there exists a positive integer N so that $d(a_n, a) < \frac{\epsilon}{2}$ for all $n \geq N$.

(b) Since a' is a limit of (a_n), the definition of a limit of a sequence states that, given $\frac{\epsilon}{2}$ (using $\frac{\epsilon}{2}$ in place of ϵ), there exists a positive integer N' so that $d(a_n, a') < \frac{\epsilon}{2}$ for all $n \geq N'$.

(c) Since m is larger than N we know that $d(a_m, a) < \frac{\epsilon}{2}$. Since m is larger than N' we know that $d(a_m, a') < \frac{\epsilon}{2}$.

(d) We now have that

$$d(a, a') \leq d(a, a_m) + d(a_m, a') < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

From this we can conclude that $a = a'$.

Continuity can also be described in terms of sequences. The basic idea is this. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point a. This means that f has a limit (as a continuous function) at a. So if we were to take any sequence (a_n) that converges to a, then the continuity of f implies that $f(a) = f(\lim a_n) = \lim f(a_n)$. That this is both a necessary condition and a sufficient condition for continuity is given in the next theorem.

Theorem 8.6. *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $a \in X$. A function $f : X \to Y$ is continuous at a if and only if $\lim f(a_n) = f(a)$ for any sequence (a_n) in X that converges to a.*

Proof. Let (X, d_X) and (Y, d_Y) be metric spaces, let $a \in X$, and let $f : X \to Y$ be a function. Assume that f is continuous at a. We will show that $\lim f(a_n) = f(a)$ for any sequence (a_n) in X that converges to a. Let (a_n) be a sequence in X that converges to a (we know such a sequence exists, namely the sequence (a)). To verify that $\lim f(a_n) = a$, let $\epsilon > 0$. The fact that f is continuous at a means that there is a $\delta > 0$ so that $d_Y(f(x), f(a)) < \epsilon$ whenever $d_X(x, a) < \delta$. Since (a_n) converges to a, we know that there exists a positive integer N such that $d_X(a_n, a) < \delta$ whenever $n \geq N$. This implies that

$$d_Y(f(a_n), f(a)) < \epsilon \text{ whenever } n \geq N.$$

We conclude that if f is continuous at a, then $\lim f(a_n) = f(a)$ for any sequence (a_n) in X that converges to a.

The proof of the reverse implication is contained in the next activity. ■

Activity 8.2. Let (X, d_X) and (Y, d_Y) be metric spaces, let $a \in X$, and let $f : X \to Y$ be a function. We prove the remaining implication of Theorem 8.6, that f is continuous at a if $\lim f(a_n) = f(a)$ for any sequence (a_n) in X that converges to a, in this activity.

(a) To have an additional assumption with which to work, let us proceed by contradiction and assume that f is not continuous at a. Why can we then say that there is an $\epsilon > 0$ so that there is no $\delta > 0$ with the property that $d_X(x, a) < \delta$ implies $d_Y(f(x), f(a)) < \epsilon$?

(b) To create a contradiction, we will construct a sequence (a_n) that converges to a while $(f(a_n))$ does not converge to $f(a)$.

  i. Explain why we can find a positive integer K such that $\frac{1}{K} < \epsilon$.

  ii. If $k > K$, explain why there is an element $a_k \in B\left(a, \frac{1}{k}\right)$ so that $d_Y(f(a_k), f(a)) \geq \epsilon$.

  iii. For $k \leq K$, let a_k be any element in $B\left(a, \frac{1}{k}\right)$. Explain why a is a limit of (a_n).

  iv. Explain why $f(a)$ is not a limit of the sequence $(f(a_n))$. What conclusion can we draw, and why?

Activity Solution.

(a) Since f is not continuous at a, we negate the definition of continuity, negating universal and existential quantifiers. This tells us that exists an $\epsilon > 0$ such that there is no $\delta > 0$ with the property that $d_X(x, a) < \delta$ implies $d_Y(f(x), f(a)) < \epsilon$.

(b) i. The sequence $\frac{1}{k}$ of real numbers converges to 0, so we can make $\frac{1}{k}$ as close to 0 as we like. That means we can make $\frac{1}{k}$ less than ϵ if k is large enough. So there is some positive integer K such that $\frac{1}{K} < \epsilon$.

  ii. From part (a), for each $k > K$ there is no δ with the property that $d_X(x, a) < \delta$ implies $d_Y(f(x), f(a)) < \epsilon$. Thus, there must be some element a_k with $d_X(a_k, a) < \frac{1}{k}$ and $d_Y(f(a_k), f(a)) > \epsilon$. Since $d_X(a_k, a) < \frac{1}{k}$, we know that $a_k \in B\left(a, \frac{1}{k}\right)$.

  iii. Choosing any $a_k \in B\left(a, \frac{1}{k}\right)$ when $k \leq K$ gives us a sequence (a_n). We now show that a is the limit of (a_n). Let $\alpha > 0$. Let N be a positive integer such that $\frac{1}{N} < \alpha$. If

$n \geq N$, since $a_n \in B\left(a, \frac{1}{n}\right)$, we see that

$$d_X(a_n, a) < \frac{1}{n} < \frac{1}{N} < \alpha.$$

Therefore, the sequence (a_n) has a limit of a.

iv. Now we demonstrate that $(f(a_n))$ does not have $f(a)$ as a limit. Using ϵ as given, note that $n > K$ implies that $d_Y(f(a_k), f(a)) \geq \epsilon$. We can always choose n larger enough so that $\frac{1}{n} < \delta$ for any given $\delta > 0$. So there can be no δ such that $d(x, a) < \delta$ implies $d_Y(f(a_k), f(a)) < \epsilon$.

v. Assuming that f is not continuous at a led to the construction of a sequence (a_n) with limit a, but $\lim f(a_n) \neq f(a)$. This contradicts our hypothesis. We must therefore conclude that f is continuous at a.

Note that Theorem 8.6 tells us that if $f : X \to Y$ is a continuous function, then f commutes with limits. That is, if (a_n) is a sequence in X that converges to $a \in X$, then

$$f(a) = f(\lim a_n) = \lim f(a_n).$$

Summary

Important ideas that we discussed in this section include the following.

- A sequence in a metric space X is a function $f : \mathbb{Z}^+ \to X$.
- A sequence (a_n) in a metric space (X, d) has a limit L in X if, given any $\epsilon > 0$ there exists a positive integer N such that $d(a_n, L) < \epsilon$ whenever $n \geq N$.
- Let f be a function from a metric space (X, d_X) to m metric space (Y, d_Y). Then f is continuous at $a \in X$ if and only if $\lim f(a_n) = f(a)$ for any sequence (a_n) in X that converges to a.

Section 9

Closed Sets in Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What are boundary points, limit points, and isolated points of a set in a metric space? How are they related and how are they different?
- What does it mean for a set to be closed in a metric space?
- What important properties do closed sets have in relation to unions and intersections?
- How can we use closed sets to determine the continuity of a function?
- How are limit points related to sequences?
- How are boundary points related to sequences?
- What is the boundary of a set in a metric space?
- How are limit points and boundary points related to closed sets?
- What is the closure of a set in a metric space?
- How are closed sets related to sequences?

Introduction

Once we have defined open sets in metric spaces, it is natural to ask if there are closed sets. Recall that closed intervals are important in calculus because every continuous function on a closed interval attains an absolute maximum and absolute minimum value on that interval. If we have closed sets in metric spaces, we might consider if there is some result that is similar to this for continuous functions on closed sets. In this section we introduce the idea of closed sets in metric spaces and discover a few of their properties.

Every interval of the form $[a, b]$ in $\mathbb{R}$ is a closed set using the Euclidean metric. What distinguishes these closed intervals from the open intervals is that the open intervals do not contain either of their endpoints – this is what makes an open interval a neighborhood of each of its points. In general, what makes open sets open is that they do not contain their boundaries. If an open set doesn't contain its boundary, then its complement, by contrast, should contain its boundary. This leads to the definition of a closed set.

Definition 9.1. A subset C of a metric space X is **closed** if its complement $X \setminus C$ is open.

We said that open sets are open because they do not contain their boundary and closed sets are closed because they do contain their boundary. However, we did not define what we mean by boundary. The point a on the "boundary" of an open interval of the form $O = (a, b)$ in $\mathbb{R}$ with the Euclidean metric has the property that every open ball that contains a contains points in O and points not in O. This is what makes the point a lie on the boundary. We can also think of the point a as being at the very limit of the set O. This motivates the next definition.

Definition 9.2. Let X be a metric space, and let A be a subset of X. A **boundary point** of A is a point $x \in X$ such that every neighborhood of x contains a point in A and a point in $X \setminus A$.

For example, in $A = (0, 1)$ as a subset of $(\mathbb{R}, d_E)$, the number 0 is a boundary point of A because any open interval in $\mathbb{R}$ that contains 0 contains points in A and points not in A. Boundary points can arise in other ways. If $A = \{0, 1\}$ as a subset of $(\mathbb{R}, d_E)$, then 0 is again a boundary point because any open interval in $\mathbb{R}$ that contains 0 contains a point (0) in A and points not in A. However, 0 is the only point in A that is contained in any open interval. In this case we call 0 an *isolated point* of A, and in the case of the set $A = (0, 1)$ we call 0 an *accumulation point* or a *limit point* of A (the use of the word "limit" here will become clear later).

Definition 9.3. Let X be a metric space, and let A be a subset of X.

(1) An **accumulation point** or **limit point** of A is a point $x \in X$ such that every neighborhood of x contains a point in A different from x.

(2) An **isolated** of A is a point $a \in A$ such that there exists a neighborhood N of a in X with $N \cap A = \{a\}$.

Note that every boundary point is either an accumulation point or an isolated point. The proof is left to the reader.

Preview Activity 9.1.

(1) For each of the given sets A, find all boundary points, limit points, and isolated points. Then determine if the set A is a closed set in the metric space (X, d). Explain your reasoning.

   (a) $X = \mathbb{R}$, $d = d_E$, the Euclidean metric, $A = [0, 0.5)$.

   (b) $X = \{x \in \mathbb{R} \mid 0 < x \leq 1\}$, $d = d_E$, the Euclidean metric, $A = (0, 0.5]$.

(c) $X = \{a, b, c, d\}$, d is the discrete metric defined by

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

and $A = \{a, b\}$.

(2) Label each of the following statements as either true or false. If true, provide a convincing argument. If false, provide a specific counterexample.

(a) Every limit point is a boundary point.

(b) Every boundary point is a limit point.

(c) Every limit point is an isolated point

(d) Every isolated point is a limit point.

(e) Every boundary point is an isolated point.

(f) Every isolated point is a boundary point.

Activity Solution.

(1) For each of the given sets A, find all boundary points, limit points, and isolated points. Then determine if the set A is a closed set in the metric space (X, d). Explain your reasoning.

(a) The only points in X for which every neighborhood contains points in A and points not in A are 0 and 0.5, so 0 and 0.5 are the boundary points of A. If $x \in [0, 0.5]$, then every neighborhood of x contains a point in A different from x, so every x with $0 \leq x \leq 0.5$ is a limit point of A. There are no isolated points of A.

The set A is not a closed set in X. Note that $\mathbb{R} \setminus A = (-\infty, 0) \cup [0.5, \infty)$. No matter what value $0 < \delta < 0.5$ has, the open ball $B(0.5, \delta)$ contains the point $0.5 - \frac{\delta}{2}$, which is not contained in $\mathbb{R} \setminus A$. Since $\mathbb{R} \setminus A$ is not a neighborhood of 0.5, the set $\mathbb{R} \setminus A$ is not open. If follows that A is not closed in $\mathbb{R}$.

(b) The only point in X for which every neighborhood contains points in A and points not in A is 0.5, so 0.5 is the only boundary point of A. If $x \in (0, 0.5]$, then every neighborhood of x contains a point in A different from x, so every x with $0 < x \leq 0.5$ is a limit point of A. There are no isolated points of A.

The set A is a closed set in X. In this case $X \setminus A = (0, 1]$. Note that every open ball in X must be contained in X, so the open ball of radius 0.25 in X centered at 1 has the form $(0.75, 1]$, which is entirely contained in $X \setminus A$. If a is any point of $X setminus A$ other than 0, then the open ball centered at a of radius $\min\{a, 0.5 - a\}$ is contained in $X \setminus A$. So $A \setminus A$ is a neighborhood of each of its points and $X \setminus A$ is an open set. Therefore, A is a closed set in X.

(c) If $x \in X$ and $r < 1$, then $B(x, r) = \{x\}$. So A has no boundary points. For the same reason, A has no limit points, and every point of A is an isolated point.

Notice that if $x \in X \setminus = \{c, d\}$, then $B(x, 0.5) = \{a\}$. So $X \setminus A$ contains the open balls $B(c, 0.5)$ and $B(d, 0.5)$. Thus, $X \setminus A$ is a neighborhood of each of its points and $X \setminus A$ is an open set. It follows that A is a closed set in X.

(2) Label each of the following statements as either true or false. If true, provide a convincing argument. If false, provide a specific counterexample.

(a) This statement is false. Let $A = (0, 1)$ in $(\mathbb{R}, d_E)$. Then $B(0.5, 0.25)$ contains points in A different from 0.5, but no points not in A. So 0.5 is a limit point of A but not a boundary point of A.

(b) This statement is false. Consider $X = (\mathbb{R}, d_E)$ and $A = \{0\} \cup (1, 2)\}$. Note that $B(0, 0.5) \cap A = \{0\}$, so 0 is not a limit point of A. However, if $\epsilon > 0$, then $B(0, \epsilon)$ contain a point (namely 0) in A and points not in A.

(c) This statement is false. Let $A = (0, 1)$ in $(\mathbb{R}, d_E)$. If $\epsilon > 0$, the open ball $B(0, \epsilon)$ contains points in A different from 0. So 0 is a limit point of A, but not an isolated point of A.

(d) This statement is false. Let $A = \{0, 1\}$ in $(\mathbb{R}, d_E)$. Then $B(0, 0.5) \cap A = \{0\}$, so 0 is an isolated point of A, but not a limit point of A.

(e) This statement is false. Let $A = (0, 1)$ in $(\mathbb{R}, d_E)$. If $\epsilon > 0$, the open ball $B(0, \epsilon)$ contains points in A different from 0 and points not in A. So 0 is a boundary point of A, but not an isolated point of A.

(f) This statement is false. Let $X = \{0, 1, 2\}$ with the discrete metric, and let $A = \{0, 1\}$. Then $B(0, 0.5) \cap A = \{0\}$, so 0 is an isolated point. But $B(0, 0.5)$ contains no points in $X \setminus A$, so 0 is not a boundary point of A.

Closed Sets in Metric Spaces

Recall that Definition 9.1 defines a closed set in a metric space to be a set whose complement is open. We have seen that both $\emptyset$ and X are open subsets of X. We now ask the same question, this time in terms of closed sets.

Activity 9.1. Let X be a metric space.

(a) Is $\emptyset$ closed in X? Explain.

(b) Is X closed in X? Explain.

Activity Solution.

(a) Since $X \setminus \emptyset = X$ and X is open in X, it follows that $\emptyset$ is closed in X.

(b) Since $X \setminus X = \emptyset$ and $\emptyset$ is open in X, it follows that X is closed in X.

Note that a subset of a metric space can be both open and closed. We call such sets *clopen* (for closed-open). When we discussed open sets, we saw that an arbitrary union of open sets is open, but that an arbitrary intersection of open sets may not be open. Since closed sets are complements of open sets, we should expect a similar result for closed sets.

Activity 9.2. Let $X = \mathbb{R}$ with the Euclidean metric. Let $A_n = \left[\frac{1}{n}, 1 - \frac{1}{n}\right]$ for each $n \in \mathbb{Z}^+$, $n \geq 2$.

(a) What is $\bigcup_{n \geq 2} A_n$? A proof is not necessary.

(b) Is $\bigcup_{n \geq 2} A_n$ closed in $\mathbb{R}$? Explain.

Activity Solution.

(a) As we let n go to infinity, $1 - \frac{1}{n}$ will approach 1 but not reach 1, and $\frac{1}{n}$ will approach 0 but not reach 0. So $\bigcup_{n \geq 1} A_n = (0, 1)$.

(b) The complement of $\bigcup_{n \geq 1} A_n$ in $\mathbb{R}$ is $(-\infty, 0] \cup [1, \infty)$, and 0 is not an interior point. So $(-\infty, 0] \cup [1, \infty)$ is not open and $\bigcup_{n \geq 1} A_n$ is not closed in $\mathbb{R}$.

Activity 9.2 shows that an arbitrary union of closed sets is not necessarily closed. However, the following theorem tells us what we can say about unions and intersections of closed sets. The results should not be surprising given the relationship between open and closed sets.

Theorem 9.4. *Let X be a metric space.*

(1) Any intersection of closed sets in X is a closed set in X.

(2) Any finite union of closed sets in X is a closed set in X.

Proof. Let X be a metric space. To prove part 1, assume that $\{C_\alpha\}$ is a collection of closed sets in X for α in some indexing set I. DeMoivre's Theorem shows that

$$X \setminus \bigcap_{\alpha \in I} C_\alpha = \bigcup_{\alpha \in I} X \setminus C_\alpha.$$

The latter is an arbitrary union of open sets and so it an open set. By definition, then, $\bigcap_{\alpha \in I} C_\alpha$ is a closed set.

For part 2, assume that C_1, C_2, $\ldots$, C_n are closed sets in X for some $n \in \mathbb{Z}^+$. To show that $C = \bigcap_{k=1}^{n} C_k$ is a closed set, we will show that $X \setminus C$ is an open set. Now

$$X \setminus \bigcup_{i=1}^{n} C_i = \bigcap_{i=1}^{n} X \setminus C_i$$

is a finite intersection of open sets, and so is an open set. Therefore, $\bigcup_{i=1}^{n} C_i$ is a closed set. ∎

Continuity and Closed Sets

Recall that we showed that a function f from a metric space (X, d_X) to a metric space (Y, d_Y) is continuous if and only if $f^{-1}(O)$ is open for every open set O in Y. We might conjecture that a similar result holds for closed sets. Since closed sets are complements of open sets, to make this connection we will want to know how $X \setminus f^{-1}(B)$ is related to $f^{-1}(Y \setminus B)$ for $B \subset Y$.

Activity 9.3. Let f be a function f from a metric space (X, d_X) to a metric space (Y, d_Y), and let B be a subset of Y.

(a) Let $x \in X \setminus f^{-1}(B)$.

i. What does this tell us about $f(x)$?

ii. What can we conclude about the relationship between $X \setminus f^{-1}(B)$ and $f^{-1}(Y \setminus B)$?

(b) Let $x \in f^{-1}(Y \setminus B)$.

i. What does this tell us about $f(x)$?

ii. What can we conclude about the relationship between $X \setminus f^{-1}(B)$ and $f^{-1}(Y \setminus B)$?

(c) What is the relationship between $X \setminus f^{-1}(B)$ and $f^{-1}(Y \setminus B)$?

Activity Solution.

(a) Let $x \in X \setminus f^{-1}(B)$.

i. Then $x \in X$ but $x \notin f^{-1}(B)$. That means $x \in X$ but $f(x) \notin B$.

ii. Since $f(x) \notin B$, $f(x) \in (Y \setminus B)$ or $x \in f^{-1}(Y \setminus B)$. Thus, $(X \setminus f^{-1}(B)) \subseteq f^{-1}(Y \setminus B)$.

(b) Let $x \in f^{-1}(Y \setminus B)$.

i. Then $f(x) \in (Y \setminus B)$. So $f(x) \notin f^{-1}(B)$.

ii. We conclude that $f(x) \in (X \setminus f^{-1}(B))$. Thus, $f^{-1}(Y \setminus B) \subseteq (X \setminus f^{-1}(B))$.

(c) The two containments in (a) and (b) show that $X \setminus f^{-1}(B) = f^{-1}(Y \setminus B)$.

Now we can consider the issue of continuity and closed sets.

Activity 9.4. Let f be a function from a metric space (X, d_X) to a metric space (Y, d_Y).

(a) Assume that f is continuous and that C is a closed set in Y. How does the result of Activity 9.3 tell us that $f^{-1}(C)$ is closed in X?

(b) Now assume that $f^{-1}(C)$ is closed in X whenever C is closed in Y. How does the result of Activity 9.3 tell us that f is a continuous function?

Activity Solution.

(a) Since C is closed, we know that $Y \setminus C$ is open. This means that $f^{-1}(Y \setminus C)$ is also open. Activity 9.3 tell us that $f^{-1}(Y \setminus B) = X \setminus f^{-1}(C)$. The fact that $X \setminus f^{-1}(C)$ is open implies that $f^{-1}(C)$ is closed.

(b) Suppose O is an open set in Y. Then $Y \setminus O$ is a closed set. Activity 9.3 tells us that $f^{-1}(Y \setminus O) = X \setminus f^{-1}(O)$. That $X \setminus f^{-1}(O)$ is closed means that $f^{-1}(O)$ is open. Thus, f is a continuous function.

The result of Activity 9.4 is summarized in the following theorem.

Theorem 9.5. *Let f be a function from a metric space (X, d_X) to a metric space (Y, d_Y). Then f is continuous if and only if $f^{-1}(C)$ is closed in X whenever C is a closed set in Y.*

Limit Points, Boundary Points, Isolated Points, and Sequences

Recall that a limit point of a subset A of a metric space X is a point $x \in X$ such that every neighborhood of x contains a point in A different from x. You might wonder about the use of the word "limit" in the definition of limit point. The next activity should make this clear.

Activity 9.5. Let X be a metric space, let A be a subset of X, and let x be a limit point of A.

(a) Let $n \in \mathbb{Z}^+$. Explain why $B\left(x, \frac{1}{n}\right)$ must contain a point a_n in A different from x.

(b) What is $\lim a_n$? Why?

Activity Solution.

(a) The fact that $B\left(x, \frac{1}{n}\right)$ is a neighborhood of x implies that $B\left(x, \frac{1}{n}\right)$ must contain a point a_n in A different from x.

(b) Given any $\epsilon > 0$, we can choose N such that $\frac{1}{N} < \epsilon$. So $n \geq N$ implies that $d(a_n, x) < \frac{1}{n} < \frac{1}{N} < \epsilon$. So $\lim a_n = x$.

The result of Activity 9.5 is summarized in the following theorem.

Theorem 9.6. *Let X be a metric space, let A be a subset of X, and let x be a limit point of A. Then there is a sequence (a_n) in A that converges to x.*

Of course, the constant sequence (a) always converges to the point a, so every point in a set A is the limit of a sequence. With limit points there is a non-constant sequence that converges to the point. We might ask what we can say about a point $a \in A$ if the only sequences in A that converges to $a \in A$ are the eventually constant sequences (a). (By an eventually constant sequence (a_n), we mean that there is a positive integer K such that for $k \geq K$, we have $a_k = a$ for some element a.) That is the subject of our next activity.

Activity 9.6. Let (X, d) be a metric space, and let A be a subset of X.

(a) Let a be an isolated point of A. Prove that the only sequences in A that converge to a are the essentially constant sequences (a).

(b) Prove that if the only sequences in A that converges to a are the essentially constant sequences (a), then a is an isolated point of A.

Activity Solution.

(a) Suppose (a_n) is a sequence in A that converges to a. Let N be a neighborhood of a in X such that $A \cap N = \{a\}$. Since N is a neighborhood of a, there is an open ball $B(a, \epsilon)$ that is a subset of N. The fact that (a_n) converges to a means that there is a positive integer K such that $k \geq K$ implies that $d(a_k, a) < \epsilon$. Then $a_k \in (A \cap N)$ and so $a_k = a$.

(b) We prove the contrapositive. Assume that a is not an isolated point of A. We will show that there is a sequence in A that converges to a that is not eventually constant. For each positive integer n, we know that $B\left(x, \frac{1}{n}\right)$ is a neighborhood of a. Since a is not a limit point of A, it follows that $B\left(x, \frac{1}{n}\right) \cap A$ contains an element a_n of A different from a.But then (a_n) is a sequence in A that converges to a with $a_n \neq a$ for every n.

Boundary points are points that are, in some sense, situated "between" a set and its complement. We will make this idea of "between" more concrete soon.

An argument just like the one in Activity 9.5 gives us the following result about boundary points.

Theorem 9.7. *Let X be a metric space, let A be a subset of X, and let b be a boundary point of A. Then there are sequences (x_n) in $X \setminus A$ and (a_n) in A that converge to x.*

Limit Points and Closed Sets

There is a connection between limit points and closed sets. The open set $(1, 2)$ in $(\mathbb{R}, d_E)$ does not contain all of its limit points or any of its boundary points, while the closed set $[1, 2]$ contains all of its boundary and limit points. This is an important attribute of closed sets. Recall that for a limit point a of a subset A of a metric space X, there are sequences in $X \setminus A$ and A that converge to a. So, in a sense, the limit points that are not in A are the points in X that are arbitrarily close to the set A. We denote the set of limit points of A as A', and the limit points of a set can tell us if the set is closed.

Theorem 9.8. *Let C be a subset of a metric space X, and let C' be the set of limit points of C. Then C is closed if and only if $C' \subseteq C$.*

Proof. Let X be a metric space, and let C be a subset of X. First we assume that C is closed and show that C contains all of its limit points. Let $x \in X$ be a limit point of C. We proceed by contradiction and assume that $x \notin C$. Then $x \in X \setminus C$, which is an open set. This implies that there is an $\epsilon > 0$ so that $B(x, \epsilon) \subseteq X \setminus C$. But then this neighborhood $B(x, \epsilon)$ contains no points in C, which contradicts the fact that x is a limit point of C. We conclude that $x \in C$ and C contains all of its limit points.

The converse of the result we just proved is the subject of the next activity.

■

Activity 9.7. Let C be a subset of a metric space X, and let C' be the set of limit points of C. In this activity we prove that C is closed if C contains all of its limit points. So assume $C' \subseteq C$.

(a) What do we need to do to show that C is closed?

(b) If we proceed by contradiction to prove that C is closed, what would be true about $X \setminus C$?

(c) What does the conclusion of part (b) tells us?

(d) What contradiction can we reach to conclude our proof?

Activity Solution.

(a) We need to show that $X \setminus C$ is open.

(b) To proceed by contradiction, we assume that $X \setminus C$ is not open.

(c) Then there exists $x \in X \setminus C$ such that no neighborhood of x is entirely contained in $X \setminus C$.

(d) This implies that every neighborhood of x contains a point in C and so x is a limit point of C. It follows that $x \in C$, contradicting the fact that $x \in X \setminus C$. We conclude that $X \setminus C$ is open and C is closed.

The Closure of a Set

We have seen that the interior of a set is the largest open subset of that set. There is a similar result for closed sets. For example, let $A = (0, 1)$ in $(\mathbb{R}, d_E)$. The set A is an open set, but if we union A with its limit points, we obtain the closed set $C = [0, 1]$. Moreover, The set $[0, 1]$ is the smallest closed set that contains A. This leads to the idea of the *closure* of a set.

Definition 9.9. The **closure** of a subset A of a metric space X is the set

$$\overline{A} = A \cup A'.$$

In other words, the closure of a set is the collection of the elements of the set and the limit points of the set – those points that are on the "edge" of the set. The importance of the closure of a set A is that the closure of A is the smallest closed set that contains A.

Theorem 9.10. *Let X be a metric space and A a subset of X. The closure of A is a closed set. Moreover, the closure of A is the smallest closed subset of X that contains A.*

Proof. Let X be a metric space and A a subset of X. To prove that $\overline{A}$ is a closed set, we will prove that $\overline{A}$ contains its limit points. Let $x \in \overline{A}'$. To show that $x \in \overline{A}$, we proceed by contradiction and assume that $x \notin \overline{A}$. This implies that $x \notin A$ and $x \notin A'$. Since $x \notin A'$, there exists a neighborhood N of x that contains no points of A other than x. But $A \subseteq \overline{A}$ and $x \notin \overline{A}$, so it follows that $N \cap A = \emptyset$. This implies that there is an open ball $B \subseteq N$ centered at x so that $B \cap A = \emptyset$. The fact that $x \in \overline{A}'$ means that $B \cap \overline{A}$ contains a point y in $\overline{A}$ different from x. Since $B \cap A = \emptyset$, we must have $y \in A'$. But this, and the fact that B is a neighborhood of y, means that B must contain a point of A different than y. But $B \cap A = \emptyset$, so we have reached a contradiction. We conclude that $x \in \overline{A}$ and $\overline{A}' \subseteq \overline{A}$. This shows that $\overline{A}$ is a closed set.

The proof that $\overline{A}$ is the smallest closed subset of X that contains A is left for the next activity. ∎

Activity 9.8. Let (X, d) be a metric space, and let A be a subset of X.

(a) What will we have to show to prove that $\overline{A}$ is the smallest closed subset of X that contains A?

(b) Suppose that C is a closed subset of X that contains A. To show that $\overline{A} \subseteq C$, why is it enough to demonstrate that $A' \subseteq C$?

(c) If $x \in A'$, what can we say about x?

(d) Complete the proof that $\overline{A} \subseteq C$.

Activity Solution.

(a) We need to prove that any closed subset of X that contains A also contains $\overline{A}$.

(b) Since $\overline{A} = A \cup A'$ and $A \subseteq C$, to show that $\overline{A} \subseteq C$ we only need to demonstrate that $A' \subseteq C$.

(c) Llet N be a neighborhood of x. Then N contains a point of A different than x.

(d) Since $A \subseteq C$, it follows that N contains a point of C different than x. So x is a limit point of C. The fact that C is closed means that C contains its limit points, so $x \in C$. Therefore, $A' \subseteq C$ and $\overline{A} \subseteq C$.

One consequence of Theorem 9.10 is the following.

Corollary 9.11. *A subset C of a metric space X is closed if and only if $C = \overline{C}$.*

We can also characterize closed sets as sets that contain their boundaries.

Definition 9.12. The **boundary** $\text{Bdry}(A)$ of a subset A of a metric space X is the set of all boundary points of A.

Theorem 9.13. *A subset C of a metric space X is closed if and only if C contains its boundary.*

Proof. Let X be a metric space, and let C be a subset of X. First we assume that C is closed and show that C contains its boundary. Let $x \in X$ be a boundary point of C. We proceed by contradiction and assume that $x \notin C$. Then $x \in X \setminus C$, which is an open set. This implies that there is an $\epsilon > 0$ so that $B(x, \epsilon) \subseteq X \setminus C$. But then this neighborhood $B(x, \epsilon)$ contains no points in C, which contradicts the fact that x is a boundary point of C. We conclude that $x \in C$ and C contains its boundary.

For the converse, assume that C contains its boundary. To show that C is closed, we prove that $X \setminus C$ is open. We again proceed by contradiction and assume that $X \setminus C$ is not open. Then there exists $x \in X \setminus C$ such that no neighborhood of x is entirely contained in $X \setminus C$. This implies that every neighborhood of x contains a point in C. Since $x \in X \setminus C$, we have that $x \notin C$. So every neighborhood of x contains a point in C and a point in $X \setminus C$. Thus, x is a boundary point of C. It follows that $x \in C$, contradicting the fact that $x \in X \setminus C$. We conclude that $X \setminus C$ is open and C is closed. ∎

The proof of Theorem 9.13 is left to the reader.

Recall that a boundary point of a subset A of a metric space X is a point $x \in X$ such that every neighborhood of x contains a point in A and a point in $X \setminus A$. The boundary points are those that are somehow "between" a set and its complement. For example if $A = (0, 1]$ in $\mathbb{R}$, the boundary of A is the set $\{0, 1\}$. We also have that $\overline{A} = [0, 1]$ and $\overline{\mathbb{R} \setminus A} = (-\infty, 0] \cup [1, \infty)$. Notice that $\text{Bdry}(A) = \overline{A} \cap \overline{\mathbb{R} \setminus A}$. That this is always true is formalize in the next theorem.

Theorem 9.14. *Let X be a metric space and A a subset of X. Then*

$$\textit{Bdry}(A) = \overline{A} \cap \overline{X \setminus A}.$$

Proof. Let X be a metric space and A a subset of X. To prove $\text{Bdry}(A) = \overline{A} \cap \overline{X \setminus A}$ we need to verify the containment in each direction. Let $x \in \text{Bdry}(A)$ and let N be a neighborhood of x. Then N contains a point in A and a point in $X \setminus A$. We consider the cases of $x \in A$ or $x \notin A$.

- Suppose $x \in A$. Then $x \in \overline{A}$. Also, $x \notin X \setminus A$, so N must contain a point in $X \setminus A$ different from x. That makes x a limit point of $X \setminus A$ and so $X \in \overline{A} \cap \overline{X \setminus A}$.
- Suppose $x \notin A$. Then $x \in X \setminus A \subseteq \overline{X \setminus A}$. Also, $x \notin A$, so N must contain a point in A different from x. That makes x a limit point of A and so $X \in \overline{A} \cap \overline{X \setminus A}$.

In either case we have $x \in \overline{A} \cap \overline{X \setminus A}$ and so $\text{Bdry}(A) \subseteq \overline{A} \cap \overline{X \setminus A}$.

For the reverse implication, refer to the next activity.

■

Activity 9.9. Let X be a metric space and A a subset of X. In this activity we prove that

$$\overline{A} \cap \overline{X \setminus A} \subseteq \text{Bdry}(A).$$

Let $x \in \overline{A} \cap \overline{X \setminus A}$.

(a) What must be true about x, given that x is in the intersection of two sets?

(b) Let N be a neighborhood of x. As we did in the proof of Theorem 9.14, we consider the cases $x \in A$ and $x \notin A$.

   i. Suppose $x \in A$. Why must N contain a point in A and a point not in A? What does this tell us about x?

   ii. Suppose $x \notin A$. Why must N contain a point in A and a point not in A? What does this tell us about x?

   iii. What can we conclude from parts i. and ii.?

Activity Solution.

(a) Since $x \in \overline{A} \cap \overline{X \setminus A}$, it is the case that $x \in \overline{A}$ and $x \in \overline{X \setminus A}$.

(b) i. Suppose $x \in A$. Then N contains a point, namely x, in A. Now $x \notin X \setminus A$ and $x \in \overline{X \setminus A}$ means that N contains a point in $X \setminus A$ different from x. So N contains a point in A and a point in $X \setminus A$, which implies $x \in \text{Bdry}(A)$.

   ii. Suppose $x \notin A$. Then N contains a point, namely x, in $X \setminus A$. Now $x \notin A$ and $x \in \overline{A}$ means that N contains a point in A different from x. So N contains a point in A and a point in $X \setminus A$, which implies $x \in \text{Bdry}(A)$.

   iii. In either case we have $x \in \text{Bdry}(A)$ and so $\overline{A} \cap \overline{X \setminus A} \subseteq \text{Bdry}(A)$. The two containments show that $\text{Bdry}(A) = \overline{A} \cap \overline{X \setminus A}$.

Closed Sets and Limits of Sequences

Suppose we consider a sequence (a_n) in a subset A of a metric space X that converges to a point x. Must it be the case that $x \in A$? We consider this question in the next activity.

Activity 9.10. Let $A = (0, 1)$ and $B = [0, 1]$ in $(\mathbb{R}, d_E)$. For each positive integer n, let $a_n = \frac{1}{n}$. Note that the sequence (a_n) is contained in both sets A and B.

(a) To what does the sequence (a_n) converge in $\mathbb{R}$?

(b) Is $\lim a_n$ in A?

(c) Is $\lim a_n \in B$?

(d) Name two significant differences between the sets A and B that account for the different responses in parts (b) and (c)? Respond using the terminology we have introduced in this section.

Activity Solution.

(a) The sequence (a_n) converges to 0.

(b) Since $0 \notin A$, even though (a_n) is in A and converges, the sequence (a_n) does not converge to a point in A.

(c) Since $0 \in B$, the sequence (a_n) converges to a point in B.

(d) One reason for this behavior is that B is closed and A is not. That is, B contains all of its limit points but A does not.

The result of Activity 9.10 is encapsulated in the next theorem.

Theorem 9.15. *A subset C of a metric space X is closed if and only if whenever (c_n) is a sequence in C that converges to a point $c \in X$, then $c \in C$.*

Proof. Let X be a metric space and C a subset of X. First assume that C is closed. Let (c_n) be a convergent sequence in C with $c = \lim c_n$. So either $c \in C$ or c is a limit point of C. Since C contains its limit points, either case gives us $c \in C$. So $\lim c_n \in C$.

The proof of the remaining implication is left to the next activity.

■

Activity 9.11. Let X be a metric space and C a subset of X. In this activity we will prove that if any time a sequence (c_n) in C converges to a point $c \in X$, the point c is in C, then C is a closed set.

(a) How can we show that the set C is closed in a way that might be relevant to this proof?

(b) Let c be a limit point of C. What does that tell us? Why?

(c) Complete the proof that C is a closed set.

Activity Solution.

(a) To show that C is closed, we use Theorem 9.8 and prove that C contains its limit points.

(b) Since c is a limit point of C, there is a sequence (c_n) in C that converges to c.

(c) It follows from our hypothesis that $c \in C$, so C contains its limit points and is closed.

Summary

Important ideas that we discussed in this section include the following.

- Let X be a metric space and A a subset of X.

   i. A point $x \in X$ is a boundary point of A if every neighborhood of x contains a point in A and a point in $X \setminus A$.
   ii. A point x is a limit point of A if every neighborhood of x contains a point in A different from x.
   iii. A point $a \in A$ is an isolated point of A if there is a neighborhood N of a such that $N \cap A = \{a\}$.

   Boundary points and limit points don't need to be in the set A, whereas an isolated point of A must be in A. In $A = (0, 1) \cup \{2\}$ as a subset of $= (\mathbb{R}, d_E)$, 0 is a boundary point but not an isolated point while 2 is a boundary point but not a limit point. Also, 0.5 is a limit point but neither a boundary or isolated point. With A as subset of $\mathbb{R}$ with the discrete metric, every point of A is an isolated point but no point in $\mathbb{R}$ is a boundary point or a limit point of A. So even though every boundary point is either a limit point or an isolated point, the three concepts are different.

- A subset A of a metric space X is closed if $X \setminus A$ is an open set.

- Any intersection of closed sets is closed while finite unions of closed sets are closed.

- A function f from a metric space X to a metric space Y is continuous if $f^{-1}(C)$ is a closed set in X whenever C is a closed set in Y.

- Let X be a metric space, let A be a subset of X, and let x be a limit point of A. Then there is a sequence (a_n) in A that converges to x.

- Let X be a metric space, let A be a subset of X, and let x be a boundary point of A. Then there are sequences (x_n) in $X \setminus A$ and (a_n) in A that converge to x.

- The boundary of a subset A of a metric space X is the set of boundary points of A.

- A subset A of a metric space X is closed if and only if A contains all of its limit points. Similarly, A is closed if and only if A contains all of its boundary points.

- The set of all limit points of a subset A of a metric space X is denoted by A'. The closure of A is the set $\overline{A} = A \cup A'$. The closure of A is the smallest closed set in X that contains A.

- A subset A of a metric space X is closed if and only if $\lim a_n$ is in A whenever (a_n) is a convergent sequence in A.

Section 10

Subspaces and Products of Metric Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a subspace of a metric space?
- How do we find the open and closed sets in a subspace of a metric space?
- What is a product of metric spaces and how do we make a product of metric spaces into a metric space?

Introduction

Let (X, d) be a metric space, and let A be a subset of X. We can make A into a metric space itself in a very straightforward manner. When we do so, we say that A is a *subspace* of X.

Preview Activity 10.1. To make a subset A of a metric space into a metric space, we need to define a metric on A. For us to consider A as a subspace of X, we want the metric on A to agree with the metric on X. So we define $d' : A \times A \to \mathbb{R}$ by

$$d'(a_1, a_2) = d(a_1, a_2)$$

for all $a_1, a_2 \in A$.

(1) Show that d' is a metric on A.

Since d' is a metric on A it follows that (A, d') is a metric space. The metric d' is the *restriction* of d to $A \times A$ and can also be denoted by d_A.

Definition 10.1. Let (X, d) be a metric space. A **subspace** of (X, d) is a subset A of X together with the metric d_A from $A \times A$ to $\mathbb{R}$ defined by

$$d_A(x, a) = d(x, a)$$

for all $x, a \in A$.

We might wonder what, if any, properties of the space X are inherited by a subspace.

(2) Let $(X, d) = (\mathbb{R}, d_E)$ and let $A = [0, 1]$. Let $0 < a < 1$. Is the set $[0, a)$ open in X? Is the set $[0, a)$ open in A? Explain.

(3) Let $(X, d) = (\mathbb{R}, d_E)$ and let $A = \mathbb{Z}$. What are the open subsets of A? Explain.

(4) Let $(X, d) = (\mathbb{R}^2, d_E)$, let $A = \{(x, 0) \mid x \in \mathbb{R}\}$, and let $Z = \{(x, 0) \mid 0 < x < 1\}$. Note that $Z \subset A \subset X$ and we can consider Z as a subspace of A and X, and A as a subspace of X.

(a) Explain why A is a closed subset of X.

(b) Explain why Z is an open subset of A.

(c) Is Z an open subset of X? Explain

Activity Solution.

(1) Let a_1, a_2, and a_3 be in A. Now $d'(a_1, a_2) = d(a_1, a_2) \geq 0$ with equality if and only if $a_1 = a_2$ since d is a metric. Also, $d'(a_1, a_2) = d(a_1, a_2) = d(a_2, a_1) = d'(a_2, a_1)$, so d' is symmetric. Finally, the fact that d is a metric shows that

$$d'(a_1, a_3) = d(a_1, a_3) \leq d(a_1, a_2) + d(a_2, a_3) = d'(a_1, a_2) + d'(a_2, a_3),$$

and d' satisfies the triangle inequality. Thus, d' is a metric on A.

(2) Since the set $[0, a)$ is not a neighborhood of 0 in X, the set $[0, a)$ is not open in X. However, in A, the set $[0, a)$ always contains the open ball $B(r, \min\{r, a - r\})$ for any $r \in (0, a)$, and also contains the open ball $[0, \frac{a}{2}) = B\left(0, \frac{a}{2}\right)$. So $[0, a)$ is an open neighborhood of each of its points in A and is therefore an open set in A. So the open sets in a subspace of X can be different than the open sets in X.

(3) If $a \in \mathbb{Z}$, then $B(a, 0.5) = \{a\}$ in $\mathbb{Z}$. So any point is an interior point of a subset of $\mathbb{Z}$ that contains it. Thus, every subset of $\mathbb{Z}$ is open in $\mathbb{Z}$. Note also, that d_E restricted to $\mathbb{Z}$ is the discrete metric.

(4) (a) Note that $X \setminus A = \{(x, y) \mid y \neq 0\}$. If $(a, b) \in X \setminus A$, then $B\left((a, b), \frac{b}{2}\right) \subset X \setminus A$. Thus, $X \setminus A$ is open in X and so A is closed in X.

(b) If $z \in Z$, then the open ball $B(z, \min\{z, 1 - z\})$ is contained in Z. Thus, Z is open in A.

(c) The answer is no. Let $z \in Z$ and let $r > 0$. The open ball $B(z, r)$ contains the point $\left(z, \frac{r}{2}\right)$ which is not in Z. So Z is a neighborhood of none of its points in X and is therefore not open in X.

Open and Closed Sets in Subspaces

We saw in our preview activity that a subspace does not necessarily inherit the properties of the larger space. For example, a subset of a subspace might be open in the subspace, but not open in the larger space. However, there is a connection between the open subsets in a subspace and the open sets in the larger space.

Theorem 10.2. *Let (X, d) be a metric space and A a subset of X. A subset O_A of A is open in A if and only if there is an open set O_X in X so that $O_A = O_X \cap A$.*

Proof. Let (X, d) be a metric space and A a subset of X. Let O_A be an open subset of A. So for each $a \in A$ there is a $\delta_a > 0$ so that $B_A(a, \delta_a) \subseteq O_A$, where $B_A(a, \delta_a)$ is the open ball in A centered at a of radius δ_a. Then, $O_A = \bigcup_{a \in O_A} B_A(a, \delta_a)$. Now let $B_X(a, \delta_a)$ be the open ball in X centered at a of radius δ_a, and let $O_X = \bigcup_{a \in O_A} B_X(a, \delta_a)$. Note that

$$B_A(a, \delta_a) = B_X(a, \delta_a) \cap A.$$

As a union of open balls in X, the set O_X is open in X. Now

$$O_X \cap A = \left(\bigcup_{a \in O_A} B_X(a, \delta_a) \right) \cap A = \bigcup_{a \in O_A} (B_X(a, \delta_a) \cap A) = \bigcup_{a \in O_A} B_A(a, \delta_a) = O_A.$$

So there is an open set O_X in X such that $O_A = O_X \cap A$.

For the reverse implication, see the following activity. ■

Activity 10.1. Let (X, d) be a metric space and A a subset of X. Suppose that $O_A = A \cap O_X$ for some set O_X that is open in X. In this activity we will prove that O_A is an open subset of A.

(a) Let $a \in O_A$. Explain why there must exist a $\delta > 0$ such that $B_X(a, \delta)$, the open ball in X of radius δ around a in X, is a subset of O_X.

(b) What would be a natural candidate for an open ball in A centered at a that is contained in A? Prove your conjecture.

(c) What conclusion can we draw?

Activity Solution.

(a) Since $O_A = O_X \cap A$, we also have $a \in O_X$. The fact that O_X is open means that there exists $\delta > 0$ such that $B_X(a, \delta)$ is a subset of O_X.

(b) Let d_A be the restriction of d to A. Let $B_A(a, \delta) = B_X(a, \delta) \cap A$. We will show that $B_A(a, \delta)$ is the open ball B in A centered at a of radius δ. If $t \in B_A(a, \delta)$, then $t \in A$ and $t \in B_X(a, \delta)$. So $t \in A$ and $d|_A(t, a) < \delta$. So $t \in B$. Now suppose $t \in B$. Then $t \in A$ and $d_A(t, a) = d(t, a) < \delta$. Thus, $t \in A$ and $t \in B_X(a, \delta)$. This means that $t \in B_X(a, \delta) \cap A =_A (a, \delta)$. Thus, $B_A(a, \delta) = B$ and $B_A(a, \delta)$ is the open ball in A centered at a of radius δ.

(c) Since $B_A(a, \delta)$ is an open set in A, we conclude that O_A is a neighborhood of each of its points and O_A is open in A.

We might now wonder about closed sets in a subspace. If X is a metric space and A is a subspace, then by definition a subset C_A of A is closed if and only if $C_A = A \setminus O_A$ for some set O_A that is open in A. The analogy of Theorem 10.2 is true for closed sets in subspaces.

Theorem 10.3. *Let (X, d) be a metric space and A a subset of X. A subset C_A of A is closed in A if and only if there is a closed set C_X in X so that $C_A = C_X \cap A$.*

The proof is left to the reader.

Activity Solution.

(a) Let C_A be a closed subset of A. Then $C_A = A \setminus O_A$ for some open set O_A in A. Since O_A is open in A, there exists an open set O_X in X such that $O_A = A \cap O_X$. Then $C_X = X \setminus O_X$ is a closed set in X. We will show that $C_A = A \cap C_X$.

Let $c \in C_A$. Then $c \in A$ and $c \notin O_A$. Since $c \in A$, we also have $c \in X$. Now $c \in A$ but not in O_A, so c cannot be in O_X. Thus, $c \in (X \setminus O_X) = C_X$. The fact that $c \in A$ implies that $c \in (A \cap C_X)$.

Now suppose that $c \in (A \cap C_X)$. Then $c \in A$ and $c \in C_X$. Since $c \in C_X$, it follows that $c \in (X \setminus O_X)$. Thus, $c \notin O_X$. But $O_A \subseteq O_X$, so $c \notin O_A$. We conclude that $c \in (A \setminus O_A) = C_A$. Therefore, $C_A = A \cap C_X$.

(b) Suppose that $C_A = A \cap C_X$ for some closed subset C_X of X. We will show that C_A is closed in A by showing that $C_A = A \setminus O_A$ for some open set O_A in A. The fact that C_X is closed in X means that $O_X = X \setminus C_X$ is an open set in X. Then $O_A = A \cap O_X$ is an open set in A. We demonstrate that $C_A = A \setminus O_A$.

Let $c \in C_A$. Then $c \in (A \cap C_X)$. So $c \in A$ and $c \in C_X$. Thus, $c \notin (X \setminus C_X) = O_X$. It follows that $c \notin O_A$. So $c \in A$ and $c \notin O_A$, which means that $c \in (A \setminus O_A)$.

Finally, suppose that $c \in (A \setminus O_A)$. Then $c \in A$ and $c \notin O_A$. Since $c \notin O_A$, it follows that $c \notin O_X$. So $c \in (X \setminus O_X) = C_X$ and $c \in (A \cap C_X) = C_A$. We conclude that $C_A = A \setminus O_A$.

Products of Metric Spaces

If we have two metric spaces (X, d_1) and (X_2, d_2), we might wonder if we can make the set $X_1 \times X_2$ into a metric space. A natural approach might be to define a function $d : (X_1 \times X_2) \times (X_1 \times X_2) \to \mathbb{R}$ by

$$d((x, y), (u, v)) = d_1(x, u)d_2(y, v)$$

for (x, y) and (u, v) in $X_1 \times X_2$. However, this d does not define a metric. For example, if $x \in X_1$ and $y \neq v$ in X_2, then $d((x, y), (x, v)) = 0$ even though $(x, y) \neq (x, v)$. To make a metric, we can take a clue from the Euclidean metric on $\mathbb{R} \times \mathbb{R}$. On $\mathbb{R}$, the metric has the form $d_1(x, y) = |x - y|$, while on $\mathbb{R}^2$ the metric is

$$d_E((x_1, x_2), (y_1, y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} = \sqrt{d_1(x_1, y_1)^2 + d_1(x_2, y_2)^2}.$$

So on (X, d_1) and (X_2, d_2) we could consider defining $d : (X_1 \times X_2) \times (X_1 \times X_2) \to \mathbb{R}$ by

$$d((x_1, x_2), (y_1, y_2)) = \sqrt{d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2}.$$

To prove that d is a metric, we will use the Cauchy-Schwarz Inequality. If $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are in $X_1 \times X_2$, then $d(x, y) = \sqrt{d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2}$ is nonnegative by definition. Also, the symmetry of d_1 and d_2 imply that $d(x, y) = d(y, x)$. Note that $d(x, y) = 0$ if and only if $d_1(x_1, y_1) = d_2(x_2, y_2) = 0$. But this happens if and only if $x_1 = y_1$ and $x_2 = y_2$, or if $x = y$. The last (and most difficult) property to verify is the triangle inequality.

Let $z = (z_1, z_2)$ be in $X_1 \times X_2$. Then

$$\begin{aligned}
d(x,z)^2 &= d_1(x_1,z_1)^2 + d_2(x_2,z_2)^2 \\
&\leq \left(d_1(x_1,y_1) + d_1(y_1,z_1)\right)^2 + \left(d_2(x_2,y_2) + d_2(y_2,z_2)\right)^2 \\
&= \left(d_1(x_1,y_1)^2 + d_2(x_2,y_2)^2\right) + 2\left(d_1(x_1,y_1)d_1(y_1,z_1)d_2(x_2,y_2)d_2(y_2,z_2)\right) \\
&\qquad + \left(d_1(y_1,z_1)^2 + d_2(y_2,z_2)^2\right) \\
&= d(x,y)^2 + 2\left(d_1(x_1,y_1)d_1(y_1,z_1) + d_2(x_2,y_2)d_2(y_2,z_2)\right) + d(y,z)^2 \\
&\leq d(x,y)^2 + d(y,z)^2.
\end{aligned}$$

Activity 10.2. Let $X_1 = [1, 2]$ and $X_2 = [3, 4]$ as subspaces of $\mathbb{R}^2$ using the Euclidean metric.

(a) Explain in detail what the product space $X_1 \times X_2$ looks like.

(b) If B_1 is an open ball in X_1 and B_2 is an open ball in X_2, is $B_1 \times B_2$ an open ball in $X_1 \times X_2$? Explain.

(c) If B_1 is an open ball in X_1 and B_2 is an open ball in X_2, is $B_1 \times B_2$ an open set in $X_1 \times X_2$? Explain.

(d) Find, if possible, an open subset of $X_1 \times X_2$ that is not of the form $O_1 \times O_2$ where O_1 is open in X_1 and O_2 is open in X_2.

Activity Solution.

(a) The points of the form (x_1, x_2) with $1 \leq x_1 \leq 2$ and $3 \leq x_4 \leq 4$ is the square in $\mathbb{R}^2$ with sides of length 1 parallel to the coordinate axes and center at $(1.5, 3.5)$.

(b) Let $B_1 = B(a, r)$ be an open ball in X_1 and $B_2 = B(b, s)$ be an open ball in X_2. Then $B_1 = (a-r, a+r)$ and $B_2 = (b-s, b+s)$. The set $B_1 \times B_2$ is the rectangle $R = \{(x, y) \mid a - r < x < a + r, b - s < y < b + s\}$ in $\mathbb{R}^2$. But an open ball using the Euclidean metric is an open disk, not a rectangle.

(c) Let $B_1 = (a-r, a+r)$ be an open ball in X_1 and $B_2 = (b-s, b+s)$ be an open ball in X_2. Let $(u, v) \in B_1 \times B_2$. Then $u \in B_1$ and $v \in B_2$. Let $\epsilon = \min\{r - d_1(a, u), s - d_2(b, v)\}$. Let $(w, z) \in B((u, v), \epsilon)$. Then

$$d_1(a, w) \leq d_1(a, u) + d_1(u, w) < d_1(a, u) + (r - d_1(a, u)) = r$$

and

$$d_2(b, z) \leq d_2(b, v) + d_2(v, z) < d_2(b, v) + (s - d_2(b, v)) = s.$$

So $B((u, v), \epsilon) \subseteq (B_1 \times B_2)$. This makes $B_1 \times B_2$ a neighborhood of each of its points and so $B_1 \times B_2$ is an open set.

Activity 10.2 shows that open sets in a product are more complicated than just products of open sets in the factors. We will return to product later when we consider topological spaces.

We conclude with one final comment about products. We can make the Cartesian product of any number of metric spaces into a metric space with the same construction we used for the product of two spaces.

Definition 10.4. Let (X_i, d_i) be metric spaces for i from 1 to some positive integer n. The **product metric space** (X, d) is the Cartesian product

$$X = X_1 \times X_2 \times \cdot \times X_n = \prod_{i=1}^{n} X_i$$

with metric d defined by

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

when $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are in X.

The metric d is called the *product metric* and the spaces (X_i, d_i) are called the *coordinate* or *factor* spaces of (X, d). The proof that d is a metric is essentially the same as in the $n = 2$ case, and is left to the reader.

Summary

Important ideas that we discussed in this section include the following.

- A subset A of a metric space (X, d) is a metric space, called a subspace, by using the metric $d|_{A \times A}$ on A.
- If X is a metric space and A is a subspace of X, a subset O_A of A is open in A if and only if $O_A = X \cap O$ for some open set O in X. A subset C_A of A is closed in A if $C_A = A \cap O_A$ for open set O_A in A. Alternatively, a set C_A is closed in A if $C_A = A \cap C$ for some closed set C in X.
- Let (X_i, d_i) be metric spaces for i from 1 to some positive integer n. The product metric space (X, d) is the Cartesian product

$$X = X_1 \times X_2 \times \cdot \times X_n = \prod_{i=1}^{n} X_i$$

with metric d defined by

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

when $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are in X.

Part III

Topological Spaces

Section 11

Topological Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a topology and what is a topological space?
- What important properties do open sets have in relation to unions and intersections?
- What is a basis for a topology? Why is a basis for a topology useful?
- What is a neighborhood in a topological space?
- What is an interior point and the interior of a set in a topological space?
- What is the connection between the interior of a set and open sets in a topological space?

Introduction

Many of the properties that we introduced in metric spaces (continuity, limit points, boundary) could be phrased in terms of the open sets in the space. With that in mind, we can broaden our concept of space by eliminating the metric and just defining the opens sets in the space. This produces what are called *topological spaces*.

Recall that the open sets in a metric space satisfied certain properties – that the arbitrary union and any finite intersection of opens sets is open. We will now take these properties as our axioms in defining topological spaces.

Definition 11.1. Let X be a non-empty set. A set τ of subsets of X is said to a **topology** on X if

(1) X and $\emptyset$ belong to τ,

(2) any union of sets in τ is a set in τ, and

(3) any finite intersection of sets in τ is a set in τ.[1]

A *topological space* is then any set on which a topology is defined. If X is the space and τ a topology on X, we denote the topological space as (X, τ). The elements of τ are called the *open sets* in the topological space. When the topology is clear from the context, we simple refer to X as the topological space. Some examples are in order.

Preview Activity 11.1.

(1) Suppose $X = \{a, b, c\}$. Is the set $\tau = \{a, b\}$ a topology on X? Why or why not?

(2) Suppose $X = \{a, b, c, d\}$. Is the collection of subsets consisting of $\tau = \{\{a\}, \{b\}, \{a, b\}\}$ a topology on X? Why or why not? If not, what is the smallest collection of subsets of X that need to be added to τ to make τ a topology on X?

(3) Suppose $X = \{a, b, c, d\}$. Is the collection of subsets consisting of

$$\tau = \{\emptyset, \{a\}, \{b\}, \{d\}, \{a, b\}, \{a, d\}, X\}$$

a topology on X? Why or why not? If not, what is the smallest collection of subsets of X that need to be added to τ to make τ a topology on X?.

(4) Suppose $X = \{a, b, c, d\}$. Is the collection of subsets consisting of

$$\tau = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{a, c\}, \{a, d\}, \{b, d\}, \{c, d\}, X\}$$

a topology on X? Why or why not? If not, what is the smallest collection of subsets of X that need to be added to τ to make τ a topology on X?

(5) Let F be the collection of finite subsets of $\mathbb{R}$. Let $\tau = \{\emptyset, \mathbb{R}\} \cup F$. First, list three members of F and three sets that are not in F. Next, is τ a topology on $\mathbb{R}$? Why or why not?

(6) Let $\tau = \{\emptyset, \mathbb{R}, \{0\}\}$. Is τ a topology on $\mathbb{R}$? Why or why not?

(7) Let X be a set and let $\tau = \{\emptyset, X\}$. Is τ a topology on X? Why or why not?

(8) Let X be a set and let τ be the collection of all subsets of X. Is τ a topology on X? Why or why not?

Activity Solution.

(1) The answer is no. The elements of a topology must be subsets of the space X, not elements of X.

(2) Since $\emptyset \notin \tau$, as given τ is not a topology on X. We need to add both $\emptyset$ and X to τ in order for τ to be a topology on X. Once we do that, then τ is closed under intersections and unions, so $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$ is a topology on X.

[1]The symbol τ is the Greek lowercase letter tau.

(3) Notice that $\{b\} \cup \{d\}$ is not in τ, so τ is not a topology on X. Since the single element sets $\{a\}$, $\{b\}$, and $\{d\}$ are all in τ, any union of these sets will be in τ. The fact that c is no an element of any set in τ means that

$$\tau = \{\emptyset, \{a\}, \{b\}, \{d\}, \{a,b\}, \{a,d\}, \{b,d\}, \{a,b,d\}, X\}$$

is a topology on X.

(4) Notice that $\{a,b\} \cup \{c\}$ is not in τ, so τ is not a topology on X. Since the single element sets $\{a\}$, $\{b\}$, $\{c\}$, and $\{d\}$ are all in τ, any union of these sets will be in τ. This means that every subset of X must be in τ,so

$$\tau = 2^X$$

is a topology on X.

(5) The sets $\{1\}$, $\{1,2\}$, and $\{1,2,3\}$ are all in F, while the sets $(0,1)$, $\mathbb{Z}$, and $\mathbb{Q}$ are not in F. Notice that the sets $\{n\}$ for $n \in \mathbb{Z}$ are all in F, but $\bigcup_{n\in\mathbb{Z}}\{n\} = \mathbb{Z}$ is not in F. Thus, τ is not closed under arbitrary unions and τ is not a topology on $\mathbb{R}$.

(6) By inspection we can see that τ is closed under unions and intersections, so τ is a topology on $\mathbb{R}$.

(7) By inspection we can see that τ is closed under unions and intersections, so τ is a topology on $\mathbb{R}$. This topology is called the *indiscrete* or *trivial* topology.

(8) By definition, τ is closed under unions and intersections, so τ is a topology on X. This topology is called the *discrete* topology.

Examples of Topologies

In our preview activity we saw several examples of topologies. Suppose X is a non-empty set.

- The topology consisting of all subsets of X is called the *discrete topology*.
- The topology $\{\emptyset, X\}$ is the *indiscrete topology*.
- If (X,d) is a metric space, then the collection consisting of unions of all open balls is a topology called the *metric topology*.

The discrete and indiscrete topologies are standard topologies that can be defined on any set and are often used to use to generate examples. Another standard topology is in the next activity.

Activity 11.1. Let X be any set and let τ_{FC} consist of the empty set along with all subsets O of X such that $X \setminus O$ is finite. Prove that τ_{FC} is a topology on X. (The topology τ_{FC} is called the *finite complement topology*. Note that if X is finite, then τ_{FC} is just the discrete topology.)

Activity Solution. By definition, $\emptyset \in \tau_{FC}$. Since $X \setminus X = \emptyset$ is finite, $X \in \tau_{FC}$. Let $\{O_\alpha\}$ be a collection of open sets in X for α in an indexing set I. Since $X \setminus O_\alpha$ is finite for each $\alpha \in I$, we have that

$$X \setminus \bigcup_{\alpha\in I} O_\alpha = \bigcap_{\alpha\in I}(X \setminus O_\alpha) \subseteq (X \setminus O_\beta$$

for any $\beta \in I$ is a finite set. Thus, $\bigcup_{\alpha \in I} O_\alpha \in \tau_{FC}$.

Now suppose that I is finite. Then

$$X \setminus \bigcap_{\alpha \in I} O_\alpha = \bigcup_{\alpha \in I} (X \setminus O_\alpha)$$

is a finite union of finite sets and so is finite. Thus, $\bigcap_{\alpha \in I} O_\alpha \in \tau_{FC}$. We conclude that τ_{FC} is a topology.

Bases for Topologies

It can be difficult to completely describe the open sets in a topology. Instead, we can describe the topology using a collection of sets that generate the topology. For example, if (X, d) is a metric space then the collection of open sets in X forms a topology on X, called the *metric* topology. We also saw that in a metric space, every open set in X is a union of open balls. For that reason we called the collection of open balls a *basis* for the open sets in X. We can do the same thing in any topological space.

Activity 11.2. Let $X = \{a, b, c, d\}$ and let $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{c, d\}, \{a, c, d\}, \{b, c, d\}, X)$. You may assume that τ is a topology on X. Explain why any nonempty open set in the topological space (X, τ) can be written in terms of $\{a\}$, $\{b\}$, and $\{c, d\}$.

Activity Solution. By exhaustively looking at every nonempty open set we can see that

$$\begin{aligned}
\{a\} &= \{a\} \\
\{b\} &= \{b\} \\
\{a, b\} &= \{a\} \cup \{b\} \\
\{c, d\} &= \{c, d\} \\
\{a, c, d\} &= \{a\} \cup \{c, d\} \\
X &= \{a\} \cup \{b\} \cup \{c, d\}.
\end{aligned}$$

Activity 11.2 shows that, just like the open balls in a metric space, a topology can have a collection of subsets whose unions make up all of the open sets in the topology. We do need to take a little care, though. For a collection (a basis) of sets to product all of the open sets, the basis sets we start with should be open sets. In addition, every element in the topological space should be an element of one of the basis sets, and every set in the topology (except the empty set) should be a union of sets in a basis. It also must be the case that we can ensure that any finite intersection of sets in the topology remains a set in the topology when we write the sets in the topology in terms of the sets in a basis. To make the last two conditions happen, we will see that it is enough to insist that for any point in the intersection of basis elements, there is another basis element in that intersection that contains the point. This is summarized in Theorem 11.2.

Theorem 11.2. *Let X be a set and let $\mathcal{B}$ be a collection of subsets of X such that*

(1) For each $x \in X$, there is a set in $\mathcal{B}$ that contains x.

(2) If $x \in X$ is an element of $B_1 \cap B_2$ for some $B_1, B_2 \in \mathcal{B}$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Then the set τ that consists of the empty set and unions of elements of $\mathcal{B}$ is a topology on X.

Before we prove Theorem 11.2, we will need to know one fact about the set $\mathcal{B}$.

Activity 11.3. Let X be a set and $\mathcal{B}$ a collection of subsets of X such that

(1) For each $x \in X$, there is a set in $\mathcal{B}$ that contains x.

(2) If $x \in X$ is an element of $B_1 \cap B_2$ for some $B_1, B_2 \in \mathcal{B}$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Let B_1, B_2, ..., B_n be in $\mathcal{B}$. Our goal in this activity is to extend property 2 and show that if $x \in \bigcap_{1 \leq k \leq n} B_k$, then there is a set $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq \bigcap_{1 \leq k \leq n} B_k$.

(a) Since the statement we want to prove depends on a positive integer n, we will use mathematical induction. Explain why the $n = 1$ and $n = 2$ cases are true.

(b) What is the inductive hypothesis and what do we want to prove in the inductive step?

(c) Use the inductive hypothesis and condition 2 to complete the proof of the following lemma.

Lemma 11.3. *Let X be a set and $\mathcal{B}$ a collection of subsets of X such that*

(1) For each $x \in X$, there is a set in $\mathcal{B}$ that contains x.

(2) If $x \in X$ is an element of $B_1 \cap B_2$ for some $B_1, B_2 \in \mathcal{B}$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Let B_1, B_2, ..., B_n be in $\mathcal{B}$. If $x \in \bigcap_{1 \leq k \leq n} B_k$, then there is a set $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq \bigcap_{1 \leq k \leq n} B_k$.

Activity Solution.

(a) The $n = 1$ case is true by property (1) of $\mathcal{B}$, since a set B_x that contains x is a subset of itself. The $n = 2$ case is the assumption (2).

(b) For the inductive hypothesis we assume that for some integer $n \geq 1$, if $B_1, B_2, \ldots, B_n$ are in $\mathcal{B}$ and $x \in \bigcap_{1 \leq k \leq n} B_k$, then there is a set $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq \bigcap_{1 \leq k \leq n} B_k$. We want to prove that if we have sets $B_1, B_2, \ldots, B_n, B_{n+1}$ in $\mathcal{B}$ and $x \in \bigcap_{1 \leq k \leq n+1} B_k$, then there is a set $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq \bigcap_{1 \leq k \leq n+1} B_k$.

(c) Assume the lemma is true for some $n \in \mathbb{Z}^+$ and assume that we have sets $B_1, B_2, \ldots, B_n$, B_{n+1} in $\mathcal{B}$ and $x \in \bigcap_{1 \leq k \leq n+1} B_k$. Then $x \in \bigcap_{1 \leq k \leq n} B_k$, so by our induction hypothesis there is a set $B' \in \mathcal{B}$ with $x \in B'$ and $B' \subseteq \bigcap_{1 \leq k \leq n} B_k$. Now $x \in B' \cap B_{n+1}$, so condition 2 implies that there is a set $B \in \mathcal{B}$ with $x \in B$ and $B \subseteq B' \cap B_{n+1} \subseteq \bigcap_{1 \leq k \leq n+1} B_k$. This completes our proof.

Now we can prove Theorem 11.2.

Proof of Theorem 11.2. Let X be a topological space, and let $\mathcal{B}$ and τ satisfy the given conditions. By definition, $\emptyset \in \tau$. For each $x \in X$ there is a set $B_x \in \mathcal{B}$ such that $x \in B_x$. Then $X = \bigcup_{x \in X} B_x$, and $X \in \tau$. To complete our proof that τ is a topology on X, we need to demonstrate that τ is closed under arbitrary unions and finite intersections. Let $\{U_\alpha\}$ be a collection of sets in τ for α in some indexing set I. By definition, each U_α is empty or is a union of elements of sets in $\mathcal{B}$. So either $U = \bigcup_{\alpha \in I} U_\alpha$ is empty, or is a union of sets in $\mathcal{B}$. Thus, $U \in \tau$. Now let n be a positive integer and $\{U_k\}$ a collection of sets in τ for $1 \leq k \leq n$. Let $U = U_1 \cap U_2 \cap \cdots \cap U_n$. If $U_k = \emptyset$ for any k, then $U = \emptyset$ is in τ. So suppose that $U_k \neq \emptyset$ for each k between 1 and n. Let $x \in U$. Then $x \in U_k$ for each k. Choose an m between 1 and n. Since U_m is a union of elements in $\mathcal{B}$, there exists $B_m \subseteq U_m$ with $x \in B_m$. Thus, $x \in \bigcap_{1 \leq m \leq n} B_m$. Lemma 11.3 shows that there is a set $B_x \in \mathcal{B}$ such that $x \in B_x$ and $B_x \subseteq \bigcap_{1 \leq m \leq n} B_m \subseteq \bigcap_{1 \leq m \leq n} U_m$. Then

$$U = \bigcap_{1 \leq m \leq n} U_m = \bigcup_{x \in U} B_x$$

and $U \in \tau$. Therefore, τ is a topology on X. ■

Any collection $\mathcal{B}$ of sets as given in Theorem 11.2 is given a special name.

Definition 11.4. Let X be a set. A set $\mathcal{B}$ is a **basis for a topology** (or just a **basis**) on X if

(1) For each $x \in X$, there is a set in $\mathcal{B}$ that contains x.

(2) If $x \in X$ is an element of $B_1 \cap B_2$ for some $B_1, B_2 \in \mathcal{B}$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

The elements of a basis $\mathcal{B}$ are called *basis elements* or the *basic open sets*. A basis for a topology on a set X defines a topology on X as shown in Theorem 11.2.

Definition 11.5. Let $\mathcal{B}$ be a basis for a topology on a set X. The **topology τ generated by $\mathcal{B}$** contains the empty set and all arbitrary unions of basis elements.

When the topology for a space X is clear from the context, we also call a basis for the topology a basis for X.

Activity 11.4.

(a) Let $X = \{a, b, c, d, e, f\}$ and $\tau = \{\emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}, X\}$. Is the set

$$\mathcal{B} = \{\{a\}, \{c, d\}, \{b, c, d, e, f\}\}$$

a basis for τ? If not, add the smallest number of sets that you can to $\mathcal{B}$ to make a basis for this topology.

(b) Let $X = \{a, b, c\}$ and let X have the discrete topology (the topology consisting of all subsets of X). Is $\mathcal{B} = \{\{a\}, \{c\}, \{a, b\}, \{b, c\}\}$ a basis for τ in the the discrete topology? If not, add the smallest number of sets that you can to $\mathcal{B}$ to make a basis for this topology.

(c) Find a basis for the discrete topology on any set X.

Activity Solution.

(a) Since

$$\begin{aligned} \{a\} &= \{a\} \\ \{c,d\} &= \{c,d\} \\ \{a,c,d\} &= \{a\} \cup \{c,d\} \\ \{b,c,d,e,f\} &= \{b,c,d,e,f\}X && = \{a\} \cup \{b,c,d,e,f\}, \end{aligned}$$

we conclude that $\mathcal{B}$ is a basis for τ

(b) The set $\mathcal{B}$ is not a basis for τ because $\{b\}$ is not a union of sets in $\mathcal{B}$. If we add $\{b\}$ to $\mathcal{B}$, then we have all of the single point sets. Every subset of X can be made of a union of single point sets, which then produces a basis for τ.

(c) Since every single point set is open, our basis has to include all single point sets. But vvery subset of X can be made of a union of single point sets, which explains why the collection of single points sets is a basis for τ.

Metric Spaces and Topological Spaces

Every metric space is a topological space, where the topology is the collection of open sets defined by the metric. This topology is called the *metric topology*. A natural question to ask is whether every topological space is a metric space. That is, given a topological space, can we define a metric on the space so that the open sets are exactly the sets in the topology? For example, any space with the discrete topology is a metric space with the discrete metric.

Activity 11.5. Let $X = \{a,b,c,e\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a,b\}, X\}$. Explain why there cannot be a metric $d : X \times X \to \mathbb{R}$ so that the open sets in the metric topology are the sets in τ. (Hint: Assume that such a metric exists and consider the open balls centered at c.)

Activity Solution. Assume such a metric d exists. Let $r = d(a,c)$. Then $B(c,r)$ is an open set containing c. The only such open set is X, but $d(a,c) < r$ implies that $a \notin B(c,r)$. This contradiction shows that no such metric exists.

We conclude that every metric space is a topological space, but not every topological space is a metric space. The topological spaces that can be realized as metric spaces are called *metrizable*.

Neighborhoods in Topological Spaces

Recall that we defined a neighborhood of a point a in a metric space to be a subset of the space that contains an open ball centered at X. Every open ball is an open set, so we can extend the idea of neighborhood to topological spaces.

Definition 11.6. Let (X,τ) be a topological space, and let $a \in X$. A subset N of X is a **neighborhood** of a if N contains an open set that contains a.

Let's look at some examples.

Activity 11.6. Let $X = \{a, b, c, d\}$ and let $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$.

(a) Find all of the neighborhoods of the point a.

(b) Find all of the neighborhoods of the point c.

Activity Solution.

(a) The open sets that contain a are $\{a\}$, $\{a, b\}$, and X. So the neighborhoods of a are the subsets of X that contain these open sets, or

$$\{a\}, \ \{a, b\}, \ \{a, c\}, \ \{a, d\}, \ \{a, b, c\}, \ \{a, b, d\}, \ \text{ and } \ X.$$

(b) The only open set containing c is X, so X is the only neighborhood of c.

In metric spaces, an open set was a neighborhood of each of its points. This is also true in topological spaces.

Theorem 11.7. *Let (X, τ) be a topological space. A subset O of X is open if and only if O is a neighborhood of each of its points.*

Proof. Let (X, τ) be a topological space, and let O be a subset of X. First we demonstrate that if O is open, then O is a neighborhood of each of its points. Assume that O is an open set, and let $a \in O$. Then O contains the open set O that contains a, so O is a neighborhood of a.

The reverse containment is the subject of the next activity. ■

Activity 11.7. Let (X, τ) be a topological space. Let O be a subset of X. Assume O is a neighborhood of each of its points.

(a) What do we need to do to show that O is an open set?

(b) Let $a \in O$. Why must there exist an open set O_a such that $a \in O_a \subseteq O$?

(c) Complete the proof that O is an open set.

Activity Solution.

(a) To prove that O is open, we can write O as a union of open sets.

(b) Let $a \in O$. Since O is a neighborhood of a, there exists an open set O_a such that $a \in O_a \subseteq O$.

(c) We will show that
$$\bigcup_{a \in O} O_a = O.$$
Since any union of open sets is an open set, this will verify that O is an open set. Since $O_a \subseteq O$ for each $a \in O$, it follows that $\displaystyle\bigcup_{a \in O} O_a \subseteq O$. Now let $x \in O$. Then $x \in O_x \subseteq \bigcup_{a \in O} O_a$ and $O \subseteq \bigcup_{a \in O} O_a$.

The Interior of a Set in a Topological Space

We have seen that topologies define the open sets in a topological space. As in metric spaces, open sets can be characterized in terms of their interior points. We defined interior points in metric spaces in terms of neighborhoods – the same holds true in topological spaces.

Definition 11.8. Let A be a subset of a topological space X. A point $a \in A$ is an **interior point** of A if A is a neighborhood of a.

Remember that a set is a neighborhood of a point if the set contains an open set that contains the point. By definition, every open set is a neighborhood of each of its points, so every point of an open set O is an interior point of O. Conversely, if every point of a set O is an interior point, then O is a neighborhood of each of its points and is open. This argument is summarized in the next theorem.

Theorem 11.9. *Let X be a topological space. A subset O of X is open if and only if every point of O is an interior point of O.*

The collection of interior points in a set form a subset of that set, called the *interior* of the set.

Definition 11.10. The **interior** of a subset A of a topological space X is the set

$$\text{Int}(A) = \{a \in A \mid a \text{ is an interior point of } A\}.$$

Activity 11.8.

(a) Consider $(\mathbb{R}, \tau)$, where τ is the standard topology. Let $A = (-\infty, 0) \cup (1, 2] \cup \{3\}$ in $\mathbb{R}$. What is Int(A)? What is the largest open subset of $\mathbb{R}$ contained in A?

(b) Consider $(\mathbb{R}, \tau)$, where τ is the discrete topology. Let $A = (-\infty, 0) \cup (1, 2] \cup \{3\}$ in $\mathbb{R}$. What is Int(A)? What is the largest open subset of $\mathbb{R}$ contained in A?

(c) Consider $(\mathbb{R}, \tau)$, where τ is the finite complement topology. Let $A = (-\infty, 0) \cup (1, 2] \cup \{3\}$ in $\mathbb{R}$. What is Int(A)? What is the largest open subset of $\mathbb{R}$ contained in A?

(d) Let $X = \{a, b, c, d\}$ and let

$$\tau = \{\emptyset, \{a\}, \{a, b\}, \{c\}, \{d\}, \{c, d\}, \{a, c, d\}, \{a, c\}, \{a, d\}, \{a, b, c\}, \{a, b, d\}, X\}.$$

Assume that τ is a topology on X. Let $A = \{b, c, d\}$. What is Int(A)? What is the largest open subset of X contained in A?

Activity Solution.

(a) If $x \in (-\infty, 0)$, then $B(x, |x|) \subseteq A$. If $x \in (1, 2)$ and $r = \min\{x - 1, 2 - x\}$, then $B(x, r) \subseteq A$. So every point in $(-\infty, 0) \cup (1, 2)$ is an interior point of A. For any $\epsilon > 0$, the ball $B(2, \epsilon)$ contains points larger than 2 and the ball $B(3, \epsilon)$ contains points larger than 3. So neither 2 nor 3 is an interior point of A. This makes Int$(A) = (-\infty, 0) \cup (1, 2)$. We also see that Int$(A)$ is the largest open subset of X contained in A.

(b) Since every subset of $\mathbb{R}$ is an open set, every set is a neighborhood of each of its points. That and the fact that Int$(A) \subseteq A$ by definition allows us to conclude that Int$(A) = A$. Since every set is open, the largest open set in $\mathbb{R}$ contained in A is also A.

(c) Recall that a set O is open in the finite complement topology if $\mathbb{R} \setminus O$ is finite. Since $\mathbb{R} \setminus A$ is not finite, no subset of A is open and so no subset of A can be a neighborhood of each of its points. It follows that Int$(A) = \emptyset$. By the same reasoning, A contains no open subset of $\mathbb{R}$, so the largest open subset of $\mathbb{R}$ contained in A is $\emptyset$.

(d) Since $\{c\} \subseteq A$ and $\{d\} \subseteq A$, we see that c and d are interior points of A. However, there is no open subset of A that contains b, so b is not an interior point of A. It follows that Int$(A) = \{c, d\}$, which is also the largest open subset of X contained in A.

One might expect that the interior of a set is an open set, as it was in metric spaces. This is true, but we can say even more. In Activity 11.8 we saw that in our examples that Int(A) was the largest open subset of X contained in A. That this is always true is the subject of the next theorem.

Theorem 11.11. *Let (X, d) be a topological space, and let A be a subset of X. Then interior of A is the largest open subset of X contained in A.*

Proof. Let X be a topological space, and let A be a subset of X. We need to prove that Int(A) is an open set in X, and that Int(A) is the largest open subset of X contained in A. First we demonstrate that Int(A) is an open set. Let $a \in$ Int(A). Then a is an interior point of A, so A is a neighborhood of a. This implies that there exists an open set O containing a so that $O \subseteq A$. But O is a neighborhood of each of its points, so every point in O is an interior point of A. It follows that $O \subseteq$ Int(A). Thus, Int(A) is a neighborhood of each of its points and, consequently, Int(A) is an open set.

The proof that Int(A) is the largest open subset of X contained in A is left for the next activity. ■

Activity 11.9. Let (X, d) be a topological space, and let A be a subset of X.

(a) What will we have to show to prove that Int(A) is the largest open subset of X contained in A?

(b) Suppose that O is an open subset of X that is contained in A, and let $x \in O$. What does the fact that O is open tell us? Then complete the proof that $O \subseteq$ Int(A).

Activity Solution.

(a) To prove that Int(A) is the largest open subset of X contained in A we need to prove that any open subset of X that is contained in A is a subset of Int(A).

(b) Suppose that O is an open subset of X that is contained in A, and let $x \in O$. Since O is a neighborhood of x, it follows that A is a neighborhood of x and $x \in$ Int(A). It follows that $O \subseteq$ Int(A) and Int(A) is the largest open subset of X contained in A.

One consequence of Theorem 11.11 is the following.

Corollary 11.12. *A subset O of a topological space X is open if and only if $O =$ Int(O).*

Summary

Important ideas that we discussed in this section include the following.

- A topology on a set X is a collection of open subsets of X. More specifically, a set τ of subsets of a set X is a topology on X if

  (1) X and $\emptyset$ belong to τ,

  (2) any union of sets in τ is a set in τ, and

  (3) any finite intersection of sets in τ is a set in τ.

  A topological space is a set along with a topology on the set.

- Any arbitrary union of open sets is open and any finite intersection of open sets is open in a topological space.

- It can be difficult to completely describe the open sets in a topology, and it can be difficult to work with arbitrary open sets. If a collection of simpler sets generate a topology, that collection of simpler sets is a basis for the topology. More formally set $\mathcal{B}$ is a basis for a topology on a set X if

  (1) For each $x \in X$, there is a set in $\mathcal{B}$ that contains x.

  (2) If $x \in X$ is an element of $B_1 \cap B_2$ for some $B_1, B_2 \in B$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

- A subset A of a topological space X is a neighborhood of a point $a \in A$ if there is an open set O contained in A such that $a \in O$.

- A point x is a subset A of a topological space X is an interior point of A if A is a neighborhood of x. The interior of set A is the collection of all interior points of A.

- A subset A of a topological space X is open if and only if A is equal to its interior.

Section 12

Closed Sets in Topological Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What does it mean for a set to be closed in a topological space?
- What important properties do closed sets have in relation to unions and intersections?
- What is a sequence in a topological space?
- What does it mean for a sequence to converge in a topological space?
- What is a limit of a sequence in a topological space?
- What is a limit point of a subset of a topological space? How are closed sets related to limit points?
- What is a boundary point of a subset of a topological space and what is the boundary of a subset of a topological space? How are closed sets related to boundary points?
- What does it mean for a space to be Hausdorff? What important properties do Hausdorff spaces have?

Introduction

We defined a closed set in a metric space to be the complement of an open set. Since a topology is defined in terms of open sets, we can make the same definition of closed set in a topological space. With the definition of closed set in hand, we can then ask if it is possible to define limit points, boundary, and closure in topological spaces and determine if there are corresponding properties for these ideas in topological spaces.

Definition 12.1. A subset C of a topological space X is **closed** if its complement $X \setminus C$ is open.

Preview Activity 12.1.

(1) List all of the closed sets in the indicated topological space.

(a) (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$.

(b) (X, τ) with $X = \{a, b, c, d, e, f\}$ and $\tau = \{\emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}, X\}$.

(c) (X, τ) with $X = \mathbb{R}$ and $\tau = \{\emptyset, \{0\}, \mathbb{R}\}$.

(d) (X, τ) with $X = \{a, b, c\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, X\}$. (What is the name of this topology?)

(e) (X, τ) with $X = \mathbb{Z}^+$ and $\tau = \{\emptyset, X\}$ (this topology is called the *indiscrete* or *trivial* topology).

(2) In each of the examples from part (a), find (if possible), a set that is

(a) both closed and open (if possible, find one that is not the entire set or the empty set)

(b) closed but not open

(c) open but not closed

(d) not open and not closed

(3) In $\mathbb{R}^n$ with the Euclidean metric, every single element set is closed. Does this property hold in the topological space (X, τ), where $X = \{a, b, c\}$ and $\tau = \{\emptyset, \{a\}, \{a, b\}, \{a, c\}, X\}$? Explain.

Activity Solution.

(1) List all of the closed sets in the indicated topological space.

(a) The closed sets are the complements of the open sets, so the closed sets are

$$X, \{b, c, d\}, \{a, c, d\}, \{c, d\}, \text{ and } \emptyset.$$

(b) The closed sets are the complements of the open sets, so the closed sets are

$$X, \{b, c, d, e, f\}, \{a, b, e, f\}, \{b, e, f\}, \{a\}, \text{ and } \emptyset.$$

(c) The closed sets are the complements of the open sets, so the closed sets are

$$\mathbb{R}, \mathbb{R} - \{0\}, \text{ and } \emptyset.$$

(d) Every subset is open, so this topology is the discrete topology. The closed sets are the complements of the open sets, so every subset is also closed. So the closed sets are

$$X, \{b, c\}, \{a, c\}, \{a, b\}, \{c\}, \{b\}, \{a\}, \text{ and } \emptyset.$$

(e) The closed sets are the complements of the open sets, so the closed sets are

$$X \text{ and } \emptyset.$$

(2) In each of the examples from part (a), find (if possible), a set that is

(a) In the topological space (X, τ) with $X = \{a, b, c, d, e, f\}$ and $\tau = \{\emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}, X\}$ the set $\{a\}$ is both open and closed.

(b) In the topological space (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$ the set $\{a, c, d\}$ is closed but not open.

(c) In the topological space (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$ the set $\{a, b\}$ is open but not closed.

(d) Consider the topological space (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$. The set $\{b, c\}$ is neither open nor closed.

(3) Since $X \setminus \{a\} = \{b, c\}$ and $\{b, c\}$ is not open, the single element set $\{a\}$ is not closed.

Unions and Intersections of Closed Sets

Now we have defined open and closed sets in topological spaces. In our preview activity we saw that a set can be both open and closed. As we did in metric spaces, we will call any set that is both open and closed a *clopen* (for closed-open) set.

By definition, any union and any finite intersection of open sets in a topological space is open, so the fact that closed sets are complements of open sets implies the following theorem.

Theorem 12.2. *Let X be a topological space.*

(1) Any intersection of closed sets in X is a closed set in X.

(2) Any finite union of closed sets in X is a closed set in X.

Proof. Let X be a topological space. To prove part 1, assume that C_α is a collection of closed set in X for α in some indexing set I. Then

$$X \setminus \bigcap_{\alpha \in I} C_\alpha = \bigcup_{\alpha \in I} X \setminus C_\alpha.$$

The latter is an arbitrary union of open sets and so it an open set. By definition, then, $\bigcap_{\alpha \in I} C_\alpha$ is a closed set.

For part 2, assume that C_1, C_2, …, C_n are closed sets in X for some $n \in \mathbb{Z}^+$. To show that $C = \bigcap_{k=1}^{n} C_k$ is a closed set, we will show that $X \setminus C$ is an open set. Now

$$X \setminus \bigcup_{\alpha \in I} C_\alpha = \bigcap_{\alpha \in I} X \setminus C_\alpha$$

is a finite intersection of open sets, and so is an open set. Therefore, $\bigcup_{\alpha \in I} C_\alpha$ is a closed set. ■

Limit Points and Sequences in Topological Spaces

Recall that we defined a limit point of a set A in a metric space X to be a point $x \in X$ such that every neighborhood of x contains a point in A different from x. Since we have defined neighborhoods in topological spaces, we can make the same definition.

Definition 12.3. Let X be a topological space, and let A be a subset of X. A **limit point** of A is a point $x \in X$ such that every neighborhood of x contains a point in A different from x.

The set A' of limit points of A is called the *derived set* of A.

Activity 12.1. Find the limit point(s) of the following sets

(a) $\{c, d\}$ in (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$

(b) $\{a, b\}$ in the set $X = \{a, b, c, d, e, f\}$ with topology

$$\tau = \{\emptyset, \{b\}, \{a, b, c\}, \{d, e, f\}, \{b, d, e, f\}, X\}.$$

(c) $\{a, b\} \subset X$ where $X = \{a, b, c\}$ in the discrete topology.

(d) $\{-1, 0, 1\} \subset \mathbb{Z}$ with τ the topology on $\mathbb{Z}$ with basis $\{B(n)\}$, where

$$B(n) = \begin{cases} \{n\} & \text{if } n \text{ is odd,} \\ \{n-1, n, n+1\} & \text{if } n \text{ is even.} \end{cases}$$

(This topology is called the *digital line topology* and has applications in digital processing.[1] The set $\mathbb{Z}$ with the digital line topology is called the *digital line*.)

Activity Solution.

(a) Neither a nor b is a limit point of $\{c, d\}$ since the open neighborhood $\{a, b\}$ contains no point in $\{c, d\}$ different than a or b. The only open set that contains c or d is X, so that is the only neighborhood of c or d. Since X contains a point in $\{c, d\}$ that is different than c (or d), both c and d are limit points of $\{c, d\}$.

(b) None of the points b, d, e, or f is a limit point of $\{a, b\}$ since the open neighborhood $\{b, d, e, f\}$ contains no point in $\{a, b\}$ different than b, d, e, or f. Any neighborhood of a or c must contain one of the open sets $\{a, b, c\}$ or X. So every neighborhood of a or c contains a point of $\{a, b\}$ different than a or c. Therefore, the limit points of $\{a, b\}$ are a and c and $\{a, b\}' = \{a, c\}$.

(c) For any $x \in \{a, b\}$, the open neighborhood $\{x\}$ of x does not contain any points in $\{a, b\}$ different than x. So the set $\{a, b\}$ has no limit points.

[1] See *Introduction to Topology: Pure and Applied* by Colin Adams and Robert Franzosa , Pearson Education, Inc., 2008, Sections 1.4 and 11.3.

(d) Any neighborhood of 0 must contain $B(0) = \{-1, 0, 1\}$, and so every neighborhood of 0 contains a point in $\{-1, 0, 1\}$ different from 0. Similarly, any neighborhood of 2 must contain $B(2) = \{1, 2, 3\}$, and so every neighborhood of 2 contains a point in $\{-1, 0, 1\}$ different from 2. Also, any neighborhood of -2 must contain $B(-2) = \{-3, -2, -1\}$, and so every neighborhood of -2 contains a point in $\{-1, 0, 1\}$ different from -2. Thus, $\{-2, 0, 2\} \subseteq \{-1, 0, 1\}'$. If n is odd, then the open neighborhood $B(n) = \{n\}$ contains no points of $\{-1, 0, 1\}$ different than n. So no odd integer is a limit point of $\{-1, 0, 1\}$. If n is an even integer different than -2, 0, and 2, then the open neighborhood $B(n) = \{n - 1, n, n + 1\}$ contains no points in $\{-1, 0, 1\}$. Therefore, $\{-1, 0, 1\}' = \{-2, 0, 2\}$.

In metric spaces, a set is closed if and only if it contains all of its limit points. So the corresponding result in topological spaces should be no surprise.

Theorem 12.4. *Let C be a subset of a topological space X, and let C' be the set of limit points of C. Then C is closed if and only if $C' \subseteq C$.*

Proof. Let X be a topological space, and let C be a subset of X. First we assume that C is closed and show that C contains all of its limit points. Let $x \in X$ be a limit point of C. We proceed by contradiction and assume that $x \notin C$. Then $x \in X \setminus C$, which is an open set. This means that there is a neighborhood (namely $X \setminus C$) of x that contains no points in C, which contradicts the fact that x is a limit point of C. We conclude that $x \in C$ and C contains all of its limit points.

For the converse, assume that C contains all of its limit points. To show that C is closed, we prove that $X \setminus C$ is open. We again proceed by contradiction and assume that $X \setminus C$ is not open. Then there exists $x \in X \setminus C$ such that no neighborhood of x is entirely contained in $X \setminus C$. This implies that every neighborhood of x contains a point in C and so x is a limit point of C. It follows that $x \in C$, contradicting the fact that $x \in X \setminus C$. We conclude that $X \setminus C$ is open and C is closed. ■

In metric spaces we saw that limit point of a set is the limit of a sequence of points in the set. To explore this idea in topological spaces, we define sequences in the same way we did in metric spaces.

Definition 12.5. A **sequence** in a topological space X is a function $f : \mathbb{Z}^+$ to X.

We use the same notation and terminology related to sequences as we did in metric spaces: we will write (x_n) to represent a sequence f, where $x_n = f(n)$ for each $n \in \mathbb{Z}^+$. We can't define convergence in a topological space using a metric, but we can use open sets. Recall that a sequence (x_n) in a metric space (X, d) converges to a point x in the space if, given $\epsilon > 0$ there exists a positive integer N such that $d(x_n, x) < \epsilon$ for all $n \geq N$. In other words, every open ball centered at x contains all of the entries of the sequence past a certain point. We can replace open balls with open sets and make a similar definition of convergence in topological spaces.

Definition 12.6. A sequence (x_n) in a topological space X **converges** to the point $x \in X$ if, for each open set O that contains x there exists a positive integer N such that $x_n \in O$ for all $n \geq N$.

If a sequence (x_n) converges to a point x, we call x a *limit* of the sequence (x_n).

Activity 12.2. In metric spaces, limits of sequences are unique. We may wonder if the same result is true in topological spaces. Consider the topological space (X, τ), where $X = \{a, b, c\}$ and $\tau = \{\emptyset, \{c\}, \{a, c\}, \{b, c\}, X\}$? Find all limits of all constant sequences in X.

Activity Solution. We start by finding all neighborhoods.

- The neighborhoods of a are $\{a, c\}$ and X.
- The neighborhoods of b are $\{b, c\}$ and X.
- The neighborhoods of c are $\{c\}$, $\{a, c\}$, $\{b, c\}$, and X.

Consider the sequence (a). The neighborhood $\{b, c\}$ of b does not contain a, so b is not a limit of the sequence (a). The neighborhood $\{c\}$ of c does not contain a, so c is not a limit of the sequence (a). Therefore, the only limit of the sequence (a) is a.

Now consider the sequence (b). The neighborhood $\{a, c\}$ of a does not contain b, so a is not a limit of the sequence (a). The neighborhood $\{c\}$ of c does not contain b, so c is not a limit of the sequence (b). Therefore, the only limit of the sequence (b) is b.

Finally, consider the sequence (c). Each neighborhood of a and b contains the entire sequence (c), so a and b are both limits of (c). It follows that every point in X is a limit of the sequence (a).

The result of Activity 12.2 is that sequences do not behave in topological spaces as we would expect them to. Consequently, sequences do not play the same important role in topological spaces as they do in metric spaces. However, the concept of limit point is important, as are the notions of boundary and closure in topological spaces.

Closure in Topological Spaces

Once we have a definition of limit point, we can define the closure of a set just as we did in metric spaces.

Definition 12.7. The **closure** of a subset A of a topological space X is the set

$$\overline{A} = A \cup A'.$$

In other words, the closure of a set is the collection of the elements of the set and the limit points of the set. The following theorem is the analog of the theorem in metric spaces about closures.

Theorem 12.8. *Let X be a topological space and A a subset of X. The closure of a A is a closed set. Moreover, the closure of A is the smallest closed subset of X that contains A.*

Proof. Let X be a topological space and A a subset of X. To prove that $\overline{A}$ is a closed set, we will prove that $\overline{A}$ contains its limit points. Let $x \in \overline{A}'$. To show that $x \in \overline{A}$, we proceed by contradiction and assume that $x \notin \overline{A}$. This implies that $x \notin A$ and $x \notin A'$. Since $x \notin A'$, there exists a neighborhood N of x that contains no points of A other than x. But $A \subseteq \overline{A}$ and $x \notin \overline{A}$, so it follows that $N \cap A = \emptyset$. This implies that there is an open set $O \subseteq N$ centered at x so that $O \cap A = \emptyset$. The fact that $x \in \overline{A}'$ means that $O \cap \overline{A}$ contains a point y in $\overline{A}$ different from x. Since

$O \cap A = \emptyset$, we must have $y \in A'$. But the fact that O is a neighborhood of y means that O must contain a point of A different than y, which contradicts the fact that $O \cap A = \emptyset$. We conclude that $x \in \overline{A}$ and $\overline{A}' \subseteq \overline{A}$. This shows that $\overline{A}$ is a closed set.

The proof that $\overline{A}$ is the smallest closed subset of X that contains A is left for the next activity. ■

Activity 12.3. Let (X, d) be a topological space, and let A be a subset of X.

(a) What will we have to show to prove that $\overline{A}$ is the smallest closed subset of X that contains A?

(b) Suppose that C is a closed subset of X that contains A. To show that $\overline{A} \subseteq C$, why is it enough to demonstrate that $A' \subseteq C$?

(c) If $x \in A'$, what can we say about x?

(d) Complete the proof that $\overline{A} \subseteq C$.

Activity Solution.

(a) We need to prove that any closed subset of X that contains A also contains $\overline{A}$.

(b) Since $\overline{A} = A \cup A'$, if C already contains A, then to show that $\overline{A} \subseteq C$, we only need to show that $A' \subseteq C$.

(c) If $x \in A'$, then x is a limit point of A. That means that every neighborhood of x in X contains a point in A different from x.

(d) Let $x \in A'$, and let N be a neighborhood of x. Then N contains a point of A different than x. Since $A \subseteq C$, it follows that N contains a point of C different than x. So x is a limit point of C. The fact that C is closed means that C contains its limit points, so $x \in C$. Therefore, $A' \subseteq C$ and $\overline{A} \subseteq C$.

One consequence of Theorem 12.8 is the following.

Corollary 12.9. *A subset C of a topological space X is closed if and only if $C = \overline{C}$.*

The Boundary of a Set

In addition to limit points, we also defined boundary points in metric spaces. Recall that a boundary point of a set A in a metric space X could be considered to be any point in $\overline{A} \cap \overline{X \setminus A}$. We make the same definition in a topological space.

Definition 12.10. Let (X, τ) be a topological space, and let A be a subset of X. A **boundary point** of A is a point $x \in X$ such that every neighborhood of x contains a point in A and a point in $X \setminus A$. The **boundary** of A is the set

$$\text{Bdry}(A) = \{x \in X \mid x \text{ is a boundary point of } A\}.$$

As with metric spaces, the boundary points of a set A are those points that are "between" A and its complement.

Activity 12.4. Find the boundaries of the following sets

(a) $\{c, d\}$ in (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$.

(b) $\{a, b\}$ in the set $X = \{a, b, c, d, e, f\}$ with topology $\tau = \{\emptyset, \{b\}, \{a, b, c\}, \{d, e, f\}, \{b, d, e, f\}, X\}$

(c) $\{a, b\} \subset X$ where $X = \{a, b, c\}$ in the discrete topology.

(d) $\mathbb{Z}$ in $\mathbb{R}$ with the finite complement topology τ_{FC}.

Activity Solution.

(a) Neither a nor b is a boundary point of $\{c, d\}$ since the open neighborhood $\{a, b\}$ contains no point in $\{c, d\}$. The only open set that contains c or d is X, so that is the only neighborhood of c or d. Since X contains a point in $\{c, d\}$ that is different than c (or d), and a point not in $\{c, d\}$, both c and d are boundary points of $\{c, d\}$. Therefore, Bdry$(\{c, d\}) = \{c, d\}$.

(b) None of the points d, e, or f is a boundary point of $\{a, b\}$ since the open neighborhood $\{d, e, f\}$ contains no point in $\{a, b\}$. The open neighborhood $\{b\}$ of b contains no points that are not in $\{a, b\}$, so b is not a boundary point of $\{a, b\}$. Any neighborhood of a or c must contain one of the open sets $\{a, b, c\}$ or X. So every neighborhood of a or c contains a point of $\{a, b\}$ and a point not in $\{a, b\}$. So a and c are boundary points of $\{a, b\}$. Therefore, Bdry$(\{a, b\}) = \{a, c\}$.

(c) Since $\{a\}$, $\{b\}$, and $\{c\}$ are open sets, and none of these sets contain points in both A and $X \setminus A$, the boundary of A is empty.

(d) Let $x \in \mathbb{R}$ and let O be an open set containing x. Since $\mathbb{R} \setminus O$ is finite, there must be infinitely many points in both $\mathbb{Z}$ and $\mathbb{R}$ that are not in $\mathbb{R} \setminus O$. So O contains infinitely many integers and real numbers. It follows that every point in $\mathbb{Z}$ is a boundary point and Bdry$(\mathbb{Z}) = \mathbb{R}$.

Just as with metric spaces, we can characterize the closed sets as the sets that contain their boundary.

Theorem 12.11. *A subset C of a topological space X is closed if and only if C contains its boundary.*

Proof. Let X be a topological space, and let C be a subset of X. First we assume that C is closed and show that C contains its boundary. Let $x \in X$ be a boundary point of C. We proceed by contradiction and assume that $x \notin C$. Then $x \in X \setminus C$, which is an open set. But then this neighborhood $X \setminus C$ contains no points in C, which contradicts the fact that x is a boundary point of C. We conclude that $x \in C$ and C contains its boundary.

For the converse, assume that C contains its boundary. To show that C is closed, we prove that C contains its limit points. Let x be a limits point of C. To show that $x \in C$, assume to the contrary that $x \notin C$. Then $x \in X \setminus C$, an open set. Since $X \setminus C$ is a neighborhood of each of its points, the fact that x is a limit point of C implies that $X \setminus C$ must contain a point of C, a contradiction. We conclude that $x \in C$ and C contains its limit points. Therefore, C is closed. ■

The proof of Theorem 12.11 is left to the reader.

Hausdorff Spaces

As we have seen, sequences in topological spaces do not generally behave as we would expect them to. As a result, we look for conditions on topological spaces under which sequences do exhibit some regular behavior. In our preview activity we saw that in it is possible in a topological space that single point sets do not have to be closed. In Activity 12.2, we also saw that limits of sequences in topological spaces are not necessarily unique. This type of behavior limits the results that one can prove about such spaces. As a result, we define classes of topological spaces whose behaviors are closer to what our intuition suggests.

Activity 12.5.

(a) Consider a metric space (x, d), and let x and y be distinct points in X.

i. Explain why x and y cannot both be limits of the same sequence if we can find disjoint open balls $B(x, r)$ centered at x and $B(y, s)$ centered at y such that $B_x \cap B_y = \emptyset$.

ii. Now show that we can find disjoint open balls $B(x, r)$ centered at x and $B(y, s)$ centered at y such that $B(x, s) \cap B(y, r) = \emptyset$.

(b) Return to our example from Activity 12.2 with $X = \{a, b, c\}$ and $\tau = \{\emptyset, \{c\}, \{a, c\}, \{b, c\}, X\}$. We saw that every point in X is a limit of the constant sequence (c). If $x \neq c$ in X, Explain why there are no disjoint open sets O_x containing x and O_c containing c.

Activity Solution.

(a) Consider a metric space (x, d), and let x and y be distinct points in X.

i. Suppose a sequence (x_n) has x as a limit. Then there exists a positive integer N such that $n \geq n$ implies $x_n \in B(x, r)$. But then $x_n \notin B(y, s)$ and so y cannot be a limit of the sequence (x_n).

ii. Let $r = \frac{d(x,y)}{2}$. To show that $B(x, r) \cap B(y, r) = \emptyset$, suppose $z \in (B(x, r) \cap B(y, r)$. Then $z \in B(x, r)$. The triangle inequality $d(x, y) \leq d(x, z) + d(z, y)$ shows that

$$d(z, y) \geq d(x, y) - d(x, z) > d(x, y) - r = 2r - r = r.$$

But this contradicts $z \in B(y, r)$. We conclude that $B(x, r) \cap B(y, r) = \emptyset$.

(b) The only open sets that contain a are $\{a, c\}$ and X and the only open sets that contain b are $\{b, c\}$ and X. These open sets all contain c, so there are no disjoint open sets O_x and O_c containing x and c.

It is the fact as described in Activity 12.5 that we can separate disjoint points by open sets that separates metric spaces from other spaces where limits are not unique. If we restrict ourselves to spaces where we can separate points like this, then we might expect to have unique limits. Such spaces are called *Hausdorff* spaces.

Definition 12.12. A topological space X is a **Hausdorff** space if for each pair x, y of distinct points in X, there exists open sets O_x of x and O_y of y such that $O_x \cap O_y = \emptyset$.

Activity 12.5 shows that every metric space is a Hausdorff space. Once we start imposing conditions on topological spaces, we restrict the number of spaces we consider.

Activity 12.6.

(a) Let X be any set and τ the discrete topology. Is (X, τ) Hausdorff? Justify your answer.

(b) Let (X, τ) be a Hausdorff topological space with $X = \{x, x_1, x_2, \ldots, x_n\}$ a finite set. Let $x \in X$. Is $\{x\}$ an open set? Explain. What does this say about the topology τ? (Hint: Is x a limit point of $\{x_1, x_2, \ldots, x_n\}$.)

Activity Solution.

(a) Let x and y be distinct elements in X. Since every subset of X is open, the disjoint open set $\{x\}$ and $\{y\}$ separate x and y. So (X, τ) is Hausdorff.

(b) Suppose $X = \{x, x_1, x_2, \ldots, x_n\}$. Let $A = \{x_1, x_2, \ldots, x_n\}$. We will show that A is closed by demonstrating that A contains all of its limit points. To do this we only need to show that x is not a limit point of A. Since X is Hausdorff, for each i there exists an open sets Ox_i and O_i such that $Ox_i \cap O_i = \emptyset$. Let $O = \cap Ox_i$. Then O is neighborhood of x and $O \cap O_i = \emptyset$ for every i. Thus, O does not contain any points in A. Since A is closed, it follows that $\{x\} = X \setminus A$ is open. Since every single element set is open, the topology τ is the discrete topology.

Example 12.13. There are examples of Hausdorff spaces that are not the standard metric spaces. For example, Let $K = \left\{\frac{1}{n} \mid n \text{ is a positive integer}\right\}$. We use K to make a topology on $\mathbb{R}$ with basis all open intervals of the form (a, b) and all sets of the form $(a, b) \setminus K$, where $a < b$ are real numbers. This topology is known as the K-topology on $\mathbb{R}$. Just as in $(\mathbb{R}, d_E)$, if x and y are distinct real numbers we can separate x and y with open intervals.

The reason we defined Hausdorff spaces is because they have familiar properties, as the next theorems illustrate.

Theorem 12.14. *Each single point subset of a Hausdorff topological space is closed.*

Proof. Let X be a Hausdorff topological space, and let $A = \{a\}$ for some $a \in X$. To show that A is closed, we prove that $X \setminus A$ is open. Let $x \in X \setminus A$. Then $x \neq a$. So there exist open sets O_x of x and O_a of a such that $O_x \cap O_a = \emptyset$. So $a \notin O_x$ and $O_x \subseteq X \setminus A$. Thus, every point of $X \setminus A$ is an interior point and $X \setminus A$ is an open set. This verifies that A is a closed set. ■

Theorem 12.15. *A sequence of points in a Hausdorff topological space can have at most one limit in the space.*

Proof. Let X be a Hausdorff topological space, and let (x_n) be a sequence in X. Suppose (x_n) converges to $a \in X$ and to $b \in X$. Suppose $a \neq b$. Then there exist open sets O_a of a and O_b of b such that $O_a \cap O_b = \emptyset$. But the fact that (x_n) converges to a implies that there is a positive integer N such that $x_n \in O_a$ for all $n \geq N$. But then $x_n \notin O_b$ for any $n \geq N$. This contradicts the fact that (x_n) converges to b. We conclude that $a = b$ and that the sequence (x_n) can have at most one limit in X. ■

Summary

Important ideas that we discussed in this section include the following.

- A subset C of a topological space X is closed if $X \setminus C$ is open.
- Any intersection of closed sets is closed, while unions of finitely many closed sets are closed.
- A sequence in a topological space X is a function $f : \mathbb{Z}^+$ to X.
- A sequence (x_n) in a topological space X converges to a point x in X if for each open set O containing x, there exists a positive integer N such that $x_n \in O$ for all $n \geq N$.
- If a sequence (x_n) in a topological space X converges to a point x, then x is a limit of the sequence (x_n).
- A limit point of a subset A of a topological space X is a point $x \in X$ such that every neighborhood of x contains a point in A different from x. A subset C of a topological space X is closed if and only if C contains all of its limit points.
- A boundary point of a subset A of a topological space X is a point $x \in X$ such that every neighborhood of x contains a point in A and a point in $X \setminus A$. The boundary of A is the set

  $$\text{Bdry}(A) = \{x \in X \mid x \text{ is a boundary point of } A\}.$$

  A subset C of X is closed if and only if C contains its boundary.
- A topological space X is Hausdorff if we can separate distinct points with open sets in the space. That is, if for each pair x, y of distinct points in X, there exists open sets O_x of x and O_y of y such that $O_x \cap O_y = \emptyset$. Hausdorff spaces are important because sequences have unique limits in Hausdorff spaces and single point sets are closed.

Section 13

Continuity and Homeomorphisms

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- How do we define a continuous function between topological spaces?
- What is the difference between metric equivalence and topological equivalence?
- What is a homeomorphism? What does it mean for two topological spaces to be homeomorphic?
- What is a topological invariant? Why are topological invariants useful?

Introduction

Recall that we could characterize a function f from a metric space (X, d_X) to a metric space (Y, d_Y) as continuous at $a \in X$ if $f^{-1}(N)$ is a neighborhood of a in X whenever N is a neighborhood of $f(a)$ in Y. We have defined neighborhoods in topological spaces, so we can use this characterization as our definition of a continuous function from one topological space to another.

Definition 13.1. A function f from a topological space (X, τ_X) to a topological space (Y, τ_Y) is **continuous at a point** $a \in X$ if $f^{-1}(N)$ is a neighborhood of a in X whenever N is a neighborhood of $f(a)$ in Y. The function f is **continuous** if f is continuous at each point in X.

We saw that in metric spaces, a useful characterization of continuity was in terms of open sets. It is not surprising that we have the same characterization in topological spaces. You may assume the result of Theorem 13.2 for this activity.

Theorem 13.2. *Let f be a function from a topological space (X, τ_X) to a topological space (Y, τ_Y). Then f is continuous if and only if $f^{-1}(O)$ is an open set in X whenever O is an open set in Y.*

Preview Activity 13.1.

(1) Let
$$(X, \tau_X) = (\{1, 2, 3, 4\}, \{\emptyset, \{1\}, \{2\}, \{1, 2\}, X\})$$
and let
$$(Y, \tau_Y) = (\{2, 4, 6, 8\}, \{\emptyset, \{4\}, \{6\}, \{4, 6\}, Y\}).$$
Define $f : X \to Y$ by $f(x) = 2x$.

(a) Is f continuous at 4?

(b) Is f a continuous function?

(2) Let
$$(X, \tau_X) = (\{1, 2, 3, 4\}, \{\emptyset, \{1\}, \{2\}, \{1, 2\}, X\})$$
and let
$$(Y, \tau_Y) = (\{2, 4, 6, 8\}, \{\emptyset, \{4\}, \{6\}, \{4, 6\}, Y\}).$$
Is it possible to define a continuous function from X to Y? If so, find and verify one. If not, explain why.

(3) Let $X = \{1, 2, 3, 4, 5\}$ and $\tau = \{\emptyset, \{1\}, \{3, 5\}, \{1, 3, 5\}, X\}$. Define $f : X \to X$ by $f(x) = |x - 3| + 1$. At which points is f continuous? Is f a continuous function?

(4) Let $f : (\mathbb{Z}, \tau_{FC}) \to (\mathbb{Z}, d_E)$ where $f(n) = n$ and τ_{FC} is the finite complement topology. Is f a continuous function? If f is not continuous, exhibit a specific point at which f fails to be continuous. Explain.

(5) Let $f : (\mathbb{Z}, d_E) \to (\mathbb{Z}, \tau_{FC})$ where $f(n) = n$ and τ_{FC} is the finite complement topology. Is f a continuous function? If f is not continuous, exhibit a specific point at which f fails to be continuous. Explain.

(6) It can sometimes be easier to show that a function f mapping a topological space (X, d_X) to a topological space (Y, d_Y) is continuous by working with a basis instead of all open sets. Let $\mathcal{B}$ be a basis for the topology on Y. Is it the case that if $f^{-1}(B)$ is open for every $B \in \mathcal{B}$, then f is continuous? Verify your result.

Activity Solution.

(1) Let

$(X, \tau_X) = (\{1, 2, 3, 4\}, \{\emptyset, \{1\}, \{2\}, \{1, 2\}, X\})$ and $(Y, \tau_Y) = (\{2, 4, 6, 8\}, \{\emptyset, \{4\}, \{6\}, \{4, 6\}, Y\})$.

Define $f : X \to Y$ by $f(x) = 2x$.

(a) The only neighborhood of $f(4) = 8$ is Y, and $f^{-1}(Y) = X$ is a neighborhood of 4. So f is continuous at 4.

(b) Note that $\{6\}$ is an open set in Y, but $f^{-1}(\{6\}) = \{3\}$ is not an open set in X. Therefore, f is not a continuous function.

(2) Define $f : X \to Y$ by $f(1) = 4$, $f(2) = 6$, $f(3) = 2$, and $f(4) = 8$. We have

- $f^{-1}(\emptyset) = \emptyset$

- $f^{-1}(\{4\}) = \{1\}$
- $f^{-1}(\{6\}) = \{2\}$
- $f^{-1}(\{4,6\}) = \{1,2\}$
- $f^{-1}(Y) = X$.

Since $f^{-1}(O)$ is open in X for every open set O in Y, then f is a continuous function.

(3) First note that $f(1) = f(5) = 3$, $f(2) = f(4) = 2$, and $f(3) = 1$. So the inverse images of the open sets in X are

- $f^{-1}(\emptyset) = \emptyset$
- $f^{-1}(\{1\}) = \{3\}$
- $f^{-1}(\{3,5\}) = \{1,5\}$
- $f^{-1}(\{1,3,5\}) = \{1,3,5\}$
- $f^{-1}(X) = X$

Since $f^{-1}(\{3,5\})$ is not an open set in X, we see that f is not a continuous function. We examine each point in turn to determine the continuity of f at the points in X.

- Any neighborhood of $f(1)$ must contain one of the sets $\{3,5\}$, $\{1,3,5\}$ or X. The inverse images of each of these sets contains the neighborhood $\{1\}$ of 1. So f is continuous at 1.
- The only neighborhood of $f(2)$ is X, so f is continuous at 2.
- The open set $\{1\}$ is a neighborhood of $f(3)$, but $f^{-1}(\{1\})$ does not contain any open set that contains 3. So f is not continuous at 3.
- The only neighborhood of $f(4)$ is X, so f is continuous at 4.
- The open set $\{3,5\}$ is a neighborhood of $f(5)$, but $f^{-1}(\{3,5\})$ does not contain any open set that contains 5. So f is not continuous at 5.

(4) The function f is not continuous. The topology on $\mathbb{Z}$ induced by the Euclidean metric is the discrete topology. So the set $\mathbb{Z}^+$ of positive integers is an open subset of $(\mathbb{Z}, d_E)$. However, $f^{-1}(\mathbb{Z}^+) = \mathbb{Z}^+$ is a set in $\mathbb{Z}$ whose compliment is infinite. So $f^{-1}(\mathbb{Z}^+)$ is not open in $(\mathbb{Z}, \tau_{FC})$. We conclude that f is not continuous.

Let $a \in \mathbb{Z}$. The set $\{a\}$ is an open set containing a in $(\mathbb{Z}, d_E)$, but $f^{-1}(\{a\}) = \{a\}$ can contain no subset whose complement in $\mathbb{Z}$ is finite. Therefore, f is continuous at no points in $(\mathbb{Z}, \tau_{FC})$.

(5) The topology on $\mathbb{Z}$ induced by the Euclidean metric is the discrete topology, so any subset of $(\mathbb{Z}, d_E)$ is open. Thus, $f^{-1}(O)$ is open for any open subset O of $(\mathbb{Z}, \tau_{FC})$. Therefore, f is a continuous function.

(6) Let O be an open set in Y. Then $O = \cup_{\alpha \in I} B_\alpha$, where $B_\alpha \in \mathcal{B}$ for all α in some indexing set I. We know that

$$f^{-1}(O) = \bigcup_{\alpha \in I} f^{-1}(B)_\alpha.$$

Since each $f^{-1}(B)_\alpha$ is open in X, it follows that $f^{-1}(O)$ is open in X. Thus, f is continuous.

Metric Equivalence

We have seen that we can make a set into a metric space with different metrics. For example, the spaces $(\mathbb{R}^2, d_E)$, $(\mathbb{R}^2, d_T)$, $(\mathbb{R}^2, d_M)$, and $(\mathbb{R}^2, d)$ are all metric spaces, where d_E is the Euclidean metric, d_T the taxicab metric, d_M the max metric, and d the discrete metric. But are these metric spaces really "different" metric spaces? What do we mean by "different"?

Activity 13.1. We might consider two metric spaces (X, d_X) and (Y, d_Y) to be equivalent if we can find a bijection between the two sets X and Y that preserves the metric properties. That is, find a bijective function $f : X \to Y$ such that $d_X(a, b) = d_Y(f(a), f(b))$ for all $a, b \in X$. In other words, f preserves distances.

(a) Let $X = ((0, 1), d_X)$ and $Y = ((0, 2), d_Y)$, with both d_X and d_Y the Euclidean metric. Is it possible to find a bijection $f : X \to Y$ that preserves the metric properties? Explain.

(b) Now let $X = ((0, 1), d_X)$ and $Y = ((0, 2), d_Y)$, where d_X is defined by $d_X(a, b) = 2|a-b|$ and $d_Y = d_E$. You may assume that d_X is a metric. Is it possible to find a bijection $f : X \to Y$ that preserves the metric properties? Explain.

Activity Solution.

(a) The answer is no. Suppose to the contrary that there is a bijective function $f : (0, 1) \to (0, 2)$ so that

$$d_X(a, b) = d_Y(f(a), f(b))$$

for all $a, b \in X$. Let $u, v \in Y$. The surjectivity of f implies that there exists $a, b \in X$ such that $f(a) = u$ and $f(b) = v$. Then

$$|u - v| = |f(a) - f(b)| = |a - b| \leq 1.$$

So $|u - v| \leq 1$ for all $u, v \in Y$. But 1.75 and 0.25 are in Y and

$$|1.75 - 0.25| = 1.5 > 1.$$

So no such function can exist.

(b) The answer is yes. Let $f : X \to Y$ be defined by $f(x) = 2x$. We will show that f is an injection. Let $a_1, a_2 \in X$ and assume $f(a_1) = f(a_2)$. Then $2a_1 = 2a_2$ from which it follows that $a_1 = a_2$. So f is an injection. Let $y \in Y$. Then $0 < y < 2$. So $0 < \frac{y}{2} < 1$ and $\frac{y}{2} \in X$. Since

$$f\left(\frac{y}{2}\right) = 2\left(\frac{y}{2}\right) = y,$$

we see that f is a surjection. Thus, f is a bijection.

Finally, let $a, b \in X$. Then

$$d_Y(f(a), f(b)) = |2a - 2b| = 2|a - b| = d_X(a, b).$$

So f preserves distances.

If there is a bijection between metric spaces that preserves distances, we say that the metric spaces are *metrically equivalent*.

Definition 13.3. Two metric spaces (X, d_X) and (Y, d_Y) are **metrically equivalent** if there is a bijection $f : X \to Y$ such that

$$d_X(x, y) = d_Y(f(x), f(y))$$
$$d_Y(u, v) = d_X(f^{-1}(u), f^{-1}(v))$$

for all $x, y \in X$ and $u, v \in Y$.

Any function f that preserves distances (like the one in Definition 13.3) is called an *isometry*.

Definition 13.4. A function f from a metric space (X, d_X) to a metric space (Y, d_Y) is an **isometry** if f is a bijection and

$$d_Y(f(a), f(b)) = d_X(a, b) \tag{13.1}$$

for all $a, b \in X$.

Metric equivalence is a very strong type of equivalence – the existence of an isometry does not allow for much flexibility since distances must be preserved. From a topological perspective, we are only concerned about the open sets – there are no distances. The open unit ball in $(\mathbb{R}^2, d_E)$ and the open ball in $(\mathbb{R}^2, d_M)$ (where d_E is the Euclidean metric and d_M is the max metric) are not that different as we can see in Figure 13.1. If we don't worry about preserving distances, we can stretch the open ball $B_E = B((0, 0), 1)$ in $(\mathbb{R}^2, d_E)$ along the lines $y = x$ and $y = -x$ uniformly in a way to mold it onto the unit ball $B_M = B((0, 0), 1)$ in $(\mathbb{R}^2, d_M)$. The important thing is that this stretching will preserve the open sets. This is a much more forgiving type of equivalence and maintains the central idea of topology that we have discussed – what properties of a space are not altered by stretching and bending the space. This type of equivalence that allows us to manipulate a space without fundamentally changing the open sets is called *topological equivalence*.

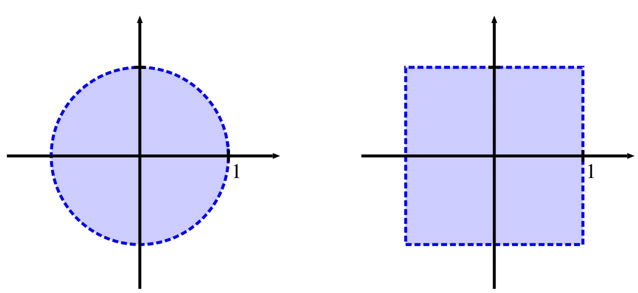


Figure 13.1: The open unit balls in $(\mathbb{R}^2, d_E)$ and $(\mathbb{R}^2, d_M)$.

Topological Equivalence

When we can deform one set into another without poking holes in the set, we consider the two sets to be equivalent from a topological perspective. Such a deformation f has to be a bijection to ensure that the two sets contain the same number of elements, continuous so that the inverse images of open sets are open, and f^{-1} must be continuous so images of open sets are open. Such a function provides a one-to-one correspondence between open sets in the two spaces. This leads to the next definition.

Definition 13.5. Two metric spaces (X, d_X) and (Y, d_Y) are **topologically equivalent** if there is a continuous bijection $f : X \to Y$ such that f^{-1} is also continuous.

Note that metric equivalence implies topological equivalence, but the reverse is not necessarily true. The function f (or f^{-1}) in Definition 13.5 is called a *homeomorphism*.

Definition 13.6. Let (X, τ_X) and (Y, τ_Y) be topological spaces. A function $f : X \to Y$ is a **homeomorphism** if f is a continuous bijection such that f^{-1} is also continuous.

If there is a homeomorphism from (X, τ_X) to (Y, τ_Y) we say that the spaces (X, τ_X) to (Y, τ_Y) are *homeomorphic* topological spaces.

It can be difficult to show directly that two metric spaces are homeomorphic, but there are ways to make the process easier in metric spaces. If f is a homeomorphism from the metric space $(\mathbb{R}^2, d_E)$ to the metric space $(\mathbb{R}^2, d_M)$, the continuity of f ensures a smooth deformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. In terms of the metrics, this means that distances cannot get distorted too much – in fact, the amount distances are distorted should be bounded. In other words, we might expect that there is a constant K so that $d_E(x, y) \leq K d_M(f(x), f(y))$ for any $x, y \in \mathbb{R}^2$. The next theorem tells us that this is a sufficient condition for topological equivalence when we work in the same underlying space.

Theorem 13.7. *Let X be a set on which two metrics d and d' are defined. If there exist positive constants K and K' so that*

$$\begin{aligned} d'(x, y) &\leq K d(x, y) \\ d(x, y) &\leq K' d'(x, y) \end{aligned}$$

for all $x, y \in X$, then (X, d) is topologically equivalent to (X, d').

Proof. Let X be a set on which two metrics d and d' are defined. Suppose there exist positive constants K and K' so that

$$\begin{aligned} d'(x, y) &\leq K d(x, y) \\ d(x, y) &\leq K' d'(x, y) \end{aligned}$$

for all $x, y \in X$. Let $i_X : (X, d) \to (X, d')$ be the identity mapping. That is, $i_X(x) = x$ for all $x \in X$. We will prove that i_X is a homeomorphism. We know that i_X is a bijection, so we only need verify that i_X and i_X^{-1} are continuous. Let $\epsilon > 0$ be given, and let $a \in X$. Let $\delta = \frac{\epsilon}{K}$. Suppose $x \in X$ so that $d(x, a) < \delta$. Then

$$d'(i_X(x), i_X(a)) = d'(x, a) \leq K d(x, a) < K\delta = K\left(\frac{\epsilon}{K}\right) = \epsilon.$$

Thus, i_X is continuous. The same argument shows that i_X^{-1} is also continuous. Therefore, i_X is a homeomorphism between (X, d) and (X, d'). ■

Activity 13.2.

(a) Are $(\mathbb{R}^2, d_T)$ and $(\mathbb{R}^2, d_M)$ topologically equivalent? Explain.

(b) Are $(\mathbb{R}^2, d_E)$ and $(\mathbb{R}^2, d_T)$ topologically equivalent? Explain.

(c) Do you expect that $(\mathbb{R}^2, d_E)$ and $(\mathbb{R}^2, d_M)$ are topologically equivalent. Explain without doing any calculations or comparisons.

Activity Solution.

(a) Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be in $\mathbb{R}^2$. Notice that

$$d_M(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\} \leq |x_1 - y_1| + |x_2 - y_2| = d_T(x, y).$$

Also,

$$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| \leq 2\max\{|x_1 - y_1|, |x_2 - y_2|\} = 2d_M(x, y).$$

So $(\mathbb{R}^2, d_T)$ and $(\mathbb{R}^2, d_M)$ are topologically equivalent.

(b) Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be in $\mathbb{R}^2$. Notice that

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

Also,

$$d_T(x, y) = |x_1 - y_1| + |x_2 - y_2| = \sqrt{(x_1 - y_1)^2} + \sqrt{(x_2 - y_2)^2} \leq \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} + \sqrt{(x_1 - y_1)^2 + (x_2 -$$

So $(\mathbb{R}^2, d_T)$ and $(\mathbb{R}^2, d_T)$ are topologically equivalent.

(c) If $f : (\mathbb{R}^2, d_T) \to (\mathbb{R}^2, d_M)$ and $g : (\mathbb{R}^2, d_E) \to (\mathbb{R}^2, d_T)$ are homeomorphisms, then we know that $f \circ g$ is a continuous bijection as is $(f \circ g)^{-1}$. So it must be the case that $(\mathbb{R}^2, d_E)$ and $(\mathbb{R}^2, d_M)$ are topologically equivalent.

Relations

We use the word "equivalent" deliberately when talking about metric or topological equivalence. Recall that equivalence is a word used with relations, and that a relation is a way to compare two elements from a set. We are familiar with many relations on sets, "$<$", "$=$", "$\geq$" on the integers, for example.

Definition 13.8. A *relation* on a set S is a subset R of $S \times S$.

For example, the subset $R = \{(a, a) : a \in \mathbb{Z}\}$ of $\mathbb{Z} \times \mathbb{Z}$ is the relation we call equals. If R is a relation on a set S, we usually suppress the set notation and write $a \sim b$ if $(a, b) \in R$ and say that a is related to b. In this case we often refer to $\sim$ as the relation instead of the set R. Sometimes we use familiar symbols for special relations. For example, we write $a = b$ if $(a, b) \in R = \{(a, a) : a \in \mathbb{Z}\}$.

When discussing relations, there are three specific properties that we consider.

- A relation $\sim$ on a set S is *reflexive* if $a \sim a$ for all $a \in S$.
- A relation $\sim$ on a set S is *symmetric* if whenever $a \sim b$ in S we also have $b \sim a$.
- A relation $\sim$ on a set S is *transitive* if whenever $a \sim b$ and $b \sim c$ in S we also have $a \sim c$.

When we use the word "equivalence", we are referring to an equivalence relation.

Definition 13.9. An **equivalence relation** is a relation on a set that is reflexive, symmetric, and transitive.

Activity 13.3.

(a) Explain why metric equivalence is an equivalence relation.

(b) Explain why topological equivalence is an equivalence relation.

Activity Solution.

(a) Explain why metric equivalence is an equivalence relation.

(b) Explain why topological equivalence is an equivalence relation.

Equivalence relations are important because an equivalence relation on a set S partitions the set into a disjoint union of equivalence classes. Since topological equivalence is an equivalence relation, we can treat the spaces that are topologically equivalent to each other as being essentially the same space from a topological perspective.

Topological Invariants

Homeomorphic topological spaces are essentially the same from a topological perspective, and they share many properties, but not all. The properties they share are called *topological invariants* or *topological properties*.

Definition 13.10. A property of a topological space X is a **topological property** (or **topological invariant**) if every topological space homeomorphic to X has the same property.

Activity 13.4. Which of the following are topological invariants? That is for topological spaces (X, τ_X) and (Y, τ_Y), if X and Y are homeomorphic space and X has the property, does it follow that Y must also have that property?

(a) X has the indiscrete topology

(b) X has the discrete topology

(c) X has the finite complement topology

(d) X contains the number 2

(e) X contains exactly 13 elements

Activity Solution.

(a) If O is an open set in T, then $f^{-1}(O)$ is open in X. This means that $f^{-1}(O)$ is either $\emptyset$ or X. It follows that O is either $\emptyset$ or $O = Y$. Thus, having the indiscrete topology is a topological invariant.

(b) Since there is a homeomorphism f from X to Y, the image of any open set in X is open in Y. The fact that f is an injection means that $\{f(x) \mid x \in X\}$ runs over all single point sets in Y, and these sets must be open. So having the discrete topology is a topological property.

(c) Let f be a homeomorphism from X to Y. Let O be an open set in Y. We know that $f^{-1}(Y \setminus O) = X \setminus f^{-1}(O)$. Since $X \setminus f^{-1}(O)$ is open, the set $X \setminus f^{-1}(O)$ is finite. So $f^{-1}(Y \setminus O)$ is finite. The fact that f is an injection means that $Y \setminus (O)$ is finite. So every open set O in Y satisfies $Y \setminus O$ is finite. Now we have to prove the converse.

Suppose B is a subset of Y such that $Y \setminus B$ is finite. Then $f^{-1}(Y \setminus B)$ is finite. This implies that $X \setminus f^{-1}(B)$ is finite and so $f^{-1}(B)$ is open in X. Since f^{-1} is continuous and f is a bijection, $f(f^{-1}(B)) = B$ is open in Y. So having the finite complement topology is a topological property.

(d) Containing a specific element is not a topological property. For example, $X = \{1, 2\}$ and $Y = \{a, b\}$, both with the discrete topology, are homeomorphic, but $2 \notin Y$.

(e) Since a homeomorphism f is a bijection, the number of elements in a set is preserved under f. So the number of elements in a set is a topological invariant.

Summary

Important ideas that we discussed in this section include the following.

- A function f from a topological space X to a topological space Y is continuous if $f^{-1}(O)$ is open in X whenever O is open in Y.

- Two metric spaces (X, d_X) and (Y, d_Y) are metrically equivalent if there is a bijection $f : X \to Y$ such that

$$d_X(x, y) = d_Y(f(x), f(y))$$
$$d_Y(u, v) = d_X(f^{-1}(u), f^{-1}(v))$$

  for all $x, y \in X$ and $u, v \in Y$. That is, X and Y are metrically equivalent if there is a isometry f from X to Y such that f^{-1} is also an isometry. Topological equivalence is a less stringent condition. Two topological spaces X and Y are topologically equivalent if there is a continuous function f from X to Y such that f^{-1} is also continuous. That is, X and Y are topologically equivalent if there is a homeomorphism between X to Y.

- A homeomorphism between topological spaces X and Y is a continuous function f from X to Y such that f^{-1} is also continuous. Two topological spaces X and Y are homeomorphic if there is a homeomorphism $f : X \to Y$.

- A topological invariant is any property that topological space X has that must also be a property of any topological space homeomorphic to X. We can sometimes use topological invariants to determine if two topological spaces are not homeomorphic.

Section 14

Subspaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a subspace of a topological space?
- How do we define the subspace topology?
- What are relatively open and closed sets?
- To what kind of spaces is $\mathbb{R}$ with the standard topology homeomorphic?

Introduction

We have seen that a subset A of a metric space (X, d_X) is a subspace of X using the restriction of the metric d_X to A. We do not have a metric in general topological spaces, so that approach can't be duplicated. But, we proved that the open sets in a subspace A of a metric space (X, d_X) are exactly the intersections of open sets in X with A. That idea can be transferred to topological spaces.

To make a subspace A of a topological space (X, τ) into a topological space, we need to define a topology on A.

Preview Activity 14.1. Let (X, τ) be a topological space and A a non-empty subset of X. It is reasonable to use the open sets in X to define open sets in A. More specifically, we might consider a subset O_A of A to be open in A if O_A is the intersection of A with some open set in X, as illustrated in Figure 14.1. With this in mind we define τ_A as

$$\tau_A = \{O \cap A \mid O \in \tau\}.$$

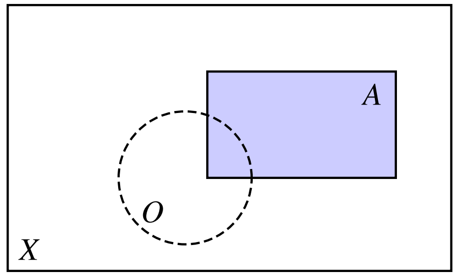


Figure 14.1: A potentially open subset in a subspace.

(1) Show that τ_A is a topology on A.

The result of item 1 is that any subset of a topological space (X, τ) is also a topological space with topology τ_A.

Definition 14.1. Let (X, τ) be a topological space. A **subspace** of (X, τ) is a non-empty subset A of X together with the topology

$$\tau_A = \{O \cap A \mid O \in \tau\}.$$

(2) For each of the following, X is a topological space and τ is a topology on X.

(a) Let $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$. Consider the subset $A = \{b, c\}$ and list the open sets in the subspace topology τ_A. Now consider $Z = \{a, b\}$. What is the name of the subspace topology τ_Z on this subset of X?

(b) Consider $X = \mathbb{R}$ with τ the indiscrete topology. What are the open sets in the subspace topology on $[1, 2]$? Now generalize to any non-empty set in the indiscrete topology.

(c) Let $X = \{a, b, c, d, e, f, g, h, i\}$ with τ the discrete topology. What are the open sets in the subspace topology on $\{a, b, d\}$. Now generalize to any non-empty set in the discrete topology.

(d) Let $X = \{a, b, c, d, e, f\}$ with $\tau = \{\emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}, X\}$. What are the open sets in the subspace $A = \{a, b, e\}$? Is every open set in A an open set in X? Explain.

(e) Let $X = \mathbb{Z}$ with $\tau = \tau_{FC}$ the finite complement topology. What are the open sets in the subspace topology on $A = \{0, 19, 37, 5284\}$? Can you generalize this to the subspace topology on any finite subset of $\mathbb{Z}$?

(f) Let $X = \mathbb{Z}$ with $\tau = \tau_{FC}$ the finite complement topology. What are the open sets in the subspace topology on the even integers? Can you generalize this to the subspace topology on any infinite subset of $\mathbb{Z}$?

Activity Solution.

(1) Since $X \in \tau$ and $X \cap A = A$, we have that $A \in \tau_A$. Also, $\emptyset \in \tau$ and $\emptyset \cap A = \emptyset$. Thus, $\emptyset \in \tau_A$. Now let $U_\alpha \in \tau_A$ for all α in some indexing set I. For each α there exists $O_\alpha \in \tau$ such that $U_\alpha = O_\alpha \cap A$. Then

$$\bigcup_{\alpha \in I} U_\alpha = \bigcup_{\alpha \in I} A \cap O_\alpha = A \cap \bigcup_{\alpha \in I} O_\alpha.$$

But $\bigcup_{\alpha \in I} O_\alpha \in \tau$, so it follows that $\bigcup_{\alpha \in I} U_\alpha \in \tau_A$. Thus, τ_A is closed under arbitrary unions. Now let $U_1, U_2, \ldots, U_n$ be in τ_A for some positive integer n. For each k there exists $O_k \in \tau$ such that $U_k = A \cap O_k$. Then Then

$$\bigcap_{k=1}^{n} U_k = \bigcap_{k=1}^{n} A \cap O_k = A \cap \bigcap_{k=1}^{n} O_k.$$

But $\bigcap_{k=1}^{n} O_k \in \tau$, so it follows that $\bigcap_{k=1}^{n} U_k \in \tau_A$. Thus, τ_A is closed under finite intersections. We conclude that τ_A is a topology on A.

(2) For each of the following, X is a topological space and τ is a topology on X.

(a) The elements of τ_A are the intersections of the sets in τ with A. So

$$\tau_A = \{\emptyset, \{b\}, A\}.$$

The elements of τ_Z are the intersections of the sets in τ with Z. So

$$\tau_Z = \{\emptyset, \{a\}, \{b\}, \{a, b\}, Z\}.$$

So τ_Z is the discrete topology on Z.

(b) The only open sets in the indiscrete topology on a set X are $\emptyset$ and X. So if A is a subset of a topological space X with the indiscrete topology, then the induced topology on A is also the indiscrete topology.

(c) If X is a topological space with the discrete topology, then every subset of X is open. Thus, if A is a subset of X, then $\tau_A = 2^A$.

(d) The elements of τ_A are the intersections of the sets in τ with A. So

$$\tau_A = \{\emptyset, \{a\}, \{b, e\}, A\}.$$

The set $\{b, e\}$ is relatively open, but not open in X.

(e) Recall that the non-empty open sets in $\mathbb{Z}$ are those sets whose complements are finite. Note that the $\mathbb{Z} \setminus \{5\}$ is in τ_{FC}. So if $n \in A$, then $(\mathbb{Z} \setminus \{5\}) \cap \{n\} = \{n\}$. So every subset of A is open and τ_A is the discrete topology. The same argument shows that if A is any finite subset of $\mathbb{Z}$, then the induced topology is the discrete topology.

(f) Let $\mathbb{E}$ be the set of even integers. We claim that the induced topology is again the finite complement topology. Let F be a finite subset of $\mathbb{E}$. Then $\mathbb{Z} \setminus F$ is in τ_{FC} and so $(\mathbb{Z} \setminus F) \cap \mathbb{E} = \mathbb{E} \setminus F$ is in the subspace topology. Conversely, if $O \in \tau_{FC}$, then $\mathbb{Z} \setminus O$ is finite. Since $\mathbb{E} \subset \mathbb{Z}$, we have $(\mathbb{E} \setminus O) \subseteq (\mathbb{Z} \setminus O)$ and

$$\mathbb{E} \setminus (\mathbb{E} \cap O) = \mathbb{E} \setminus O$$

is a finite set. So the only elements in the subsapce topology are those whose complements in $\mathbb{E}$ are finite.

The same argument will show that if X is an infinite topological space with the finite complement topology, then the induced topology on any infinite subset is the finite complement topology.

The Subspace Topology

In our preview activity, we saw that the intersection of the open sets in a topological space X with any non-empty subset A of X forms a topology for A. We then have A as a subspace of X.

The topology τ_A in Definition 14.1 is called the *subspace topology*, the *induced topology*, or the *relative topology*. In our preview activity we saw that sets that are open in a subspace A of a topological space X need not be open in X. So we call the sets in τ_A *relatively open*.

Once we have defined relatively open sets, we can then consider how to define relatively closed sets.

Activity 14.1. Let (X, τ) be a topological space, and let A be a subset of X.

(a) Since (A, τ_A) is a topological space using the subspace topology, how is a closed set in A defined? Such a set will be called *relatively closed*.

(b) How might we expect a relatively closed set in A to be related to a closed set in X? State and prove a theorem for this result.

Activity Solution.

(a) A relatively closed set is the complement in A of a relatively open set.

(b)

Theorem 14.2. *Let (X, τ) be a topological space, and let A be a subset of X. A set $C_A \in A$ is relatively closed if and only if $C_A = C \cap A$ for some closed set C in X.*

Proof. Let (X, τ) be a topological space, and let A be a subset of X. Let C_A be a relatively closed set in A. Then $A \setminus C_A$ is a relatively open set in A. Thus, $A \setminus C_A = A \cap O$ for some open set O in X. Let $C = X \setminus O$. Since O is open in X we know that C is closed in X. We will show that $C_A = A \cap C$. Now

$$C_A = A \setminus (A \setminus C_A) = A \setminus (A \cap O) = A \setminus O = A \cap (X \setminus O) = A \cap C$$

as desired.

For the converse, let C be a closed set in X and suppose $C_A = C \cap A$. To show that C_A is relatively closed, notice that

$$A \setminus C_A = A \setminus (C \cap A) = A \setminus C = A \cap (X \setminus C).$$

Now $X \setminus C$ is an open set, so $A \setminus C_A$ is a relatively open set. We conclude that C_A is a relatively closed set. ∎

Bases for Subspaces

Recall that a basis $\mathcal{B}$ for a topological space is a collection of sets that generate all of the open sets through unions. If we have a basis $\mathcal{B}$ for a topological space (X, τ), and if A is a subspace of X, we might ask if we can find a basis $\mathcal{B}_A$ from $\mathcal{B}$ in a natural way.

Activity 14.2. Let (X, τ) be a topological space with basis $\mathcal{B}$, and let A be a subspace of X.

(a) There is a natural candidate to consider as a basis $\mathcal{B}_A$ for A. How do you think we should define the elements in $\mathcal{B}_A$?

(b) Recall that a set $\mathcal{B}$ is a basis for a topological space X if

(1) For each $x \in X$, there is a set in $\mathcal{B}$ that contains x.

(2) If $x \in X$ is an element of $B_1 \cap B_2$ for some $B_1, B_2 \in B$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Show that your set from (a) is a basis for the induced topology on A.

Activity Solution.

(a) It is reasonable to define $\mathcal{B}_A$ as

$$\mathcal{B}_A = \{B \cap A \mid B \in \mathcal{B}\}.$$

(b) Let $a \in A$. Then $a \in X$ so there is a set $B \in \mathcal{B}$ such that $a \in B$. Then $a \in B \cap A$, so $\mathcal{B}_A$ satisfies the first condition of a basis.

Now suppose $a \in A$ and that $a \in B_1 \cap B_2$ for some B_1 and B_2 in $\mathcal{B}_A$. By definition of $\mathcal{B}_A$, there are sets C_1 and C_2 in $\mathcal{B}$ such that $B_1 = A \cap C_1$ and $B_2 = A \cap C_2$. The fact that $\mathcal{B}$ is a basis for X means that there is a set $C_3 \in \mathcal{B}$ such that $a \in C_3 \subseteq C_1 \cap C_2$. Let $B_3 = A \cap C_3$. Then $a \in B_3$ and

$$B_3 = A \cap C_3 \subseteq A \cap (C_1 \cap C_2) = (A \cap C_1) \cap (A \cap C_2) = B_1 \cap B_2.$$

It follows that $\mathcal{B}_A$ is a basis for the relative topology on A.

Open Intervals and $\mathbb{R}$

If we think of a homeomorphism as allowing us to stretch or bend a space, it is reasonable to think that we could stretch an open interval of the form (a, b) infinitely in both directions without altering the nature of the open sets. That is, we should expect that $\mathbb{R}$ with the standard topology is homeomorphic to (a, b) with the subspace topology.

Activity 14.3. Let a and b be real numbers with $a < b$. To show that $\mathbb{R}$ is homeomorphic to (a, b), we need a continuous bijection from $\mathbb{R}$ to (a, b) whose inverse is also continuous.

(a) First we demonstrate that $(0, 1)$ and $\mathbb{R}$ are homeomorphic. Let $f : (0, 1) \to \mathbb{R}$ be defined by

$$f(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right).$$

i. Explain why f maps $(0, 1)$ to $\mathbb{R}$.
ii. Explain why f is an injection.
iii. Explain why f is a surjection.
iv. Explain why f and f^{-1} are continuous. (Hint: Use a result from calculus.)

(b) The result of (a) is that $\mathbb{R}$ and $(0, 1)$ are homeomorphic spaces. To complete the argument that $\mathbb{R}$ is homeomorphic to (a, b), define a function $g : (0, 1) \to (a, b)$ and explain why your g is a homeomorphism.

Activity Solution.

(a) Let $f(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$.

i. Since $\tan(x)$ is defined on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, we have that f is defined for

$$\begin{gathered} -\tfrac{\pi}{2} < \pi\left(x - \tfrac{1}{2}\right) < \tfrac{\pi}{2} \\ -\tfrac{1}{2} < x - \tfrac{1}{2} < \tfrac{1}{2} \\ 0 < x < 1. \end{gathered}$$

The output of the tangent function is always a real number, so f maps $(0, 1)$ into $\mathbb{R}$.

ii. If $f(a) = f(b)$, then

$$\begin{aligned} \tan\left(\pi\left(a - \frac{1}{2}\right)\right) &= \tan\left(\pi\left(b - \frac{1}{2}\right)\right) \\ \tan^{-1}\left(\tan\left(\pi\left(a - \frac{1}{2}\right)\right)\right) &= \tan^{-1}\left(\tan\left(\pi\left(b - \frac{1}{2}\right)\right)\right) \\ \pi\left(a - \frac{1}{2}\right) &= \pi\left(b - \frac{1}{2}\right) \\ a &= b. \end{aligned}$$

Thus, f is an injection.

iii. If $y \in \mathbb{R}$, then

$$f\left(\frac{1}{2} + \frac{1}{\pi}\arctan(y)\right) = y,$$

so f maps $(0, 1)$ onto $\mathbb{R}$.

iv. Results from calculus apply to $\mathbb{R}$ using the standard metric. Since f is a composite of differentiable functions, we know that f is continuous. Note also that f is invertible with

$$f^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x).$$

We see that f^{-1} is also a differentiable function, therefore continuous. We conclude that $\mathbb{R}$ is homeomorphic to $(0, 1)$.

(b) Define $g : (0, 1) \to (a, b)$ by $g(x) = a + (b - a)x$. As a linear function we know that g is continuous. The fact that $a < b$ implies that g is increasing and invertible. Since $a + (b - a)(0) = a$ and $a + (b - a)(1) = b$, the increasing continuous nature of g shows that g maps $(0, 1)$ onto (a, b). The inverse of g is also linear, so g^{-1} is continuous. We conclude that g is a homeomorphism. The transitivity of begin homeomorphic shows that $\mathbb{R}$ is homeomorphic to any open interval of the form (a, b).

It is left to the reader to show that $\mathbb{R}$ is also homeomorphic to any interval of the form (a, ∞) or $(-\infty, b)$. Later we will determine if $\mathbb{R}$ is homeomorphic to intervals of the form $[a, b)$, $(a, b]$, $[a, \infty)$ or $(-\infty, b]$.

Summary

Important ideas that we discussed in this section include the following.

- A subspace of a topological space is any nonempty subset of the topological space endowed with the subspace topology.
- An open subset in the subspace topology for a subset A of a topological space X is any set of the form $O \cap A$, where O is an open set in X.
- The relatively open sets are the open sets in a subspace topology. The relatively closed sets are complements of the relatively open sets in a subspace topology. That is, a relatively closed set in the subspace A of a topological space X are the sets of the form $A \cap C$, where C is a closed set in X.
- The topological space $\mathbb{R}$ with the standard topology is homeomorphic to any open interval as well as open intervals of the form (a, ∞) or $(-\infty, b)$ for any real numbers a and b.

Section 15

Quotient Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a quotient topology?
- What is a quotient space?
- What are two examples of familiar quotient spaces?

Introduction

Take the interval $X = [0, 1]$ in $\mathbb{R}$ and bend it to be able to glue the endpoints together. The resulting object is a circle. By identifying the endpoints 0 and 1 of the interval, we are able to create a new topological space. We can view this gluing or identifying of points in the space X in a formal way that allows us to recognize the resulting space as a quotient space.

The Quotient Topology

Given a topological space X and a surjection f from X to a set Y, we can use the topology on X to define a topology on Y. This topology on Y identifies points in X through the function f. The resulting topology on Y is called a *quotient* topology. The quotient topology gives us a way of creating a topological space which models gluing and collapsing parts of a topological space.

Preview Activity 15.1. Let $X = \{1, 2, 3, 4, 5, 6\}$ and let $\tau = \{\emptyset, \{1, 2\}, \{4, 6\}, \{1, 2, 4, 6\}, X\}$. Let $Y = \{a, b, c, d\}$ and define $p : X \to Y$ by

$$p(1) = b,\ p(2) = a,\ p(3) = c,\ p(4) = d,\ p(5) = c, \text{ and } p(6) = a.$$

Our goal in this activity is to define a topology on Y that is related to the topology on X via p.

(a) We know the sets in X that are open. So let us consider the sets U in Y such that $p^{-1}(U)$ is open in X. Define σ to be this set. That is

$$\sigma = \{U \subseteq Y \mid p^{-1}(U) \in \tau\}.$$

Find all of the sets in σ.

(b) Show that σ is a topology on Y.

(c) Explain why $p : (X\tau) \to (Y, \sigma)$ is continuous.

(d) Show that σ is the largest topology on Y for which p is continuous. That is, if σ' is a topology on Y with $\sigma \subset \sigma'$, then $p : (X, \tau) \to (Y, \sigma')$ is not continuous.

Activity Solution.

(a) The only possibilities for sets U are those with $p^{-1}(U)$ being $\emptyset$, $\{1, 2\}$, $\{4, 6\}$, $\{1, 2, 4, 6\}$, and X. Since p is a surjection, the only time $p^{-1}(U) = \emptyset$ is when $U = \emptyset$. Also, $p^{-1}(U) = X$ when $X = Y$. We consider the remaining cases in turn.

- Since $p(1) = p(6) = b$ and $p(2) = a$, $p^{-1}(U)$ can never be $\{1, 2\}$.
- Since $p(4) = d$, and $p(2) = p(6) = a$, $p^{-1}(U)$ can never be $\{4, 6\}$.
- Since $p(1) = b$, $p(2) = p(6) = a$, and $p(4) = d$, $p^{-1}(U) = \{1, 2, 4, 6\}$ when $U = \{a, b, d\}$.

Thus,

$$\sigma = \{\emptyset, \{a, b, d\}, Y\}.$$

(b) By inspection we can see that unions and intersections of sets in σ remain in σ, so σ is a topology on Y.

(c) By definition, if U is in σ, then $p^{-1}(U)$ is in τ. So the inverse image of any open set is open under p and p is continuous.

(d) Suppose σ' is a topology on Y with $\sigma \subset \sigma'$. Then there exists $O \in \sigma' \setminus \sigma$. If $p^{-1}(O)$ were open, then we would have $O \in \sigma$ be definition. But this is a contradiction. So $p^{-1}(O)$ is not open and p is not continuous from $(X\tau)$ to (Y, σ').

Quotient Spaces

As we saw in our preview activity, if we have a surjection p from a topological space (X, τ) to a set Y, we were able to define a topology on Y by making the open sets the sets $U \subseteq Y$ such that $p^{-1}(U)$ is open in X. This is how we will create what is called the *quotient topology*. Before we can define the quotient topology, we need to know that this construction always makes a topology.

Activity 15.1. Let (X, τ_X) be a topological space, let Y be a set, and let $p : X \to Y$ be a surjection. Let

$$\tau_Y = \{U \subseteq Y \mid p^{-1}(U) \in \tau_X\}.$$

(a) Why are $\emptyset$ and Y in τ_Y?

(b) Let $\{U_\beta\}$ be a collection of sets in τ_Y for β in some indexing set J.

i. Show that $\bigcup_{\beta \in J} U_\beta$ is in τ_Y.

ii. If J is finite, show that $\bigcap_{\beta \in J} U_\beta$ is in τ_Y.

(c) What conclusion can we draw about τ_Y?

Activity Solution.

(a) Since $p^{-1}(\emptyset) = \emptyset \in \tau_X$ we have that $\emptyset \in \tau_Y$. The fact that p is a surjection means that $p^{-1}(Y) = X \in \tau_X$. So $Y \in \tau_Y$.

(b) Let $\{U_\beta\}$ be a collection of sets in τ_Y for β in some indexing set J.

i. By definition, $p^{-1}(U_\beta \in \tau_X$ for every $\beta \in J$. Since $p^{-1}\left(\bigcup_{\beta \in J} U_\beta\right) = \bigcup_{\beta \in J} p^{-1}(U_\alpha)$ is in τ_X it follows that $\bigcup_{\beta \in J} U_\beta$ is in τ_Y.

ii. Since $p^{-1}\left(\bigcap_{\beta \in J} U_\beta\right) = \bigcap_{\beta \in J} p^{-1}(U_\alpha)$ is in τ_X it follows that $\bigcap_{\beta \in J} U_\beta$ is in τ_Y.

(c) We can conclude that τ_Y is a topology on Y.

Activity 15.1 allows us to define the quotient topology.

Definition 15.1. Let (X, τ_X) be a topological space, let Y be a set, and let $p : X \to Y$ be a surjection. The **quotient topology** on Y is the set

$$\{U \subseteq Y \mid p^{-1}(U) \in \tau_X\}.$$

The function p is then called a **quotient map** and Y a **quotient space**.

Activity 15.2.

(a) Let $X = \mathbb{R}$ with standard topology, let $Y = \{-1, 0, 1\}$, and define $p : X \to Y$ by

$$p(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Find all of the open sets in the quotient topology.

(b) Let $X = \mathbb{R}$ with standard topology, let $Y = [0, 1)$, and define $p : X \to Y$ by

$$p(x) = x - \lfloor x \rfloor,$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x, (For example $\lfloor 1.2 \rfloor = 1$, and so $p(1.2) = 1.2 - 1 = 0.2$. Be careful, note that $\lfloor -0.7 \rfloor = -1$.) Determine the sets in the quotient topology. (Hint: The graph of p on $[-2, 2]$ is shown in Figure 15.1.)

Activity Solution.

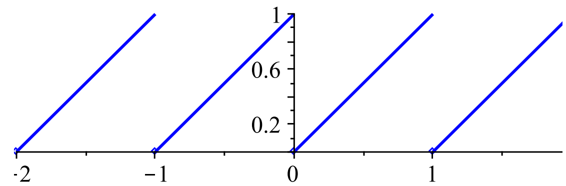


Figure 15.1: The graph of $p(x) = x - \lfloor x \rfloor$.

(a) We know that $\emptyset$ and Y are open. We consider all of the remaining subsets of Y. Note that

$p^{-1}(\{-1\}) = (-\infty, 0), p^{-1}(\{0\}) = \{0\}, p^{-1}(\{1\}) = (0, \infty, 0), p^{-1}(\{-1, 0\}) = (-\infty, 0], p^{-1}(\{-1, 1\}$

The only subsets of Y that have open inverse images are $\{-1\}$, $\{1\}$, and $\{-1, 1\}$. So the quotient topology is

$$\{\emptyset, \{-1\}, \{1\}, \{-1, 1\}, Y\}.$$

(b) The graph shows that for y in Y, we have

$$p^{-1}(y) = \{k + y \mid k \in \mathbb{Z}\}.$$

So $p^{-1}(B) = \{k + b \mid k \in \mathbb{Z} \text{ and } b \in B\}$. The only way for $p^{-1}(B)$ to be open in X is for $p^{-1}(B)$ to be a union of open intervals. This will happen only when B is a union of open intervals in Y. Since $p^{-1}((a, b)) = \bigcup_{k \in \mathbb{Z}} (k + a, k + b)$ when $0 \leq a < b < 1$, the quotient topology is

$$\{\emptyset, Y\} \cup \{(a, b) \mid 0 \leq a < b < 1\}.$$

Another perspective of the quotient topology utilizes the fact that any equivalence relation on a set X partitions X into a union of disjoint equivalence classes $[x] = \{y \in X \mid y \sim x\}$. There is a natural surjection q from X to the space of equivalence classes given by $q(x) = [x]$. We investigate this perspective in the next activity.

Activity 15.3. Let $X = \{a, b, c, d, e, f\}$ and let $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, \{a, b, c, d\}, X\}$. Then (X, τ) is a topological space. Let $A = \{a, b, c\}$ and $B = \{d, e, f\}$. Define a relation $\sim$ on X such that $x \sim y$ if x and y are both in A or both in B. Assume that $\sim$ is an equivalence relation. The sets A and B are the equivalence classes for this relation. That is $A = [a] = [b] = [c]$ and $B = [d] = [e] = [f]$. Let $X^* = \{A, B\}$. Then we can define $p : X \to X^*$ by sending $x \in X$ to the set to which it belongs. That is, $p(x) = [x]$ for $x \in X$, or

$$p(a) = A, p(b) = A, p(c) = A, p(d) = B, p(e) = B, \text{ and } p(f) = B.$$

Determine the sets in the quotient topology on X^*.

Activity Solution. Since $X^* = \{A, B\}$, there are only limited possibilities for open sets. We consider each in turn:

$$\begin{aligned} p^{-1}(\emptyset) &= \emptyset \\ p^{-1}(\{A\}) &= \{a, b, c\} \\ p^{-1}(\{B\}) &= \{d, e, f\} \\ p^{-1}(X^*) &= X. \end{aligned}$$

From this list we see that the quotient topology on X^* is $\{\emptyset, \{A\}, X^*\}$.

The partition of X in Activity 15.3 into the disjoint union of sets A and B defines an equivalence relation on X where $x \sim y$ if x and y are both in the same set A or B. That is, $a \sim b \sim c$ and $d \sim e \sim f$. In this context, the sets A and B are equivalence classes – $A = [a]$ and $B = [d]$, where $[x]$ is the equivalence class of x. This leads to a general construction.

If (X, τ) is a topological space and $\sim$ is an equivalence relation on X, we can let $X/\sim$ be the set of distinct equivalence classes of X under $\sim$. Then $p : X \to X/\sim$ defined by $p(x) = [x]$ is a surjection and $X/\sim$ has the quotient topology. The space $X/\sim$ is called a *quotient space*. The space $X/\sim$ is also called an *identification* space because the equivalence relation identifies points in the set to be thought of as the same. This allows us to visualize quotient spaces as resulting from gluing or collapsing parts of the space X.

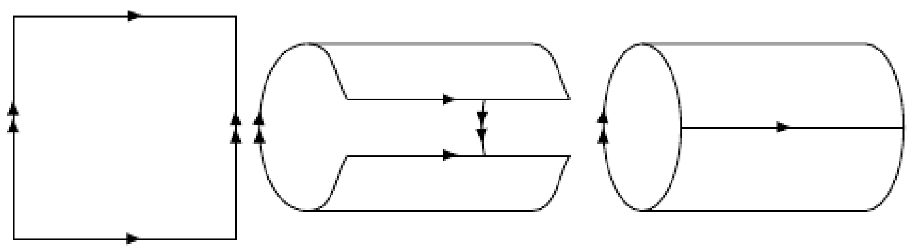


Figure 15.2: A tube as the identification space $X/\sim$.

Example 15.2. Let $I = [0, 1]$ and let $X = I \times I$ with standard topology. Define a relation $\sim$ on X by $(x, y) \sim (x, y)$ if $0 < x < 1$ and $0 \leq y \leq 1$, $(0, y) \sim (1, y)$ if $0 \leq y \leq 1$. It is straightforward to show that $\sim$ is an equivalence relation. Let us consider what the identification space $X/\sim$ looks like. The space $I \times I$ is the unit square as shown in Figure 15.2. All points in the interior of the square are identified only with themselves. However, the left side and right side are identified with each other in the same direction. Think of X as a piece of paper. We roll up the sides of the square to make the left and right sides coincide. The result is that $X/\sim$ is the cylinder as shown in Figure 15.2 (image copied from `http://i.stack.imgur.com/FJaFe.png`).

Activity 15.4. Quotient spaces can be difficult to describe. This activity presents a few more examples.

(a) Let $X = [0, 1]$ with standard topology and define an equivalence relation $\sim$ on X by $0 \sim 1$ and $x \sim x$ for all x not equal to 0 or 1. What does the quotient space $X/\sim$ look like? (Hint: Think about the relation $\sim$ as gluing the points 0 and 1 together.)

(b) Describe quotient spaces of $X = I \times I$ with standard topology given by the following equivalence relations $\sim$. (Here I is the closed interval $[0, 1]$.)

   i. $(x, y) \sim (x, y)$ if $0 < x < 1$ and $0 \leq y \leq 1$ and $(0, y) \sim (1, 1 - y)$ when $0 \leq y \leq 1$.

   ii. $(x, y) \sim (x, y)$ if $0 < x < 1$ and $0 < y < 1$, $(0, y) \sim (1, y)$ for $0 < y < 1$, $(x, 0) \sim (x, 1)$ for $0 < x < 1$, and $(0, 0) \sim (0, 1) \sim (1, 0) \sim (1, 1)$

   iii. $(x, y) \sim (x, y)$ if $0 < x < 1$ and $0 < y < 1$ and $(x, y) \sim (u, v)$ if (x, y) and (u, v) are boundary points.

Activity Solution.

(a) If we take the unit interval X and glue the points 0 and 1 together, we obtain a circle, that is the space S^1.

(b) Describe quotient spaces of $X = I \times I$ with standard topology given by the following equivalence relations $\sim$. (Here I is the closed interval $[0, 1]$.)

i. All points in the interior of the square are identified only with themselves. However, the left side and right side are identified with each other – but in the opposite direction as indicated at left in Figure 15.3. So if we roll up and twist the square to make the left and right sides coincide in the opposite direction, we obtain the Möbius strip as shown at right in Figure 15.3.

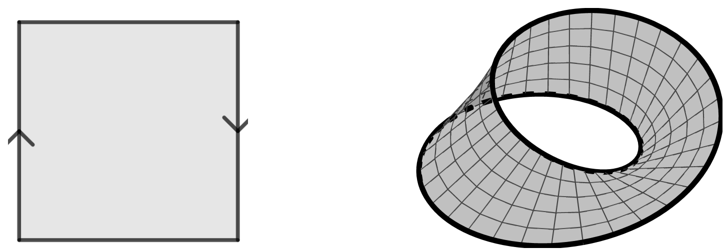


Figure 15.3: A Möbius strip as the identification space $X/\sim$.

ii. All points in the interior of the square are identified only with themselves. However, the left side and right side are identified with each other, as are the top and bottom – both in the same direction as indicated in Figure 15.2. So if we first roll up the square to identify the left and right sides, we obtain a cylinder. The top and bottom of the square become the top and bottom of the cylinder. Now bend the cylinder around to identify the top and bottom, resulting in the torus as shown at right in Figure 15.4 (image copied from `http://i.stack.imgur.com/FJaFe.png`). .

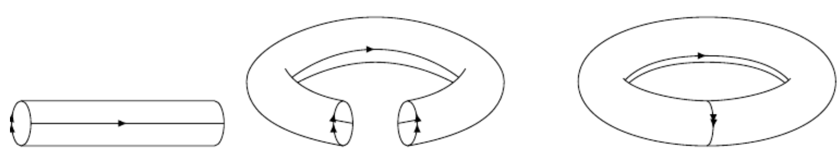


Figure 15.4: A torus as the identification space $X/\sim$.

iii. All points in the interior of the square are identified only with themselves, but the boundaries points are all identified together as indicated at left in Figure 15.5. So the boundary collapses to a point. Think of taking the square and pinching all of the sides together. What is left looks like a sphere, with the sides all identified with a single point on the sphere as shown at right in Figure 15.5.

Many other interesting identification spaces can be made. For example, if $X = I \times I$ and $\sim$ is defined by $(x, y) \sim (x, y)$ if $0 < x < 1$ and $0 < y < 1$, $(0, y) \sim (1, y)$ for $0 < y < 1$, $(x, 0) \sim$

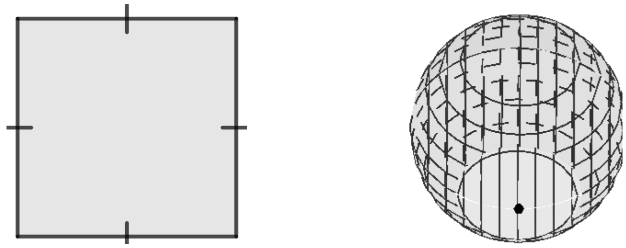


Figure 15.5: A sphere as the identification space $X/\sim$.

$(1-x, 1)$ for $0 < x < 1$, then the resulting identification space $X/\sim$ is a Klein bottle. A nice illustration of this can be seen at https://plus.maths.org/content/introducing-klein-bottle.

Identifying Quotient Spaces

Suppose X is a topological space and Y is a set, and let $p : X \to Y$ be a surjection. We can define a relation $\sim_p$ on X by $x \sim_P y$ if and only if $p(x) = p(y)$. It is straightforward to show that $\sim_p$ is an equivalence relation. From this we can see that our two approaches to defining the quotient topology and quotient spaces are really the same.

Oftentimes we have a topological space X and a relation $\sim$ on X, and we would like to have an effective way to be able to identify the quotient space $X/\sim$ as homeomorphic to some familiar topological space Y. That is, we want to be able to show that there is a homeomorphism f from $X/\sim$ to Y.

Example 15.3. Consider the following situation. Let $X = \mathbb{R}$ with the standard topology and define the relation $\sim$ on $\mathbb{R}$ by $x \sim y$ if $x - y \in \mathbb{Z}$. It is straightforward to show that $\sim$ is an equivalence relation. We will see that $\mathbb{R}/\sim$ is homeomorphic to the circle $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ as a subspace of $\mathbb{R}^2$ with the standard topology. Since every point on the unit circle has the form $(\cos(t), \sin(t))$ for some real number t, we might try defining $f : (\mathbb{R}/\sim) \to S^1$ by $f([t]) = (\cos(t), \sin(t))$. However, we have that $0 \sim 1$, which means that $[0] = [1]$, but $f([0]) \neq f([1])$ and so f is not well-defined. Another option might be $f([t]) = (\cos(2\pi t), \sin(2\pi t))$. In this case, if $x \sim y$, then $2\pi x$ and $2\pi y$ differ by a multiple of 2π and so $f([x]) = f([y])$. We could then attempt to show that f is a homeomorphism.

The following theorem encapsulates the above example.

Theorem 15.4. *Let X and Y be sets and let $\sim$ be an equivalence relation on X. Let f be a function from X to Y such that $f(x_1) = f(x_2)$ whenever $x_1 \sim x_2$ in X. Let $X/\sim$ be the set of equivalence classes of X under the relation $\sim$, and let $p : X \to (X/\sim)$ be the standard map defined by $p(x) = [x]$. The function $\overline{f}$ mapping $X/\sim$ to Y defined by $\overline{f}([x]) = f(x)$ for every $x \in X$ is the unique function that satisfies*

$$f = \overline{f} \circ p.$$

Activity 15.5. Theorem 15.4 is a statement about sets and functions, and there is no topology involved. We prove the theorem in this activity. Use the conditions stated in Theorem 15.4.

(a) Show that $\overline{f}$ is well-defined. That is, show that whenever $[x_1] = [x_2]$ in $X/\sim$, then $\overline{f}([x_1]) = \overline{f}([x_2])$.

(b) Prove that $f = \overline{f} \circ p$.

(c) Show that the uniqueness of $\overline{f}$ comes from the equation $f = \overline{f} \circ p$.

Activity Solution.

(a) The fact that $f(x_1) = f(x_2)$ whenever $x_1 \sim x_2$ in X makes $\overline{f}$ well-defined. That is, if $[x_1] = [x_2]$ in $X/\sim$, then $x_1 \sim x_2$ and so

$$\overline{f}([x_1]) = f(x_1) = f(x_2) = \overline{f}([x_2]).$$

(b) To demonstrate that $f = \overline{f} \circ p$, let $x \in X$. Then

$$(\overline{f} \circ p)(x) = \overline{f}(p(x)) = \overline{f}([x]) = f(x).$$

(c) For uniqueness, if $g : (X/\sim) \to Y$ satisfies $g \circ p = f$, then $g(p(x)) = g([x]) = f(x)$ for every $x \in X$ and $g = \overline{f}$.

Now we present a final result that can be very helpful when working with quotient spaces.

Theorem 15.5. *Let X be a topological space and let $\sim$ be an equivalence relation on X. Consider the set $X/\sim$ to be a topological space with the quotient topology, and let $p : X \to (X/\sim)$ be the standard surjection defined by $p(x) = [x]$. Let Y be a topological space with $f : X \to Y$ a continuous function such that $f(x_1) = f(x_2)$ whenever $x_1 \sim x_2$ in X. Then $\overline{f} : (X/\sim) \to Y$ defined by $\overline{f}([x]) = f(x)$ is the unique continuous function satisfying $f = \overline{f} \circ p$.*

Proof. The existence of $\overline{f}$ as the unique function satisfying $f = \overline{f} \circ p$ was established in Theorem 15.4. All that remains is to show that $\overline{f}$ is continuous. Let O be an open set in Y. Since f is continuous, we know that $f^{-1}(O)$ is open in X. If $x_1 \in f^{-1}(O)$ and $x_1 \sim x_2$, then $x_2 \in f^{-1}(O)$ as well. Thus, we can write $f^{-1}(O)$ as

$$f^{-1}(O) = \bigcup_{x \in f^{-1}(O)} [x].$$

That is, $f^{-1}(O)$ is a union of equivalence classes. Now $\overline{f}([x]) = f(x)$, so if $x \in f^{-1}(O)$, then $[x] \in \overline{f}^{-1}(O)$. Thus,

$$f^{-1}(O) = \bigcup_{x \in f^{-1}(O)} [x] = \bigcup_{[x] \in \overline{f}^{-1}(O)} [x] = \overline{f}^{-1}(O).$$

We conclude that $\overline{f}^{-1}(O)$ is open in X and $\overline{f}$ is continuous. ∎

Now we will see how to use Theorem 15.5 to establish a homeomorphism from a quotient space of a given topological space to another topological space

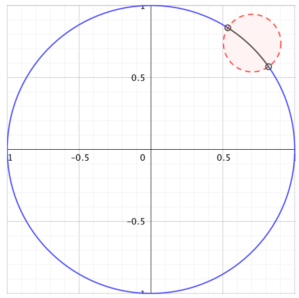


Figure 15.6: A basis element for S^1.

Example 15.6 (Continued.)**.** We return to the example with $X = \mathbb{R}$ under the standard topology with equivalence relation $\sim$ defined by $x \sim y$ if $x - y \in \mathbb{Z}$. Our goal is to show that $\mathbb{R}/\sim$ is homeomorphic to the circle $Y = S^1$.

Step 1. Define a continuous surjection $f : X \to Y$ that respects the relation. That is, we need to ensure that $f(x_1) = f(x_2)$ whenever $x_1 \sim x_2$ in X. We saw earlier that the function f defined by $f(t) = (\cos(2\pi t), \sin(2\pi t))$ respects the relation. Since every point on the unit circle is of the form $(\cos(\theta), \sin(\theta))$ for some real number θ, choosing $t = \frac{\theta}{2\pi})$ makes $f(t) = (\cos(\theta), \sin(\theta))$ and f is a surjection. Now we need to demonstrate that f is continuous. A collection of basic open sets in S^1 can be found by intersecting S^1 with open balls in $\mathbb{R}^2$ as illustrated in Figure 15.6. We can see that the basic open sets are arcs of the form $\widehat{ab}$ for a and b in S^1. Suppose $a = (\cos(2\pi A), \sin(2\pi A))$ and $b = (\cos(2\pi B), \sin(2\pi B))$ for angles A and B. Then $f^{-1}(\widehat{ab})$ is the union of intervals $(A + 2\pi k, B + 2\pi k)$ for $k \in \mathbb{Z}$. As a union of open intervals, we have that $f^{-1}(\widehat{ab})$ is open in X. We have now found a continuous surjection from X to Y that respects the relation.

Step 2. Find a continuous function from $X/\sim$ to Y. Theorem 15.5 tells us that the function $\overline{f} : (X/\sim) \to Y$ defined by $f([t]) = f(t)$ is continuous. So $\overline{f}$ is our candidate to be a homeomorphism.

Step 3. Show that $\overline{f}$ is a bijection. Let $y \in Y$. The fact that f is a surjection means that there is a $t \in \mathbb{R}$ such that $f(t) = y$. It follows that $\overline{f}([t]) = f(t) = y$ and $\overline{f}$ is a surjection. To demonstrate that $\overline{f}$ is an injection, suppose $\overline{f}([s]) = \overline{f}([t])$ for some $s, t \in \mathbb{R}$. Then $(\cos(2\pi s), \sin(2\pi s)) = f(s) = f(t) = (\cos(2\pi t), \sin(2\pi t))$. It must be the case then that $2\pi s$ and $2\pi t$ differ by a multiple of 2π. That is, $2\pi s - 2\pi t = 2\pi k$ for some integer k. From this we have $s - t = k \in \mathbb{Z}$, and so $s \sim t$. This makes $[s] = [t]$ and we conclude that $\overline{f}$ is an injection.

Step 4. Show that $\overline{f}$ is a homeomorphism. At this point we already know that $\overline{f}$ is a continuous bijection, so the only item that remains is to show that $\overline{f}(\overline{O})$ is open whenever $\overline{O}$ is open in $X/\sim$. Let $p : X \to (X/\sim)$ be the standard map. Let $\overline{O}$ be a nonempty open set in $X/\sim$. Then $O = p^{-1}(\overline{O})$ is open in X. Thus, O is a union of open intervals. Let (a, b) be an interval contained in O. From the definition of f we have that $f(a, b)$ is the open arc $\widehat{f(a)f(b)}$, which is open in Y. So $f(O)$ is a union of open arcs in Y, which makes $f(O)$ open in Y. Now $f(O) = (\overline{f} \circ p)(O) = \overline{f}(p(O)) = \overline{f}(\overline{O})$, and $\overline{f}(\overline{O})$ is open in Y. We conclude that $\overline{f}$ is a homeomorphism from $X/\sim$ to $\S^1$, and so S^1 is a quotient space of $\mathbb{R}$.

Summary

Important ideas that we discussed in this section include the following.

- Let (X, τ_X) be a topological space, let Y be a set, and let $p : X \to Y$ be a surjection. The quotient topology on Y is the set
$$\{U \subseteq Y \mid p^{-1}(U) \in \tau_X\}.$$
- The function p is a quotient topology as in the previous bullet is called a quotient map and the space Y is a quotient space.
- A circle, a Möbius strip, a torus, and a sphere can all be realized as quotient spaces.

Section 16

Compact Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a cover of a subset of a topological space? What is an open cover?
- What is a subcover of a cover?
- What is a compact subset of a topological space?
- What is one application of compactness?
- How do we characterize the compact subsets of $\mathbb{R}^n$? What theorem provides this characterization?

Introduction

Closed and bounded intervals have important properties in calculus. Recall, for example, that every real-valued function that is continuous on a closed interval $[a, b]$ attains a maximum and minimum value on that interval. The question we want to address in this section is if there is a corresponding characterization for subsets of topological spaces that ensure that continuous real-valued functions with domains in topological spaces attain maximum and minimum values. The property that we will develop is called compactness.

The word "compact" might bring to mind a notion of smallness, but we need to be careful with the term. We might think that the interval $(0, 0.5)$ is small, but $(0, 0.5)$ is homeomorphic to $\mathbb{R}$, which is not small. Similarly, we might think that the interval $[-10000, 10000]$ is large, but this interval is homeomorphic to the "small" interval $[-0.00001, 0.000001]$. As a result, the concept of compactness does not correspond to size, but rather representation, in a way.

Since a topology defines open sets, topological properties are often defined in terms of open sets. Let us consider an example to see if we can tease out some of the details we will need to get a useful notion of compactness. Consider the open interval $(0,1)$ in $\mathbb{R}$. Suppose we write $(0,1)$ as a union of open balls. For example, let $O_n = \left(\frac{1}{n}, 1 - \frac{1}{n}\right)$ for $n \in \mathbb{Z}^+$ and $n \geq 3$. Notice that $(0,1) = \bigcup_{n \geq 3} O_n$. Any collection of open sets whose union is $(0,1)$ is called an *open cover* of $(0,1)$. Working with a larger number of sets is generally more complicated than working with a smaller number, so it is reasonable to ask if we can reduce the number of sets in our open cover of $(0,1)$ and still cover $(0,1)$. In particular, working with a finite collection of sets is preferable to working with an infinite number of sets (we can exhaustively check all of the possibilities in a finite setting if necessary). Notice that $O_n \subset O_{n+1}$ for each n, so we can eliminate many of the sets in this cover. However, if we eliminate enough sets so that we are left with only finitely many, then there will be a maximum value of n so that O_n remains in our collection. But then $\frac{1}{2n}$ will not be in the union of our remaining collection of sets. As a result, we cannot find a finite collection of the O_n whose union contains $(0,1)$.

Let's apply the same idea now to the set $[0,1]$. Suppose we extend our open cover $\{O_n\}$ to be an open cover of the closed interval $[0,1]$ by including two additional open balls: $O_0 = B(0, 0.5)$ and $O_1 = B(1, 0.5)$. Now the sets O_0, O_1, and O_4 is a finite collection of sets that covers $[0,1]$. So even though the interval $[0,1]$ is "larger" than $(0,1)$ in the sense that $(0,1) \subset [0,1]$ we can represent $[0,1]$ in a more efficient (that is finite) way in terms of open sets than we can the interval $(0,1)$. This is the basic idea behind compactness.

Definition 16.1. A subset A of a topological space X is **compact** if for every set I and every family of open sets $\{O_\alpha\}$ with $\alpha \in I$ such that $A \subseteq \bigcup_{\alpha \in I} O_\alpha$, there exists a finite subfamily $\{O_{\alpha_1}, O_{\alpha_2}, \ldots, O_{\alpha_n}\}$ such that $A \subseteq \bigcup_{i=1}^{n} O_{\alpha_i}$.

There is some terminology associated with Definition 16.1.

Definition 16.2. A **cover** of a subset A of a topological space X is a collection $\{S_\alpha\}$ of subsets of X for α in some indexing set I so that $A \subseteq \bigcup_{\alpha \in I} S_\alpha$. In addition, if each set S_α is an open set, then the collection $\{S_\alpha\}$ is an **open cover** for A.

Definition 16.3. A **subcover** of a cover $\{S_\alpha\}_{\alpha \in I}$ of a subset A of a topological space X is a collection $\{S_\beta\}$ for $\beta \in J$, where J is a proper subset of I such that $A \subseteq \bigcup_{\beta \in J} S_\beta$. In addition, if J is a finite set, the subcover $\{S_\beta\}_{\beta \in J}$ is a **finite subcover** of $\{S_\alpha\}_{\alpha \in I}$.

So the sets O_0, O_1, and O_4 in our previous example form a finite subcover of the open cover $\{O_n\}_{n \geq 3}$.

We can restate the definition of compactness in the following way: a subset A of a topological space X is compact if every open cover of A has a finite subcover of A.

Preview Activity 16.1. Determine if the subset A of the topological space X is compact. Either prove A is compact by starting with an arbitrary infinite cover and demonstrating that there is a finite subcover, or find a specific infinite cover and prove that there is no finite subcover.

(1) $A = \{-2, 3, e, \pi, 456875\}$ in $X = \mathbb{R}$ with the Euclidean topology. Generalize this example.

(2) $A = (0,1]$ in $X = \mathbb{R}$ with the Euclidean topology.

(3) $A = \left\{\frac{1}{n} \mid n \in \mathbb{Z}^+\right\}$ in $X = \mathbb{R}$ with the Euclidean topology.

(4) $A = \mathbb{Z}^+$ in $X = \mathbb{R}$ with the Euclidean topology.

(5) $A = \mathbb{Z}^+$ in $X = \mathbb{R}$ with the finite complement topology.

(6) $A = \mathbb{R}$ in $X = \mathbb{R}$ with the Euclidean topology.

Activity Solution.

(1) Let X be a topological space and let A be a finite subset of X. We will demonstrate that A is compact. Suppose $\mathcal{C} = \{O_\alpha\}_{\alpha \in I}$ is an open cover of A with indexing set I. For each $a \in A$, let O_a be an element of $\mathcal{C}$ with $a \in O_a$. Then $\{O_a\}_{a \in A}$ is a finite subset of $\mathcal{C}$ that covers A.

(2) For each $n \in Z^+$, let $O_n = \left(\frac{1}{n}1 + \frac{1}{n}\right)$. Then $\mathcal{C} = \{O_n\}_{n \in \mathbb{Z}^+}$ is an open cover of A. To show that $\mathcal{C}$ contains no finite subcover of A, proceed by contradiction and assume that there is a positive integer k and a finite collection U_1, U_2, ..., U_k of sets in $\mathcal{C}$ such that $A \subseteq \cup_{i=1}^k U_i$. For each i let $U_i = (a_i, b_i)$, and let $m = \min\{a_i \mid \mid 1 \leq i \leq k\}$. Then $\frac{1}{2a_m}$ is in A, but not in $\cup_{i=1}^k U_i$. So $\mathcal{C}$ contains no finite subcover of A and A is not compact.

(3) The set A is not compact, with an argument just like the one in part 2.

(4) The set A is not compact. For each $n \in \mathbb{Z}^+$, let $O_n = \left(n - \frac{1}{2}, n + \frac{1}{2}\right)$. Then $\mathcal{C} = \{O_n\}_{n \in \mathbb{Z}^+}$ is an open cover of $\mathbb{Z}^+$. If $n \in Z^+$ and we remove the set O_n from the collection $\mathcal{C}$, then the union of the remaining sets in $\mathcal{C}$ will not contain n. So $\mathcal{C}$ contains no finite cover of A.

(5) We show that A is compact. Let $\mathcal{C} = \{O_\alpha\}_{\alpha \in I}$ be an open cover of A with indexing set I. Choose an $\alpha \in I$ and let $U_0 = O_\alpha$. We know that $\mathbb{R} \setminus U_0$ is finite. Let $\{x_1, x_2, \ldots, x_k\} = \mathbb{R} \setminus U_0$. For each i there is a set U_i in $\mathcal{C}$ such that $x_i \in U_i$. Then $A \subseteq \mathbb{R} = \cup_{i=0}^k U_i$ and so $\mathcal{C}$ contains a finite subcover of A. Therefore, A is compact. This argument shows that any nonempty subset of any space with the finite complement topology is compact.

(6) The set $\mathbb{R}$ is not compact. Let $x \in \mathbb{R}$ and let $O_x = (-x, x)$ (with $O_0 = \emptyset$). Then $\mathcal{C} = \{O_x\}_{x \in \mathbb{R}}$ is an open cover of $\mathbb{R}$. If $\mathcal{C}$ contains a finite subcover $\{U_i\}_{1 \leq i \leq k}$ with $U_i = (-x_i, x_i)$, then the set $\{x_i\}_{1 \leq i \leq k}$ has an upper bound M. But then $M + 1$ is not in $\cup_{1 \leq i \leq k} U_i$. So $\mathcal{C}$ contains no finite subcover of $\mathbb{R}$ and $\mathbb{R}$ is not compact.

Compactness and Continuity

In our preview activity we learned about compactness – the analog of closed intervals from $\mathbb{R}$ in topological spaces. Recall that a subset A of a topological space X is compact if every open cover of A has a finite sub-cover. As we will see, the definition of compactness is exactly what we need to ensure results of the type that continuous real-valued functions with domains in topological spaces attain maximum and minimum values on compact sets.

We might expect that compact sets have certain properties, but we must be careful which ones we assume.

Activity 16.1. Let $X = \{a, b, c, d\}$ and give X the topology $\tau = \{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}, X\}$.

(a) Explain why every finite subset of a topological space must be compact.

(b) Find, if possible, a subset of X that is compact and open. If no such subset exists, explain why.

(c) If A is a compact subset of X, must A be open? Explain.

(d) Find, if possible, a subset of X that is compact and closed. If no such subset exists, explain why.

(e) If A is a compact subset of X, must A be closed? Explain.

Activity Solution.

(a) Suppose that X is a topological space and $A = \{a_1, a_2, \ldots, a_n\}$ is a finite subset of X. If $\mathcal{O} = \{O_\alpha\}_{\alpha \in I}$ is an open cover of A, then for each $1 \leq k \leq n$ there is an α_k such that $a_k \in O_{\alpha_k}$. So $\{O_{\alpha_k}\}_{1 \leq k \leq n}$ is a finite subcover of $\mathcal{O}$. So every finite subset of a topological space is compact.

(b) The set $\{a\}$ is both open and compact in X.

(c) The set $\{c\}$ is compact but not open in X.

(d) The set $\{d\}$ is closed and compact in X.

(e) The set $\{a\}$ is compact, but not closed because $\{a\}^c$ is not open.

The message of Activity 16.1 is that compactness by itself is not related to closed or open sets. We will see later, though, that in some reasonable circumstances, compact sets and closed sets are related. For the moment, we take a short detour and ask if compactness is a topological property.

Activity 16.2. Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be continuous. Assume that A is a compact subset of X. In this activity we want to determine if $f(A)$ must be a compact subset of Y.

(a) What do we need to show to prove that $f(A)$ is a compact subset of Y? Where do we start?

(b) If we have an open cover of $f(A)$ in Y, how can we find an open cover $\{U_\alpha\}$ for A? Be sure to verify that what you claim is actually an open cover of A.

(c) What do we know about any open cover of A?

(d) Complete the proof of the following theorem.

Theorem 16.4. *Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be continuous. If A is a compact subset of X, then $f(A)$ is a compact subset of Y.*

Activity Solution.

(a) To prove that $f(A)$ is a compact subset of Y, let $\{O_\alpha\}$ be a collection of open subsets of Y for α in some indexing set I such that $f(A) \subseteq \bigcup_{\alpha \in I} O_\alpha$. We will show that there is a positive integer n and a finite collection O_{α_1}, O_{α_2}, …, O_{α_n}, of sets in $\{O_\alpha\}_{\alpha \in I}$ that cover $f(A)$.

(b) Since f is continuous, the sets $U_\alpha = f^{-1}(O_\alpha)$ are open in X for each $\alpha \in I$. Also, if $a \in A$, then $f(a) \in \bigcup_{\alpha \in I} O_\alpha$. So $f(a) \in O_\beta$ for some $\beta \in I$. Then $a \in f^{-1}(O_\beta) \subseteq f^{-1}\left(\bigcup_{\alpha \in I} U_\alpha\right)$. This implies that $a \in \bigcup_{\alpha \in I} U_\alpha$ and so $A \subseteq \bigcup_{\alpha \in I} U_\alpha$. Thus, the sets U_α for $\alpha \in I$ form an open cover of A.

(c) The fact that A is compact means that there is a positive integer n and a finite collection $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n}$, of sets in $\{U_\alpha\}_{\alpha \in I}$ that cover A.

(d) We will prove that $f(A) \subseteq \bigcup_{i=1}^{n} O_{\alpha_i}$, which will complete our proof that every open cover of $f(A)$ in Y has a finite sub-cover.

Let $b \in f(A)$. Then $f^{-1}(b) \in A \subseteq \bigcup_{i=1}^{n} U_{\alpha_i}$. It follows that $f^{-1}(b) \in U_{\alpha_i}$ for some $1 \leq i \leq n$. Then $b \in f\left(U_{\alpha_i}\right) \subseteq)_{\alpha_i}$. Therefore,

$$f(A) \subseteq \bigcup_{i=1}^{n} O_{\alpha_i}.$$

This verifies that every open cover of $f(A)$ has a finite subcover.

A consequence of Activity 16.2 is that compactness is a topological property.

Corollary 16.5. *Let (X, τ_X) and (Y, τ_Y) be homeomorphic topological spaces. Then a subset A of X is compact if and only if $f(A)$ is compact in Y.*

Compact Subsets of $\mathbb{R}^n$

The metric space $(\mathbb{R}^n, d_E)$ is not compact since the open cover $\{B(0,n)\}_{n \in \mathbb{Z}^+}$ has no finite subcover. Since we have already shown that $(\mathbb{R}, d_E)$ is homeomorphic to the topological subspaces (a,b), $(-\infty, b)$, and (a, ∞) for any $a, b \in \mathbb{R}$, we conclude that no open intervals are compact. Similarly, no half-closed intervals are compact. In fact, we will demonstrate in this section that the compact subsets of $(\mathbb{R}^n, d_E)$ are exactly the subsets that are closed and bounded. The first step is contained in the next activity.

Activity 16.3. We have seen that compact sets can be either open or closed. However, in certain situations compact sets must be closed. We investigate that idea in this activity. Let A be a compact subset of a Hausdorff topological space X. We will examine why A must be a closed set.

(a) To prove that A is a closed set, we consider the set $X \setminus A$. What property of $X \setminus A$ will ensure that A is closed? How do we prove that $X \setminus A$ has this property?

(b) Let $x \in X \setminus A$. Assume that A is a non-empty set (why can we make this assumption)? For each $a \in A$, why must there exist disjoint open sets O_{xa} and O_a with $x \in O_{xa}$ and $a \in O_a$?

(c) Why must there exist a positive integer n and elements a_1, a_2, $\ldots$, a_n in A such that the sets $O_{a_1}, O_{a_2}, \ldots, O_{a_n}$ form an open cover of A?

(d) Now find an open subset of $X \setminus A$ that has x as an element. What does this tell us about A?

Activity Solution.

(a) To show that A is closed, we need to show that $X \setminus A$ is open. To show that $X \setminus A$ is open, we can demonstrate that $X \setminus A$ is a neighborhood of each of its points.

(b) Since $\emptyset$ is always closed, if $A = \emptyset$ we are done. So we can assume that $A \neq \emptyset$. Let $x \in X \setminus A$. Since X is Hausdorff, we can separate points by disjoint open sets. So for each $a \in A$, there exist disjoint open sets O_{xa} and O_a with $x \in O_{xa}$ and $a \in O_a$.

(c) The collection $\{O_a \mid a \in A\}$ is an open cover of A. The fact that A is compact means that there is a finite subcover for this cover. That is, there exist a positive integer n and elements $a_1, a_2, \ldots, a_n$ in A such that the sets $O_{a_1}, O_{a_2}, \ldots, O_{a_n}$ form an open cover of A.

(d) Let $O = \bigcap_{k=1}^{n} O_{xa_k}$. Note that if $a \in A$, then $a \in O_{a_k}$ for some k. But $O_{xa_k} \cap O_{a_k} = \emptyset$, so $a \notin O_{xa_k}$. Since $O \subseteq O_{xa_k}$, it is the case that $a \notin O$. Thus, $O \cap A = \emptyset$ and $O \subseteq X \setminus A$. We know that $x \in O_{xa_k}$ for every k, so $x \in O$. A finite union of open sets is open, so O is open. This means that $X \setminus A$ is a neighborhood of each of its points and is therefore open. So A is closed.

The result of Activity 16.3 is summarized in Theorem 16.6.

Theorem 16.6. *If A is a compact subset of a Hausdorff topological space, then A is closed.*

Theorem 16.6 tells us something about compact subsets of $(\mathbb{R}^n, d_E)$. Since every metric space is Hausdorff, we can conclude the following corollary.

Corollary 16.7. *If A is a compact subset of $(\mathbb{R}^n, d_E)$, then A is closed.*

To classify the compact subsets of $(\mathbb{R}^n, d_E)$ as closed and bounded, we need to discuss what it means for a set in $\mathbb{R}^n$ to be bounded. The basic idea is straightforward – a subset of $\mathbb{R}^n$ is bounded if it doesn't go off to infinity in any direction. In other words, a subset A of $\mathbb{R}^n$ is bounded if we can construct a box in $\mathbb{R}^n$ that is large enough to contain it. Thus, the following definition.

Definition 16.8. A subset A of $\mathbb{R}^n$ is **bounded** if there exists $M > 0$ such that $A \subseteq Q^n_M$, where

$$Q^n_M = \{(x_1, x_2, \ldots, x_n) \mid -M \leq x_i \leq M \text{ for every } 1 \leq i \leq n\}.$$

The set Q^n_M in Definition 16.8 is called the *standard n-dimensional cube of size M*. A standard 3-dimensional cube of size M is shown in Figure 16.1.

An important fact about standard n-cubes is that they are compact subsets of $\mathbb{R}^n$.

Theorem 16.9. *Let $n \in \mathbb{Z}^+$. The standard n-dimensional cube of size M is a compact subset of $\mathbb{R}^n$ for any $M > 0$.*

Proof. We proceed by contradiction and assume that there is an $n \in \mathbb{Z}^+$ and a positive real number M such that Q^n_M is not compact. So there exists an open cover $\{O_\alpha\}$ with α in some indexing set I of Q^n_M that has no finite sub-cover. Let $Q_0 = Q^n_M$ so that Q_0 is an n-cube with side length $2M$. Partition Q_0 into 2^n uniform sub-cubes of side length $M = \frac{2M}{2}$ (a picture for $n = 2$ is shown at left in Figure 16.2). Let Q_0' be one of these sub-cubes. The collection $\{O_\alpha \cap Q_0'\}_{\alpha \in I}$ is an open cover

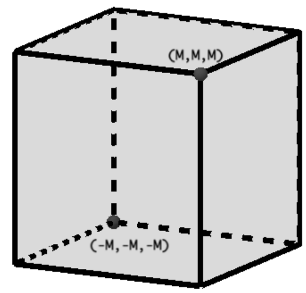


Figure 16.1: A standard 3-cube Q_M^3.

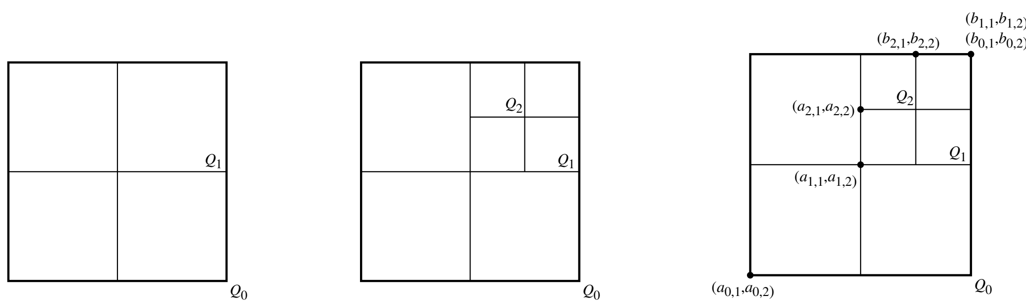


Figure 16.2: Left : Q_1. Middle: Q_2. Right: Labeling the corners.

of Q_0' in the subspace topology. If each of these open covers has a finite sub-cover, then we can take the union of all of the O_αs over all of the finite sub-covers to obtain a finite sub-cover of $\{O_\alpha\}_{\alpha \in I}$ for Q_0. Since our cover $\{O_\alpha\}_{\alpha \in I}$ for Q_0 has no finite sub-cover, we conclude that there is one sub-cube, Q_1, for which the open cover $\{O_\alpha \cap Q_1\}_{\alpha \in I}$ has no finite sub-cover. Now we repeat the process and partition Q_1 into 2^n uniform sub-cubes of side length $\frac{M}{2} = \frac{2M}{2^2}$. The same argument we just made tells us that there is a sub-cube Q_2 of Q_1 for which the open cover $\{O_\alpha \cap Q_2\}_{\alpha \in I}$ has no finite sub-cover (an illustration for the $n = 2$ case is shown at middle in Figure 16.2). We proceed inductively to obtain an infinite nested sequence

$$Q_0 \supset Q_1 \supset Q_2 \supset Q_3 \supset \cdots \supset Q_k \supset \cdots$$

of cubes such that for each $k \in \mathbb{Z}$, the lengths of the sides of cube Q_k are $\frac{M}{2^{k-1}} = \frac{2M}{2^k}$ and the open cover $\{O_\alpha \cap Q_k\}_{\alpha \in I}$ of Q_k has no finite sub-cover. Now we show that $\bigcap_{k=1}^\infty Q_k \neq \emptyset$.

For $i \in \mathbb{Z}^+$, let $Q_i = [a_{i,1}, b_{i,1}] \times [a_{i,2}, b_{i,2}] \times \cdots [a_{i,n}, b_{i,n}]$. That is, think of the point $(a_{i,1}, a_{i,2}, \ldots, a_{i,n})$ as a lower corner of the cube and the point $(b_{i,1}, b_{i,2}, \ldots, b_{i,n})$ as an upper corner of the n-cube Q_i (a labeling for $n = 2$ and i from 1 to 3 is shown at right in Figure 16.2). Let $q = (\sup\{a_{i,1}\}, \sup\{a_{i,2}\}, \ldots, \sup\{a_{i,n}\})$. We will show that $q \in \bigcap_{k=1}^\infty Q_k$. Fix $r \in \mathbb{Z}^+$. We need to demonstrate that

$$q \in Q_r = \{(x_1, x_2, \ldots, x_n) \mid a_{r,s} \leq x_s \leq b_{r,s} \text{ for each } 1 \leq s \leq n\}.$$

For each s between 1 and n we have

$$a_{r,s} \leq \sup\{a_{i,s}\} \tag{16.1}$$

because $\sup\{a_{i,s}\}$ is an upper bound for all of the $a_{i,s}$. The fact that our cubes are nested means that

$$\begin{aligned} &a_{1,s} \leq a_{2,s} \leq \cdots, \\ &b_{1,s} \geq b_{2,s} \geq \cdots, \\ &a_{i,s} \leq b_{i,s} \end{aligned} \tag{16.2}$$

for every i and s. Since $\sup\{a_{i,s}\}$ is the least upper bound of all of the $a_{i,s}$, property (16.2) shows that $\sup\{a_{i,s}\} \leq b_{i,s}$ for every i. Thus, $\sup\{a_{i,s}\} \leq b_{r,s}$ and so $a_{r,s} \leq \sup\{a_{i,s}\} \leq b_{r,s}$. This shows that $q \in Q_k$ for every k. Consequently, $q \in \bigcap_{k=1}^{\infty} Q_k$ and $\bigcap_{k=1}^{\infty} Q_k$ is not empty. (The fact that the side lengths of our cubes are converging to 0 implies that $\bigcap_{k=1}^{\infty} Q_k = \{q\}$, but we only need to know that $\bigcap_{k=1}^{\infty} Q_k$ is not empty for our proof.)

Since $\{O_\alpha\}_{\alpha \in I}$ is a cover for Q_0, there must exist an $\alpha_q \in I$ such that $q \in O_{\alpha_q}$. The set O_{α_q} is open, so there exists $\epsilon_q > 0$ such that $B(q, \epsilon_q) \subseteq O_{\alpha_q}$. The maximum distance between points in Q_k is the distance between the corner points $(a_{k,1}, a_{k,2}, \ldots, a_{k,n})$ and $(b_{k,1}, b_{k,2}, \ldots, b_{k,n})$, where each length $b_{k,s} - a_{k,s}$ is $\frac{M}{2^{k-1}}$. The distance formula tells us that this maximum distance between points in Q_k is

$$D_k = \sqrt{\sum_{s=1}^{n} \left(\frac{M}{2^{k-1}}\right)^2} = \sqrt{n\left(\frac{M}{2^{k-1}}\right)^2} = \frac{M}{2^{k-1}}\sqrt{n}.$$

Now choose $K \in \mathbb{Z}^+$ such that $D_K < \epsilon_q$. Then if $x \in Q_K$ we have $d_E(q, x) < D_K$ and $x \in B(q, \epsilon_q)$. So $Q_K \subseteq B(q, \epsilon_q)$. But $B(q, \epsilon_q) \subseteq O_{\alpha_q}$. So the collection $\{O_{\alpha_q} \cap Q_K\}$ is a sub-cover of $\{O_\alpha \cap Q_K\}_{\alpha \in I}$ for Q_K. But this contradicts the fact this open cover has no finite sub-cover. The assumption that led us to this contradiction was that Q_0 was not compact, so we conclude that the standard n-dimensional cube of size M is a compact subset of $\mathbb{R}^n$ for any $M > 0$. ■

One consequence of Theorem 16.9 is that any closed interval $[a, b]$ in $\mathbb{R}$ is a compact set. But we can say even more – that the compact subsets of $\mathbb{R}^n$ are the closed and bounded subsets. This will require one more intermediate result about closed subsets of compact topological spaces.

Activity 16.4. Let X be a compact topological space and C a closed subset of X. In this activity we will prove that C is compact.

(a) What does it take to prove that C is compact?

(b) Use an open cover for C and the fact that C is closed to make an open cover for X.

(c) Use the fact that X is compact to complete the proof of the following theorem.

Theorem 16.10. *Let X be a compact topological space. Then any closed subset of X is compact.*

Activity Solution.

(a) We need to show that every open cover of C has a finite subcover.

(b) Let $\{O_\alpha\}$ be an open cover of C for α in some indexing set I. Since C is closed, the set $X \setminus C$ is an open set. Then $\{O_\alpha\} \cup (X \setminus C)$ is an open cover of X.

(c) Since X is compact, this open cover has a finite sub-cover. In other words, there are sets $O_{\alpha_1}, O_{\alpha_2}, \ldots, O_{\alpha_n}$ for some positive integer n such that $\alpha_i \in I$ for each i and

$$X \subseteq O_{\alpha_1} \cup O_{\alpha_2} \cup \cdots \cup O_{\alpha_n} \cup (X \setminus C).$$

But then we must have

$$C \subseteq O_{\alpha_1} \cup O_{\alpha_2} \cup \cdots \cup O_{\alpha_n}$$

and we have found a finite sub-cover of the open cover $\{O_\alpha\}$. We conclude that C is compact.

Now we can prove a major result, that the compact subsets of $(\mathbb{R}^n, d_E)$ are the closed and bounded subsets. This result is important enough that it is given a name.

Theorem 16.11 (The Heine-Borel Theorem)**.** *A subset A of $(\mathbb{R}^n, d_E)$ is compact if and only if A is closed and bounded.*

Proof. Let A be a subset of $(\mathbb{R}^n, d_E)$. Assume that A is closed and bounded. Since A is bounded, there is a positive number M such that $A \subseteq Q^n_M$. Theorem 16.9 shows that Q^n_M is compact, and then Theorem 16.10 shows that A is compact.

For the converse, assume that A is a compact subset of $\mathbb{R}^n$. We must show that A is closed and bounded. Now $(\mathbb{R}^n, d_E)$ is a metric space, and so Hausdorff. Theorem 16.6 then shows that A is closed. To conclude our proof, we need to demonstrate that A is bounded. For each $k > 0$, let

$$O_k = \{(x_1, x_2, \ldots, x_n) \mid -k < x_i < k \text{ for every } 1 \leq i \leq n\}.$$

That is, O_k is the open k-cube in $\mathbb{R}^n$. Next let

$$U_k = O_k \cap A$$

for each k. Since $\bigcup_{k>0} O_k = \mathbb{R}^n$, it follows that $\{U_k\}_{k>0}$ is an open cover of A. The fact that A is compact means that there is a finite collection $U_{k_1}, U_{k_2}, \ldots, U_{k_n}$ of sets in $\{U_k\}_{k>0}$ that cover A. Let $K = \max\{k_i \mid 1 \leq i \leq n\}$. Then $U_{k_i} \subseteq U_K$ for each i, and so $A \subseteq U_K \subset Q^n_K$. Thus, A is bounded. This completes the proof that if A is compact in $\mathbb{R}^n$, then A is closed and bounded. ■

You might wonder whether the Heine-Borel Theorem is true in any metric space.

Activity 16.5. A subset A of a metric space (X, d) is bounded if there exists a real number M such that $d(a_1, a_2) \leq M$ for all $a_1, a_2 \in A$. (This is equivalent to our definition of a bounded subset of $\mathbb{R}^n$ given earlier, but works in any metric space.) Explain why $\mathbb{Z}$ as a subset of $(\mathbb{R}, d)$, where d is the discrete metric, is closed and bounded but not compact.

Activity Solution. Let $M = 1$. For any $a_1, a_2 \in \mathbb{Z}$, we have that $d(a_1, a_2) \leq 1$. So $\mathbb{Z}$ is bounded. Every subset of a metric space with the discrete metric is closed, so $\mathbb{Z}$ is closed. But $\mathbb{Z}$ is not compact since open cover with no finite sub-cover is $\{\{n\}\}_{n \in \mathbb{Z}}$.

An Application of Compactness

As mentioned at the beginning of this section, compactness is the quality we need to ensure that continuous functions from topological spaces to $\mathbb{R}$ attain their maximum and minimum values.

Activity 16.6. In this activity we prove the following theorem.

Theorem 16.12. *A continuous function from a compact topological space to the real numbers assumes a maximum and minimum value.*

(a) Let X be a compact topological space and $f : X \to \mathbb{R}$ a continuous function. What does the continuity of f tell us about $f(X)$ in $\mathbb{R}$?

(b) Why can we conclude that the set $f(X)$ has a least upper bound M? Why must M be an element of $f(X)$?

(c) Complete the proof of Theorem 16.12.

Activity Solution.

(a) Since f is continuous, we know that $f(X)$ is compact in $\mathbb{R}$ by Activity 16.2.

(b) By the Heine-Borel Theorem, it follows that $f(X)$ is closed and bounded. Since $f(X)$ is bounded, there is a least upper bound M for $f(X)$. Now M is a limit point of $f(X)$ and $f(X)$ is closed, so $M \in f(X)$.

(c) Similarly, $f(X)$ contains its greatest lower bound m. Thus, f assumes a maximum value M and a minimum value m in $\mathbb{R}$.

Summary

Important ideas that we discussed in this section include the following.

- A cover of a subset A of a topological space X is any collection of subsets of X whose union contains A. An open cover is a cover consisting of open sets.
- A subcover of a cover of a set A is a subset of the cover such that the union of the sets in the subcover also contains A.
- A subset A of a topological space is compact if every open cover of A has a finite subcover.
- A continuous function from a compact topological space to the real numbers must attain a maximum and minimum value.
- The Heine-Borel Theorem states that the compact subsets of $\mathbb{R}^n$ are exactly the subsets that are closed and bounded.

Section 17

Connected Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a connected subset of a topological space?
- What is a separation of a subset of a topological space? Why are separations useful?
- What are the connected subsets of $\mathbb{R}$?
- What is a connected component of a topological space?
- What is an application of connectedness?
- What is a cut set of a topological space? Why are cut sets useful?

Introduction

The term "connected" should bring up images of something that is one piece, not separated. There is more than one way we can interpret the notion of connectedness in topological spaces. For example, we might consider a topological space to be connected if we can't separate it into disjoint pieces in any non-trivial way. As another possibility, we might consider a topological space to be connected if there is always a path from one point in the spaced to another, provided we define what "path' means. These are different notions of connectedness, and we focus on the first notion in this section.

Connectedness is an important property, and one that we encounter in the calculus. For example, we will see in this section that the Intermediate Value Theorem relies on connected subsets of $\mathbb{R}$. To define a connected set, we will need to have a way to understand when and how a set can be separated into different pieces. Since a topology is defined by open sets, when we want to separate objects we will do so with open sets. This is similar to the idea behind Hausdorff spaces, except that we now want to know if a set can be separated in some way rather than separating points.

As an example to motivate the definition, consider the sets $X = (0,1) \cup (1,2)$ and $Y = [1,2]$ in $\mathbb{R}$ with the Euclidean topology. Notice that we can write X as the union of two disjoint open sets $X_1 = (0,1)$ and $X_2 = (1,2)$. So we shouldn't think of X as being connected. However, if we attempt to write Y as a union of two subsets, say $Y_1 = [1, 1.5)$ and $Y_2 = [1.5, 2]$, it is impossible for both of these subsets to be open and disjoint. So Y is a set we should consider to be connected. This is the notion of connectedness that we wish to investigate.

Definition 17.1. A subset A of a topological space (X, τ) is **connected** if A cannot be written as the union of two disjoint, nonempty, proper, relatively open subsets in the subspace topology. A topological space X is connected if X is a connected subset of X.

If a set A is not connected, we say that A is disconnected.

Preview Activity 17.1. Can the subset A of the topological space X be written as the union of two disjoint non-empty proper open sets?

(1) The set $A = \{a, b\}$ in (X, τ) with $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$.

(2) The set $A = \{a, b, c\}$ in (X, τ) with $X = \{a, b, c, d, e, f\}$ and
$$\tau = \{\emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}, X\}.$$

(3) The set $A = X$ with $X = \{a, b, c, d\}$ and
$$\tau = \{\emptyset, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, b, c\}, \{b, c, d\}, X\}.$$

(4) The set $A = \{d, f\}$ in $X = \{a, b, c, d, e, f\}$ with the discrete topology. Generalize your findings.

(5) The set $A = \{a, c, d\}$ in $X = \{a, b, c, d, e\}$ with the indiscrete topology. Generalize your findings.

(6) The set $A = \mathbb{Z}$ in $X = \mathbb{R}$ with the finite complement topology. Generalize your findings.

(7) The set $A = X$ in $X = \{x \in \mathbb{R} \mid 1 \leq x \leq 2 \text{ or } 3 < x < 4\}$ with the subspace metric topology from $(\mathbb{R}, d_E)$.

(8) The set $A = X$ in $X = \left\{(x, y) \in \mathbb{R}^2 \mid y = \frac{1}{x} \text{ or } y = 0\right\}$ with open sets
$$\tau = \{U \cap X \mid U \text{ is open in the Euclidean Topology on } \mathbb{R}^2\}.$$

Activity Solution.

(1) The answer is yes. Note that $A = \{a\} \cup \{b\}$, where $\{a\} = A \cap \{a\}$ and $\{b\} = A \cap \{b\}$ are relatively open sets.

(2) The answer is yes. Both $\{a\} = A \cap \{a\}$ and $\{b, c\} = A \cap \{b, c, d, e, f\}$ are relatively open sets and $A = \{a\} \cup \{b, c\}$.

(3) The answer is no. The only disjoint open sets of X are $\{b\}$ and $\{c\}$ and their union is not X.

(4) The answer is yes. Since every subset of X is open, we can write A as a disjoint union of the open sets $\{d\}$ and $\{f\}$. This argument can be applied to any subset of X containing two or more elements. To see why, let (X, τ) be a topological space where τ is the discrete topology, and let A be a subset of X with two or more elements. Let $a \in A$ and let $O = \cup_{\substack{x \in A \\ x \neq a}}$. Note that O is open and A is the disjoint union of $\{a\}$ with O.

(5) The are no proper open sets in X, so no proper relatively open sets. Thus, A cannot be written as a disjoint union of nonempty proper open sets. This applies to any space with the indiscrete topology.

(6) We have seen that the subspace topology of an infinite subset of a topological space with the finite complement topology is the finite complement topology. Let U and V be relatively open sets in $\mathbb{Z}$. Then $\mathbb{Z} \setminus (U \cap V)$ is a finite set. This means that $U \cap V$ is an infinite set. So either U is infinite or V is infinite, and there are no disjoint relatively open sets in $\mathbb{Z}$. Thus $\mathbb{Z}$ cannot be written as a disjoint union of nonempty proper open sets.

Recall that if X is a finite set, the finite complement topology on X is just the discrete topology. We discussed the discrete topology in problem 3. If X is an infinite set with the finite complement topology, then the same argument as above shows that no subset of X can be written as a disjoint union of nonempty proper relatively open sets.

(7) Notice that $[1, 2] = (0, 2.5) \cap X$ and $(3, 4) = (3, 4) \cap X$, so both $[1, 2]$ and $(3, 4)$ are open in X. Then X can be written as a disjoint union of nonempty proper open sets as $X = [1, 2] \cup (3, 4)$.

(8) Let $Q_1 = \{(x, y) \mid | \ x > 0 \text{ and } y > 0\}$ and let $Q_4 = \{(x, y) \mid | \ x < 0 \text{ and } y < 0\}$. Then Q_1 and Q_4 are open sets in $\mathbb{R}^2$, so $O_1 = Q_1 \cap A$ and $O_2 = Q_4 \cap A$ are relatively open in A. Note that O_1 is the portion of the graph of $y = \frac{1}{x}$ in the first quadrant and O_2 is the portion of the graph of $y = \frac{1}{x}$ in the fourth quadrant. Now let $b \in \mathbb{R}$. The distance from the point $(0, b)$ to the graph of $y = \frac{1}{x}$ is positive, so let ϵ_b be that distance. Then the set $\{(0, b) \mid | \ x \in \mathbb{R}\}$ can be written as $O = \cup_{b \in \mathbb{R}} B((0, b), \epsilon_b)$. The sets O_1, O_2, and O are disjoint and relatively open, and $A = (O_1 \cup O_2) \cup O$.

Connected Sets

As we learned in our preview activity, connected sets are those sets that cannot be separated into a union of disjoint open sets. Another characterization of connectedness is established in the next activity.

Activity 17.1. Let (X, τ) be a topological space.

(a) Assume that X is a connected space, and let A be a subset of X that is both open and closed. What happens if we combine A and $X \setminus A$? What does the fact that X is connected tell us about A?

(b) Now assume that the only subsets of X that are both open and closed are $\emptyset$ and X. Must it follow that X is connected? Prove your assertion.

(c) Summarize the result of this activity into a theorem of the form "A topological space (X, τ) is connected if and only if ... ".

Activity Solution.

(a) The set $X = A \cup (X \setminus A)$ expresses X as a disjoint union of two subsets. Since X is connected, A cannot be a proper subset of X, so $A = \emptyset$ or $A = X$.

(b) Suppose that A and B are disjoint open subsets of X such that $X = A \cup B$. It follows that $B = X \setminus A$ and so A is closed. Since A is both open and closed, we can conclude that $A = \emptyset$ or $A = X$. This implies that there are no proper disjoint open subsets of X whose union is X. Therefore, X is connected.

(c) The appropriate theorem is

theorem A topological space (X, τ) is connected if and only if the only sets that are both open and closed in X are $\emptyset$ and X.

A standard example of a connected topological space is the metric space $(\mathbb{R}, d_E)$.

Theorem 17.2. *The metric space $(\mathbb{R}, d_E)$ is a connected topological space.*

Proof. We proceed by contradiction and assume that there are non-empty open sets U and V such that $\mathbb{R} = U \cup V$ and $U \cap V = \emptyset$. Let $a \in U$ and $b \in V$. Since $U \cap V = \emptyset$, we know that $a \neq b$. Without loss of generality we can assume $a < b$. Let $U' = U \cap [a, b]$ and let $V' = V \cap [a, b]$. The set V' is bounded below by a, so $x = \inf\{v \mid v \in V'\}$ exists. Since $\mathbb{R} = U \cup V$ it must be the case that $x \in U$ or $x \in V$.

Suppose $x \in U$. The fact that U is an open set implies that there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$. But then $B(x, \epsilon) \cap V = \emptyset$ and so $d(x, v) \geq \epsilon$ for every $v \in V$. This means that $x + \epsilon < v$ for every $v \in V'$, contradicting the fact that x is the greatest lower bound. We conclude that $x \notin U$.

It follows that $x \in V$. Since $a \in U$, we know that $x \neq a$. The fact that V is an open set tells us that there exists $\delta > 0$ such that $B(x, \delta) \subseteq V$. We can choose δ to ensure that $\delta < x - a$. Since $x > a$, the interval $(x - \delta, x)$ is a subset of V', and so x is not a lower bound for V.

Each possibility leads to a contradiction, so we conclude that the sets U and V cannot exist. Therefore, $(\mathbb{R}, d_E)$ is a connected topological space. ■

As you might expect, connectedness is a topological property.

Activity 17.2. Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be a continuous function. Assume that A is a connected subset of X. Our goal is to prove that $f(A)$ is a connected subset of Y.

(a) Assume to the contrary that $f(A)$ is not connected. What do we then assume about $f(A)$?

(b) Suppose that U and V form a separation of $f(A)$ in Y. Show that $R = f^{-1}(U)$ and $S = f^{-1}(V)$ form a separation of A in X. Explain how we have proved the following.

Theorem 17.3. *Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be a continuous function. If A is a connected subset of X, then $f(A)$ is a connected subset of Y.*

Activity Solution.

(a) Assuming that $f(A)$ is not connected, it follows that there is a separation U and V of $f(A)$ in Y.

(b) Since f is continuous, we know that $R = f^{-1}(U)$ and $S = f^{-1}(V)$ are open in X. First note that

$$A =\subseteq f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V) = R \cup S.$$

The fact that $U \cap f(A) \neq \emptyset$ implies that there is an element $y \in U \cap f(A)$. Thus, $y \in f(A)$ and there is an $x \in A$ with $f(x) = y$. So $x \in f^{-1}(U)$ and $x \in A$. Thus, $x \in R \cap A$. Similarly, $S \cap A \neq \emptyset$. Finally,

$$R \cap S \cap A = f^{-1}(U) \cap f^{-1}(V) \cap A = f^{-1}(U \cap V) \cap A.$$

So if $x \in R \cap S \cap A$, then $x \in f^{-1}(U \cap V)$ and $x \in A$. So $f(x) \in U \cap V \cap f(A) = \emptyset$. It follows that $R \cap S \cap A = \emptyset$.

We have then found a separation of A, which contradicts the fact that A is connected. So we conclude that $f(A)$ is connected. The theorem is just the contrapositive of the result we proved.

The fact that connectedness is preserved by continuous functions means that connectedness is a property that is shared by any homeomorphic topological spaces, as the next corollary indicates.

Corollary 17.4. *Let (X, τ_X) and (Y, τ_Y) be homeomorphic topological spaces. Then X is connected if and only if Y is connected.*

Proof. Let (X, τ_X) and (Y, τ_Y) be topological spaces and let $f : X \to Y$ be a homeomorphism. Assume that X is connected. Since f is continuous, Theorem 17.3 shows that $f(X) = Y$ is connected. The reverse implication follows from the fact that f^{-1} is a homeomorphism. ∎

Recall that $(\mathbb{R}, d_E)$ is homeomorphic to the topological subspaces (a, b), $(-\infty, b)$, and (a, ∞) for any $a, b \in \mathbb{R}$. The fact that $(\mathbb{R}, d_E)$ is connected (Theorem 17.2) allows us to conclude that all open intervals are connected. It would seem natural that all closed (or half-closed) intervals should also be connected. We address this question next. Before we get to this result, we consider an alternate formulation of connected subsets.

Consider the set $A = (-1, 0) \cup (4, 5)$ in $\mathbb{R}$. Let $U = (-2, 3)$ and $V = (2, 6)$ in $\mathbb{R}$. Note that $U' = U \cap A = (-1, 0)$ and $V' = V \cap A = (4, 5)$, and so U and V are open sets in $\mathbb{R}$ that separate the set A into two disjoint pieces. We know that U' and V' are open in A and $A = U' \cup V'$ with $U' \cap V' = \emptyset$. So to show that a subset of a topological space X is not connected, this example suggests that it suffices to find non-empty open sets U and V in X with $U \cap V \cap A = \emptyset$ and $A \subseteq (U \cup V)$. Note that it is not necessary to have $U \cap V = \emptyset$. That this works in general is the result of the next theorem.

Theorem 17.5. *Let X be a topological space. A subset A of X is disconnected if and only if there exist open sets U and V in X with*

- $A \subseteq (U \cup V)$,
- $U \cap A \neq \emptyset$,
- $V \cap A \neq \emptyset$, *and*
- $U \cap V \cap A = \emptyset$.

Proof. Let X be a topological space, and let A be a subset of X. We first assume that A is disconnected and show that there are open sets U and V in X that satisfy the given conditions. Since A is disconnected, there are non-empty open sets U' and V' in A such that $U' \cup V' = A$ and $U' \cap V' = \emptyset$. Since U' and V' are open in A, there exist open sets U and V in X so that $U' = U \cap A$ and $V' = V \cap A$. Now

$$A = U' \cup V' = (U \cap A) \cup (V \cap A) = (U \cup V) \cap A,$$

and so $A \subseteq U \cup V$. By construction, $U \cap A = U'$ and $V \cap A = V'$ are not empty. Finally,

$$U \cap V \cap A = (U \cap A) \cap (V \cap A) = U' \cap V' = \emptyset.$$

So we have found sets U and V that satisfy the conditions of our theorem.

The proof of the reverse implication is left to the next activity.

■

Activity 17.3. Let X be a topological space, and let A be a subset of X. Assume that there exist open sets U and V in X with $A \subseteq U \cup V$, $U \cap A \neq \emptyset$, $V \cap A \neq \emptyset$, and $U \cap V \cap A = \emptyset$. Prove that A is disconnected.

Activity Solution. For the reverse implication, assume that there exist open sets U and V in X with $A \subseteq U \cup V$, $U \cap A \neq \emptyset$, $V \cap A \neq \emptyset$, and $U \cap V \cap A = \emptyset$. Let $U' = U \cap A$ and $V' = V \cap A$. By construction, $U' \neq \emptyset$ and $V' \neq \emptyset$. Also,

$$U' \cap V' = (U \cap A) \cap (V \cap A) = U \cap V \cap A = \emptyset.$$

Since U and V are open in X, we know that U' and V' are open in A. To complete the proof that A is disconnected, it remains to show that $A = U' \cup V'$. Since $A \subseteq U \cup V$, it follows that

$$A = (U \cup V) \cap A = (U \cap A) \cup (V \cap A) = U' \cup V'.$$

Therefore, A is a disconnected set.

The conditions in Theorem 17.5 provide a convenient way to show that a set is disconnected, and so any pair of sets U and V that satisfy the conditions of Theorem 17.5 is given a special name.

Definition 17.6. Let X be a topological space, and let A be a subset of X. A **separation** of A is a pair of non-empty open subsets U and V of X such that

- $A \subseteq (U \cup V)$,

- $U \cap A \neq \emptyset$,
- $V \cap A \neq \emptyset$, and
- $U \cap V \cap A = \emptyset$.

Connected Subsets of $\mathbb{R}$

With Theorem 17.5 in hand, we are just about ready to show that any interval in $\mathbb{R}$ is connected. Let us return for a moment to our example of $A = (-1, 0) \cup (4, 5)$ in $\mathbb{R}$. It is not difficult to see that if U and V are a separation of A, then the subset $(-1, 0)$ must be entirely contained in either U or in V. The reason for this is that $(-1, 0)$ is a connected subset of A. This result is true in general.

Activity 17.4. Let X be a topological space, and let A be a subset of X. Assume that U and V form a separation of A. Let C be a connected subset of A. In this activity we want to prove that $C \subseteq U$ or $C \subseteq V$.

(a) If we proceed by contradiction, what additional assumption(s) do we make?

(b) What conclusion can we draw about the sets $U' = U \cap C$ and $V' = V \cap C$?

(c) Complete the proof of the following lemma.

Lemma 17.7. *Let X be a topological space, and let A be a subset of X. Assume that U and V form a separation of A. If C is a connected subset of A, then $C \subseteq U$ or $C \subseteq V$.*

Activity Solution.

(a) To prove that $C \subseteq U$ or $C \subseteq V$, we proceed by contradiction and assume that $C \not\subseteq U$ and $C \not\subseteq V$.

(b) Then $U' = U \cap C \neq \emptyset$ and $V' = V \cap C \neq \emptyset$, and so U' and V' form a separation of C in X.

(c) Part (b) implies that C is disconnected, a contradiction. We conclude that $C \subseteq U$ or $C \subseteq V$.

Now we can prove that any interval in $\mathbb{R}$ is connected. Since $[a, b]$, $[a, b)$, and $(a, b]$ are all sets that lie between (a, b) and $\overline{(a, b)}$, we can address their connectedness all at once with the next result.

Theorem 17.8. *Let X be a topological space and C a connected subset of X. If A is a subset of X and $C \subseteq A \subseteq \overline{C}$, then A is connected in X.*

Proof. Let X be a topological space and C a connected subset of X. Let A be a subset of X such that $C \subseteq A \subseteq \overline{C}$. To show that A is connected, assume to the contrary that A is disconnected. Then there are non-empty open subsets U and V of X that form a separation of A. Lemma 17.7 shows that $C \subseteq U$ or $C \subseteq V$. Without loss of generality we assume that $C \subseteq U$. Since $U \cap V \cap A = \emptyset$, it follows that

$$C \cap V = (C \cap A) \cap V = C \cap (A \cap V) \subseteq U \cap A \cap V = \emptyset.$$

Since $A \cap V \neq \emptyset$, there is an element $x \in A \cap V$. Since $x \notin C$ and $x \in A \subseteq \overline{C}$, it must be the case that x is a limit point of C. Since V is an open neighborhood of x, it follows that $V \cap C \neq \emptyset$. This contradiction allows us to conclude that A is connected. ■

One consequence of Theorem 17.8 is that any interval of the form $[a, b)$, $(a, b]$, $[a, b]$, $(-\infty, b]$, or $[a, \infty)$ in $\mathbb{R}$ is connected. This prompts the question, are there any other subsets of $\mathbb{R}$ that are connected?

Activity 17.5. Let A be a subset of $\mathbb{R}$.

(a) Let $A = \{a\}$ be a single point subset of $\mathbb{R}$. Is A connected? Explain.

(b) Now suppose that A is a subset of $\mathbb{R}$ that contains two or more points. Assume that A is not an interval. Then there must exist points a and b in A and a point c in $\mathbb{R} \setminus A$ between a and b. Use this idea to find a separation of A. What can we conclude about A?

(c) Explicitly describe the connected subsets of $(\mathbb{R}, d_E)$.

Activity Solution.

(a) Suppose U and V form a separation of A. Since $A \subseteq (U \cup V)$, it follows that $a \in U$ or $a \in V$. so $A \subseteq U$ or $A \subseteq V$ and A is connected.

(b) Let $U = (-\infty, c)$ and $V = (c, \infty)$. Since $c \notin A$, $A \subseteq (U \cup V)$. By definition, $A \cap U \cap V = \emptyset$, $A \cap U \neq \emptyset$, and $A \cap V \neq \emptyset$. So U and V form a separation of A and A is not connected.

(c) The connected subsets of $\mathbb{R}$ are the intervals and the single point sets.

Components

As Activity 17.5 demonstrates, spaces like $A = (1, 2) \cup (3, 4)$ are not connected. Even so, A is made of two connected subsets $(1, 2)$ and $(3, 4)$. These connected subsets are called *components*.

Definition 17.9. A subspace C of a topological space X is a **component** (or **connected component**) of X if C is connected and there is no larger connected subspace of X that contains C.

As an example, if $X = (1, 2) \cup [4, 10) \cup \{-1, 15\}$, then the components of X are $(1, 2)$, $[4, 10)$, $\{-1\}$ and $\{15\}$. As the next activity shows, we can always partition a topological space into a disjoint union of compenents.

Activity 17.6. Let (X, τ) be a nonempty topological space. We can isolate the connected subsets of $A = (1, 2) \cup (3, 4)$ by identifying points that lie in the same connected subset. In other words, define a relation $\sim$ on $A = (1, 2) \cup (3, 4)$ as follows: $a \sim b$ if a and b are elements of the same connected subset of A.

(a) Show that if $x \in X$, then $\{x\}$ is connected.

(b) Suppose that X is a topological space and $\{A_\alpha\}$ for α in some indexing set I is a collection of connected subsets of X. Let $A = \bigcup_{\alpha \in I} A_\alpha$. Suppose that $\bigcap_{\alpha \in I} A_\alpha \neq \emptyset$. Show that A is a connected subset of X. (Hint: Assume a separation and use Lemma 17.7.)

(c) Part (a) shows that every element in x belongs to come connected subset of X. So we can write X as a union of connected subsets. But there is probably overlap. To remove the overlap, we define the following relation $\sim$ on X:

For x and y in X, $x \sim y$ if x and y are contained in the same connected subset of X.

As with any relation, we ask if $\sim$ is an equivalence relation.

i. Is $\sim$ a reflexive relation? Why or why not?
ii. Is $\sim$ a symmetric relation? Why or why not?
iii. Is $\sim$ a transitive relation? Why or why not?

Activity Solution.

(a) Suppose $\{x\} = U \cup V$ with U and V nonempty relatively open sets and $U \cap V = \emptyset$. Since $\{x\} = U \cup V$, it is the case that $x \in U$ or $x \in V$. But if $x \in U$, then $V = \emptyset$, and if $x \in V$, then $U = \emptyset$. Either contradicts the assumption that U and V are nonempty.

(b) Suppose U and V form a separation of A. Let $a \in \bigcap_{\alpha \in I} A_\alpha$. Let $U' = U \cap A$ and $V' = V \cap A$. The fact that $U' \cap V' = \emptyset$ means that $a \in U$ or $a \in V$, but not both. Without loss of generality, assume that $a \in U$. Let $\alpha \in I$. Since A_α is connected, Lemma 17.7 tells us that A_α is a subset of U or a subset of V. But $a \in U$ and $a \in A_\alpha$, so it must be that $A_\alpha \subseteq U$. But α was an arbitrary element of I, so we can conclude that $A \subseteq U$. But then $V \cap A = \emptyset$, a contradiction. Therefore, A must be a connected subset of X.

i. Let $x \in X$. Since $\{x\}$ is connected, we have $x \sim x$.
ii. Let x and y be in X with $x \sim y$. Then x and y both lie in some connected subset C. Thus y and x are both in C and so $y \sim x$.
iii. Suppose $x \sim y$ and $y \sim z$ for some x, y, and z in X. Then x and y are in some connected subset C_1 of X. Similarly, y and z are in some connected subset C_2 of X. But C_1 and C_2 share the point y, so $C_1 \cup C_2$ is connected by pat (b). Thus, x and z are elements of the same connected subset of X and $x \sim z$.

Activity 17.6 shows that the relation $\sim$ is an equivalence relation, and so partitions the underlying topological space X into disjoint sets. If $x \in X$, then the equivalence class of x is a connected subset of X. There can be no larger connected subset of X that contains x, since equivalence classes are disjoint or the same. So the equivalence classes are exactly the connected components of X. The components of a topological space X satisfy several properties.

- Each $a \in X$ is an element of exactly one connected component C_a of X.
- A component C_a contains all connected subsets of X that contain a. Thus, C_a is the union of all connected subsets of X that contain a.
- If a and b are in X, then either $C_a = C_b$ or $C_a \cap C_b = \emptyset$.
- Every connected subset of X is a subset of a connected component.
- Every connected component of X is a closed subset of X.
- The space X is connected if and only if X has exactly one connected component.

Two Applications of Connectedness

The Intermediate Value Theorem from calculus tells us that if f is a continuous function on a closed interval $[a, b]$, then f assumes all values between $f(a)$ and $f(b)$. This is a result that seems obvious with a picture, and we take the result for granted. Now we can prove this theorem.

Theorem 17.10 (The Intermediate Value Theorem)**.** *Let X be a topological space and A a connected subset of X. If $f : A \to \mathbb{R}$ is a continuous function, then for any $a, b \in A$ and any $y \in \mathbb{R}$ between $f(a)$ and $f(b)$, there is a point $x \in A$ such that $f(x) = y$.*

Activity 17.7. In this activity we prove the Intermediate Value Theorem. Let X be a topological space and A a connected subset of X. Assume that $f : A \to \mathbb{R}$ is a continuous function, and let $a, b \in A$.

(a) Explain why we can assume that $a \neq b$.

(b) Explain what happens if $y = f(a)$ or $y = f(b)$.

(c) Now assume that $f(a) \neq f(b)$. Without loss of generality, assume that $f(a) < f(b)$. Why can we say that $f(A)$ is an interval?

(d) How does the fact that $f(A)$ is an interval complete the proof?

Activity Solution.

(a) If $a = b$, then the only y between $f(a)$ and $f(b)$ is $f(a)$, which is a value attained by f on $[a, b]$.

(b) If $y = f(a)$, then we take $c = a$. If $y = f(b)$, then we take $c = b$.

(c) Since A is connected in X, Theorem 17.3 shows that $f(A)$ is a connected subset of $\mathbb{R}$. We proved that the connected subsets of $\mathbb{R}$ are single point sets or intervals. Since $f(a) < f(b)$, we know that $f(A)$ is not a single point set, so $f(A)$ is an interval.

(d) Since $f(A)$ is an interval, it contains all values between $f(a)$ and $f(b)$. So if $f(a) < y < f(b)$, then $y \in f(A)$.

Fixed point theorems are important in mathematics. A fixed point of a function f is an input c so that $f(c) = c$. There are many fixed point theorems – one of the most well-known is the Brouwer Fixed Point Theorem that shows that every continuous function from a closed ball B in $\mathbb{R}^n$ to itself must have a fixed point. We can use the Intermediate Value Theorem to prove this result in $\mathbb{R}$.

Activity 17.8. In this activity we prove the following theorem.

Theorem 17.11. *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a, b] \to [a, b]$ be a continuous function. Then there is a number $c \in [a, b]$ such that $f(c) = c$.*

So let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a, b] \to [a, b]$ be a continuous function.

(a) Explain why we can assume that $a < f(a)$ and $f(b) < b$.

(b) Define $g : [a, b] \to \mathbb{R}$ by $g(x) = x - f(x)$.

i. Why is g a continuous function?

ii. What can we say about $g(a)$ and $g(b)$? Use the Intermediate Value Theorem to complete the proof.

Activity Solution.

(a) If $f(a) = a$ or $f(b) = b$, then f has a fixed point and we are done. So we can assume that $f(a) \neq a$ and $f(b) \neq b$. Since $f(a)$ and $f(b)$ are in $[a, b]$, it follows that $a < f(a)$ and $f(b) < b$.

(b) Define $g : [a, b] \to \mathbb{R}$ by $g(x) = x - f(x)$.

i. The fact that g is a difference of two continuous functions implies that g is a continuous function.

ii. Now $f(a) > a$ implies $g(a) = a - f(a) < 0$ and $f(b) < b$ implies that $g(b) = b - f(b) > 0$. The Intermediate Value Theorem then tells us that there is a number $c \in [a, b]$ such that $g(c) = 0$. But then

$$0 = g(c) = c - f(c)$$

and $f(c) = c$. Thus we demonstrated that f must have a fixed point.

Cut Sets

It can be difficult to determine if two topological spaces are homeomorphic. We can sometimes use topological invariants to determine if spaces are not homeomorphic. For example, if X is connected and Y is not, then X and Y are not homeomorphic. But just because two spaces are connected, it does not automatically follow that the spaces are homeomorphic. For example consider the spaces $(0, 2)$ and $[0, 2)$. Both are connected subsets of $\mathbb{R}$. If we remove a point, say 1, from the set $(0, 2)$ the resulting space $(0, 1) \cup (1, 2)$ is no longer connected. The same result is true if we remove any point from $(0, 2)$. However, if we remove the point 0 from $[0, 2)$ the resulting space $(0, 2)$ is connected. So the spaces $(0, 2)$ and $[0, 2)$ are fundamentally different in this respect, and so are not homeomorphic. Any set that we can remove from a connected set to obtain a disconnected set is called a *cut set*.

Definition 17.12. A subset S of a connected topological space X is a **cut set** of X if the set $X \setminus S$ is disconnected. A point p in a connected topological space X is a **cut point** if $X \setminus \{p\}$ is disconnected.

For example, 1 is a cut point of the space $(0, 2)$. Once we have a new property, we then ask if that property is a topological invariant.

Theorem 17.13. *Let X and Y be connected topological spaces and let $f : X \to Y$ be continuous. If $S \subset X$ is a cut set, then $f(S)$ is a cut set of Y.*

Proof. Let X and Y be topological spaces with $f : X \to Y$ a homeomorphism. Let S be a cut set of X. Let U and V form a separation of $X \setminus S$. We will demonstrate that $f(U)$ and $f(V)$ form a separation of $Y \setminus f(S)$, which will prove that $f(S)$ is a cut set of Y. Since f^{-1} is continuous, the sets $f(U)$ and $f(V)$ are open sets in Y. Next we prove that $X \setminus f(S) \subseteq f(U) \cup f(V)$. Let $y \in Y \setminus f(S)$. Since f is a surjection, there exists an $x \in X$ with $f(x) = y$. The fact that $y \notin f(S)$ means that $x \notin f(S)$. So $x \in X \setminus S \subseteq U \cup V$. If $x \in U$, then $f(x) = y \in f(U)$ and if $x \in V$, then $x = f(y) \in f(V)$. So $Y \setminus f(S) \subseteq f(U) \cup f(V)$.

Now we demonstrate that $f(U) \cap (Y \setminus f(S)) \neq \emptyset$ and $f(V) \cap (Y \setminus f(S)) \neq \emptyset$. Since U and V form a separation of $X \setminus S$, we know that $U \cap (X \setminus S) \neq \emptyset$ and $V \cap (X \setminus S) \neq \emptyset$. Let $x \in U \cap (X \setminus S)$. Then $x \in U$ and $x \notin S$. So $f(x) \in f(U)$ and the fact that f is an injection implies that $f(x) \notin f(S)$. Thus, $f(x) \in f(U) \cap (Y \setminus f(S))$. The same argument shows that $x \in V \cap (X \setminus S)$ implies that $f(x) \in f(V) \cap (Y \setminus f(S))$. So $f(U) \cap (Y \setminus f(S)) \neq \emptyset$ and $f(V) \cap (Y \setminus f(S)) \neq \emptyset$.

Finally, we show that $f(U) \cap f(V) \cap (Y \setminus f(S)) = \emptyset$. Suppose $y \in f(U) \cap f(V) \cap (Y \setminus f(S))$. Let $x \in X$ such that $f(x) = y$. Since f is an injection, we know that $f(x) \in f(U)$ means $x \in U$. so $x \in U \cap V$. The fact that $y \in Y \setminus f(S)$ means that $y \notin f(S)$. Thus, $x \notin S$. So $x \in X \setminus S$. We then have $x \in U \cap V \cap (X \setminus S) = \emptyset$. It follows that $f(U) \cap f(V) \cap (Y \setminus f(S)) = \emptyset$. Therefore, $f(U)$ and $f(V)$ form a separation of $Y \setminus f(S)$ and $f(S)$ is a cut set of Y.

■

Activity 17.9.

(a) Explain why being a cut set is a topological invariant.

(b) Use the idea of cut sets/points to explain why the unit circle in $\mathbb{R}^2$ is not homeomorphic to the interval $[0, 1]$ in $\mathbb{R}$. Note: the unit circle is the set $\{(x, y) \mid x^2 + y^2 = 1\}$. Draw pictures to illustrate your explanation. (A formal proof is not necessary, but you need to provide a convincing justification.)

(c) Consider the following subsets of $\mathbb{R}^2$ in the subspace topology:

$$A = \{(x, y) \mid x^2 + y^2 = 1\} \quad \text{and} \quad B = \{(x, 0) \mid -1 \leq x \leq 1\}.$$

Is $A \cup B$ homeomorphic to A? (A formal proof is not necessary, but you need to provide a convincing justification.)

Activity Solution.

(a) If $f : X \to Y$ is a homeomorphism, then our proof shows that S is a cut set in X if and only if $f(S)$ is a cut set in Y. Thus, being a cut set is a topological property. The fact that being a cut point is a topological property just follows from the fact that a cut point p is equivalent to a cut set $\{p\}$.

(b) Every interior point of the interval $[0, 1]$ is a cut point, but the unit circle has no cut points.

(c) The answer is no. Any set of two distinct points is a cut set of A, but $A \cup B$ has only one set of two points that is a cut set.

We have seen that topological equivalence is an equivalence relation, which partitions the collection of all topological spaces into disjoint homeomorphism classes. Topological invariants can then help us identify the classes to which different spaces belong. In general, though, it can be more difficult to prove that two spaces are homeomorphic that that are not.

Activity 17.10. Consider the spaces $S_1 = \mathbb{R}$, $S_2 = (0, 1)$ in $\mathbb{R}$, $S_3 = [-1, 1]$ in $\mathbb{R}$, the line segment S_4 in $\mathbb{R}^2$ between the points $(0, 0)$ and $(2, 2)$, the space S_5 determine by the letter X, and the space S_6 determined by the letter Y in $\mathbb{R}^2$. Identify the distinct homeomorphism classes determined by these six spaces. No formal proofs are necessary, but you need to give convincing arguments.

Activity Solution. We have already shown that $\mathbb{R}$ and $(0, 1)$ are homeomorphic spaces. Thus, $[\mathbb{R}] = [(0, 1)]$. Since S_3 and S_4 contain points that are not cut points, we conclude that neither space is in $[\mathbb{R}]$. Removing the center point in S_5 cuts S_5 into four connected components, and no point in $\mathbb{R}$ does the same. So $S_5 \notin [\mathbb{R}]$. Similarly, removing the middle point from S_6 cuts S_6 into three connected components. So $S_6 \notin [\mathbb{R}]$. Therefore, only S_1 and $S - 2$ belong to $[\mathbb{R}]$.

The function $f : S_3 \to S_4$ defined by $f(x) = (x + 1, x + 1)$ is linear in both components, so is a homeomorphism. Thus, $[S_3] = [S_4]$. For the same reason as above, neither S_5 nor S_6 is in $[S_3]$. Therefore, only S_3 and S_4 are in $[S_3]$.

Finally, there is no point in S_5 that cuts S_5 into three connected components, so $[S_5] \neq [S_6]$. We conclude that the distinct classes are $[S_1]$, $[S_3]$, $[S_5]$, and $[S_6]$.

Summary

Important ideas that we discussed in this section include the following.

- A subset A of a topological space (X, τ) is connected if A cannot be written as the union of two disjoint, nonempty, proper, relatively open subsets in the subspace topology. A topological space X is connected if X is a connected subset of X.

- A separation of a subset A of a topological space X is a pair of non-empty open subsets U and V of X such that

  - $A \subseteq (U \cup V)$,
  - $U \cap A \neq \emptyset$,
  - $V \cap A \neq \emptyset$, and
  - $U \cap V \cap A = \emptyset$.

  Showing that a set has a separation can be a convenient way to show that the set is disconnected.

- The connected subsets of $\mathbb{R}$ are the intervals and the single point sets.

- A subspace C of a topological space X is a connected component of X if C is connected and there is no larger connected subspace of X that contains C.

- One application of connectedness is the Intermediate Value Theorem that tells us that if A is a connected subset of a topological space X and if $f : A \to \mathbb{R}$ is a continuous function, then for any $a, b \in A$ and any $y \in \mathbb{R}$ between $f(a)$ and $f(b)$, there is a point $x \in A$ such that $f(x) = y$.

- A subset S of a connected topological space X is a cut set of X if the set $X \setminus S$ is disconnected, while a .point p in X is a cut point if $X \setminus \{p\}$ is disconnected. The property of being a cut set or a cut point is a topological invariant, so we can sometimes use cut sets and cut points to show that topological spaces are not homeomorphic.

Section 18

Path Connected Spaces

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is a path in a topological space?
- What is a path connected subset of a topological space?
- What is a path connected component of a topological space?
- What is a locally path connected space?
- What connections are there between connected spaces and path connected spaces?

Introduction

We defined connectedness in terms of separability by open sets. There are other ways to look at connectedness. For example, the subset $(0, 1)$ is connected in $\mathbb{R}$ because we can draw a line segment (which we will call a *path*) between any two points in $(0, 1)$ and remain in the set $(0, 1)$. So we might alternatively consider a topological space to be connected if there is always a path from one point in the spaced to another. Although this is a new notion of connectedness, we will see that path connectedness and connectedness are related.

Intuitively, a space is path connected if there is a path in the space between any two points in the space. To formalize this idea, we need to define what we mean by a path. Simply put, a path is a continuous curve between two points. We can therefore define a path as a continuous function.

Definition 18.1. Let X be a topological space. A **path** from point a to point b in X is a continuous function $p : [0, 1] \to X$ such that $p(0) = a$ and $p(1) = b$.

With the notion of path, we can now define path connectedness.

Definition 18.2. A subspace A of a topological space X is **path connected** if, given any $a, b \in A$ there is a path in A from a to b.

Preview Activity 18.1.

(1) Is $\mathbb{R}$ with the Euclidean metric topology path connected? Explain.

(2) Is $\mathbb{R}$ with the finite complement topology path connected? Explain.

(3) Let $A = \{b, c\}$ in (X, τ) with $X = \{a, b, c, d, e, f\}$ and

$$\tau = \{\emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}, X\}.$$

Is A connected? Is A path connected? Explain.

Activity Solution.

(1) The answer is yes. Let $a, b \in \mathbb{R}$, and define $p : [0, 1] \to \mathbb{R}$ by $p(x) = a + (b - a)x$. Since p is a linear function, we have shown in previous work that p is continuous. Note that $p(0) = a$ and $p(1) = b$. So $\mathbb{R}$ is path connected.

(2) Let $a, b \in \mathbb{R}$. Define $p : [0, 1] \to \mathbb{R}$ by $p(0) = a + (b - a)x$. Then $p(0) = a$ and $p(1) = b$. We consider the cases $a = b$ and $a \neq b$.

If $a = b$, then p is the constant function $p(x) = a$. Let O be an open set in $\mathbb{R}$. If $a \in O$, then $p^{-1}(O) = [0, 1]$. If $a \notin O$, then $p^{-1}(O) = \emptyset$. In any case, $p^{-1}(O)$ is open and so p is a continuous function.

Now assume $a \neq b$. Then p is an injection. To demonstrate that p is continuous, let O be an open set in $\mathbb{R}$. So $\mathbb{R} \setminus O$ is finite. This implies that $p^{-1}(\mathbb{R} \setminus O)$ is finite. But $p^{-1}(\mathbb{R} \setminus O) = [0, 1] \setminus p^{-1}(O)$. Any finite set is closed in $[0, 1]$, so $p^{-1}(O)$ is open and p is continuous. Therefore, $\mathbb{R}$ with the finite complement topology is path connected.

(3) The only open sets that contain b are $\{b, c, d, e, f\}$ and X. But these two sets do not form a separation of A, so A is connected. It is also true that A is path connected. The constant function $p : [0, 1] \to A$ defined by $p(x) = b$ is a path from b to b. Similarly, there is a constant path from c to c. So the only question is whether there is a path from b to c. The open sets in A are $\emptyset$, $\{c\}$, and A. Let $p : [0, 1] \to A$ be defined by $p(0) = b$, and $p(x) = c$ for all $x \neq 0$. Since $p^{-1}(\emptyset) = \emptyset$, $p^{-1}(\{c\}) = (0, 1]$ and $p^{-1}(A) = [0, 1]$, all of which are open in $[0, 1]$, we see that p is a continuous function. Therefore, A is also path connected.

Path Connectedness

As with every new property we define, it is natural to ask if path connectedness is a topological property.

Activity 18.1. In this activity we prove Theorem 18.3.

Theorem 18.3. *Let X and Y be topological spaces and let $f : X \to Y$ be a continuous function. If A is a path connected subspace of X, then $f(A)$ is a path connected subspace of Y.*

Assume that X and Y are topological spaces, $f : X \to Y$ is a continuous function, and $A \subseteq X$ is path connected. To prove that $f(A)$ is path connected, we choose two elements u and v in $f(A)$. It follows that there exist elements a and b in A such that $f(a) = u$ and $f(b) = v$.

(a) Explain why there is a continuous function $p : [0, 1] \to A$ such that $p(0) = a$ and $p(1) = b$.

(b) Determine how p and f can be used to define a path $q : [0, 1] \to f(A)$ from u to v. Be sure to explain why q is a path. Conclude that $f(A)$ is path connected.

Activity Solution.

(a) Since A is path connected, there is a path p in A from a to b. That is, $p : [0, 1] \to A$ is a continuous function such that $p(0) = a$ and $p(1) = b$.

(b) Let $q : [0, 1] \to f(A)$ be defined by $q = f \circ p$. As a composite of continuous functions we know that q is a continuous function. Also, $q(0) = (fp)(0) = f(a) = u$ and $q(1) = (fp)(1) = f(b) = v$. Thus, q is a path in $f(A)$ from u to v and $f(A)$ is path connected.

A consequence of Theorem 18.3 is the following.

Corollary 18.4. *Path connectedness is a topological property.*

path connected

Path Connectedness as an Equivalence Relation

We saw that we could define an equivalence relation using connected subsets of a topological space, which partitions the space into a disjoint union of connected components. We might expect to be able to do something similar with path connectedness. The main difficulty will be transitivity. As illustrated in Figure 18.1, if we have a path p from a to b and a path q from b to c, it appears that we can just follow the path p from a to b, then path q from b to c to have a path from a to c. But there are two problems to consider: how do we define this path as a function from $[0, 1]$ into our space, and how do we know the resulting function is continuous. The next lemma will help.

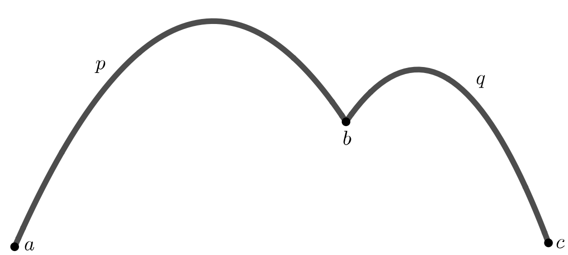


Figure 18.1: A path from a to c.

Lemma 18.5 (The Gluing Lemma)**.** *Let A and B be closed subsets of a space $X = A \cup B$, and let $f : A \to Y$ and $g : B \to Y$ be continuous functions into a space Y such that $f(x) = g(x)$ for all $x \in (A \cap B)$. Then the function $h : X \to Y$ defined by*

$$h(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B \end{cases}$$

is continuous.

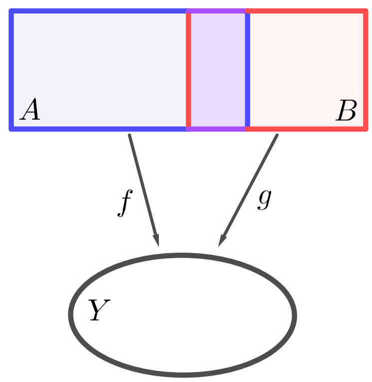


Figure 18.2: The Gluing Lemma.

Proof. Let A and B be closed subsets of a space $X = A \cup B$, and let $f : A \to Y$ and $g : B \to Y$ be continuous functions into a space Y such that $f(x) = g(x)$ for all $x \in (A \cap B)$ as illustrated in Figure 18.2. Define $h : X \to Y$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B. \end{cases}$$

To show that h is continuous, let C be a closed subset of Y. Then

$$h^{-1}(C) = \{x \in X \mid h(x) \in C\} = \{x \in A \mid f(x) \in C\} \cup \{x \in B \mid g(x) \in C\} = f^{-1}(C) \cup g^{-1}(C).$$

Since f is continuous, $f^{-1}(C)$ is closed in the subspace topology on A and since g is continuous $g^{-1}(C)$ is closed in the subspace topology on B. So $f^{-1}(C) = A \cap D$ and $g^{-1}(C) = B \cap E$ for some closed sets D and E of X. The fact that A is closed in X implies that $A \cap D$ is closed in X. Similarly, the fact that B is closed in X implies that $B \cap E$ is closed in X. Thus,

$$h^{-1}(C) = f^{-1}(C) \cup g^{-1}(C) = (A \cap D) \cup (B \cap E)$$

is a finite union of closed sets in X and so is closed in X. Since $h^{-1}(C)$ is closed for every closed set in Y, it follows that h is continuous. ■

We can use the Gluing Lemma to create a path from a to c given a path from a to b and a path from b to c.

Activity 18.2. Use the Gluing Lemma to explain why the path product given in the following definition is actually a path from $p(0)$ to $q(1)$.

Definition 18.6. Let p be a path from a to b and q a path from b to c in a space X. The **path product** $q * p$ is the path in X defined by

$$(q * p)(x) = \begin{cases} p(2x) & \text{for } 0 \leq x \leq \frac{1}{2} \\ q(2x-1) & \text{for } \frac{1}{2} \leq x \leq 1. \end{cases}$$

Activity Solution. The Gluing Lemma allows us to piece together the paths p and q. Let $A = \left[0\frac{1}{2}\right]$, $B = \left[\frac{1}{2}, 1\right]$, $f(x) = p(2x)$, and $g(x) = q(2x-1)$. Then $A \cap B = \left\{\frac{1}{2}\right\}$. Note that $f\left(\frac{1}{2}\right) = p(1) = b = q(0) = g\left(\frac{1}{2}\right)$. It follows that h as in the Gluing Lemma is a path from a to c.

Now we can show that path connectedness defines an equivalence relation on a topological space.

Activity 18.3. Let (X, τ) be a topological space. Define a relation on X as follows:

$$a \sim b \text{ if there is a path in } X \text{ from } a \text{ to } b. \tag{18.1}$$

(a) Explain why $\sim$ is a reflexive relation.

(b) Explain why $\sim$ is a symmetric relation.

(c) Explain why $\sim$ is a transitive relation.

Activity Solution.

(a) Let $a \in X$. The constant function $p : [0, 1] \to X$ defined by $p(x) = a$ for all $xinX$ is a path from a to a. So $a \sim a$ and $\sim$ is a reflexive relation.

(b) Let $b \in X$ such that $a \sim b$. Then there is a path $p : [0, 1] \to X$ from a to b. The function $q : [0, 1] \to X$ defined by $q(x) = p(1 - x)$ is a path from b to a. So $b \sim a$ and $\sim$ is symmetric.

(c) Assume $a \sim b$ and let $c \in X$ such that $b \sim c$. So there is a path p from a to b and a path q from b to c. The path product $q * p$ is a path from a to c, which makes $a \sim c$. We conclude that $\sim$ is transitive and an equivalence relation.

Since $\sim$ as defined in (18.1) is an equivalence relation, the relation partitions X into a union of disjoint equivalence classes. The equivalence class of an element $[a]$ is called a *path-component* of X, and is the largest path connected subset of X that contains a.

Definition 18.7. The **path-component** of an element a in a topological space (X, τ) is the largest path connected subset of X that contains a.

Path Connectedness and Connectedness

Path connectedness and connectedness are different concepts, but they are related. In this section we will show that any path connected space must also be connected. We will see later that the converse is not true except in finite topological spaces.

Theorem 18.8. *If a topological space X is path connected, then X is connected.*

Proof. Suppose that X is path connected. Let $a \in X$ and for any $x \in X$ let p_x be a path from a to x. Let $C_x = p_x([0,1])$. Now C_x is the continuous image of the connected set $[0,1]$ in $\mathbb{R}$, so C_x is connected. Also, $p_x(0) = a \in C_x$ and $p_x(1) = x \in C_x$. Thus, every set C_x contains a and so $\bigcap_{x \in X} C_x$ is not empty. Therefore,

$$C = \bigcup_{x \in X} C_x$$

is a connected subset of X. But every $x \in X$ is in a C_x, so $C = X$. We conclude that X is connected. ■

In the following sections we explore the reverse implication in Theorem 18.8 – that is, does connectedness imply path connectedness.

Path Connectedness and Connectedness in Finite Topological Spaces

In this section we will demonstrate that connectedness and path connectedness are equivalent concepts in finite topological spaces. In the following section, we prove that path connectedness and connectedness are not equivalent in infinite topological spaces. Throughout this section, we assume that X is a finite topological space. We begin with an example to motivate the main ideas.

Activity 18.4. Let $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, b, c\}, \{b, c, d\}, X\}$. Assume that τ is a topology on X.

(a) Is X connected? Explain.

(b) For each $x \in X$, let U_x be the intersection of all open sets that contain x (we call U_x a *minimal neighborhood* of x).

Definition 18.9. For $x \in X$, the **minimal neighborhood** U_x of x is the intersection of all open sets that contain x.

Find U_x for each $x \in X$.

(c) We will see that the minimal neighborhoods of X are path connected. Here we will illustrate with U_d.

i. Let $p : [0,1] \to X$ be defined by

$$p(t) = \begin{cases} b & \text{if } 0 \leq t < \frac{1}{2} \\ d & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Show that p is a path in U_d from b to d.

ii. Let $p : [0,1] \to X$ be defined by

$$p(t) = \begin{cases} c & \text{if } 0 \leq t < \frac{1}{2} \\ d & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Show that p is a path in U_d from c to d.

iii. Explain why U_d is path connected.

Activity Solution.

(a) The answer is yes. The only proper open set that contains d is $\{b, c, d\}$, and $\{a\}$ is not an open set, so there is no pair of disjoint, non-empty open subsets of X whose union is X.

(b) The singleton sets $\{b\}$ and $\{c\}$ are open, so $U_b = \{b\}$ and $U_c = \{c\}$. We take each other element in turn:

- $U_a = \{a, b\} \cap \{a, b, c\} \cap X = \{a, b\}$,
- $U_d = \{b, c, d\} \cap X = \{b, c, d\}$.

i. We check the inverse images of each open set in X:
   - $p^{-1}(\emptyset) = \emptyset$
   - $p^{-1}(\{b\}) = \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{c\})\emptyset$
   - $p^{-1}(\{a, b\}) = \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{b, c\} = \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{a, b, c\}) = \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{b, c, d\}) = [0, 1]$
   - $p^{-1}(X) = [0, 1]$.

   So $p^{-1}(O)$ is open for every open set O in X. Therefore, p is continuous.

ii. We check the inverse images of each open set in X:
   - $p^{-1}(\emptyset) = \emptyset$
   - $p^{-1}(\{b\}) = \emptyset$
   - $p^{-1}(\{c\}) \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{a, b\}) = \emptyset$
   - $p^{-1}(\{b, c\} = \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{a, b, c\}) = \left[0, \frac{1}{2}\right)$
   - $p^{-1}(\{b, c, d\}) = [0, 1]$
   - $p^{-1}(X) = [0, 1]$.

   So $p^{-1}(O)$ is open for every open set O in X. Therefore, p is continuous.

iii. The answer is yes. The constant function from $[0, 1]$ to U_d defined by $p(x) = d$ for all $x \in [0, 1]$ is a path from d to d. There is a path p_{bd} from b to d and a path p_{cd} from c to d. The function p_{db} defined by $p_{db}(t) = p_{bd}(1 - t)$ is a path from d to b. The composite, $p_{db}p_{cd}$ is then a path from c to b and there is a path between any two points in U_d. We conclude that U_d is path connected.

The terminology in Definition 18.9 is apt. Since every neighborhood N of a point $x \in X$ must contain an open set O with $x \in O$, it follows that $U_x \subseteq O \subseteq N$. So every neighborhood of $x \in X$ has U_x as a subset. In addition, when X is finite, the set U_x is a finite intersection of open sets, so the sets U_x are open sets (this is not true in general in infinite topological spaces – you should find an example where U_x is not open). In Activity 18.4 we saw that U_x was path connected for a particular x in one example. The next activity shows that this result is true in general in finite topological spaces.

Activity 18.5. Let X be a finite topological space, and let $x \in X$. In this activity we demonstrate that U_x is path connected. Let $y \in U_x$ and define $p : [0, 1] \to X$ by

$$p(t) = \begin{cases} y & \text{if } 0 \leq t < \frac{1}{2} \\ x & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

To prove that p is continuous, let O be an open set in X. We either have $x \in O$ or $x \notin O$.

(a) Suppose $x \in O$. Why must y also be in O? What, then, is $p^{-1}(O)$?

(b) Now suppose $x \notin O$. There are two cases to consider.

i. What is $p^{-1}(O)$ if $y \in O$?

ii. What is $p^{-1}(O)$ if $y \notin O$?

(c) Explain why p is a path from y to x.

(d) Show that we can find a path between any two points in U_x. Conclude that U_x is path connected.

Activity Solution.

(a) Every open set containing x contains U_x. Since $y \in U_x$, it follows that $y \in O$. Suppose $x \in O$. Why must y also be in O? Since $p(t) \in O$ for every t in $[0, 1]$, we conclude that $p^{-1}(O) = [0, 1]$.

(b) Now suppose $x \notin O$.

i. If $y \in O$, then $p^{-1}(O) = \left[0, \frac{1}{2}\right)$.

ii. If $y \notin O$, then $p^{-1}O = \emptyset$.

(c) So $p^{-1}(O)$ is open in $[0, 1]$ for every open set O in X. Therefore, p is continuous and is a path from y to x.

(d) We have seen that there is a path p_{yx} from y to x for any $y \in U_x$. Let a and b be points in U_x. There is a path p_{ax} from a to x and a path p_{bx} from b to x. Since points being path connected is a symmetric relation, there is a path p_{xb} from x to b. Then path product $p_{xb} * p_{ax}$ is then a path from a to b in U_x. Since path-connectedness is reflexive and symmetric, we conclude that there is a path between any two points in U_x, so U_x is a path connected set.

The sets U_x collectively form the space X, and each of the U_x is a path connected subspace. So every point in X is contained in some neighborhood with a path connected subset containing x. Spaces with this property are called *locally path connected*.

Definition 18.10. A topological space (X, τ) is **locally path connected at** x if every neighborhood of x contains a path connected neighborhood with x as an element. The space (X, τ) is **locally path connected** if X is locally path connected at every point.

If X is a finite topological space, for any $x \in X$ the set U_x is the smallest open set containing x. This means that any neighborhood of N of x will contain U_x as a subset. Thus, a finite topological space is locally path connected (this is not true in general of infinite topological spaces). One consequence of a locally path connected space is the following.

Lemma 18.11. *A space X is locally path connected if and only if for every open set O of X, each path component of O is open in X.*

Proof. Let X be a locally path connected topological space. We first show that for every open set O in X, every path component of O is open in X. Let O be an open set in X and let P be a path component of O. Let $p \in P$. Since X is locally path connected, the neighborhood O of x contains a path connected neighborhood Q of p. The fact that $p \in Q$ and P is a path component of O implies that $Q \subseteq P$. Thus, P contains a neighborhood of p and P is open.

Now we show that if for every open set O in X the path components of O are open in X, then X is locally path connected. Let $x \in X$ and let N be a neighborhood of x. Then N contains an open set U with $x \in U$. Let P be the path component in U that contains x. Now P is path connected and, by hypothesis, P is open in X and so is a path connected neighborhood of x. Thus, N contains a path connected neighborhood of x and X is locally path connected at every point.

■

Since X is open in X whenever X is a topological space, a natural corollary of Lemma 18.11 is the following.

Corollary 18.12. *Let X be a locally path connected topological space. Then every path component of X is open in X.*

Since there are only finitely many open sets in the finite space X, any arbitrary intersection of open sets in X just reduces to a finite intersection. So the intersection of any collection of open sets in X is again an open set in X. We will show that X is a union of path connected components, which will ultimately allow us to prove that if X is connected, then X is also path connected.

Activity 18.6. Let X be a locally path connected topological space. In this activity we will prove that the components and path-components of X are the same.

(a) Let $x \in X$, and let C be the component of X containing x and P be the path-component of X containing x. Show that $P \subseteq C$.

(b) To complete the proof that $P = C$, proceed by contradiction and assume that $C \neq P$. Let Q be the union of all path components of X that are different from P and that intersect C. Each such path-component is connected, and is therefore a subset of C. So $C = P \cup Q$. Explain why P and Q form a separation of C. (Hint: How do we use the fact that X is locally path connected?)

Activity Solution.

(a) Since P is path connected, Theorem 18.8 shows that P is connected. The fact that C is the largest connected subset of X containing x implies that $P \subseteq C$.

(b) Lemma 18.11 shows that P is an open set, and that Q is an open set. Since path components are either equal or disjoint, we have that $P \cap Q = \emptyset$. Therefore, if $C \neq P$, then P and Q form a separation of C. This contradicts the fact that C is connected. We conclude that $C = P$ and that the components and path components of X are the same.

We can now complete our main result of this section.

Theorem 18.13. *Let X be a finite topological space. Then X is connected if and only if X is path connected.*

Proof. Let X be a finite topological space. Theorem 18.8 demonstrates that if X is path connected, then X is connected. For the reverse implication, assume that X is path connected. Then X is composed of a single path-component, $P = X$. Since the path-components and components of X are the same, we conclude that $P = X$ is a component of X and that X is connected. ∎

Path Connectedness and Connectedness in Infinite Topological Spaces

Given that connectedness and path connectedness are equivalent in finite topological spaces, a reasonable question now is whether the converse of Theorem 18.8 is true in arbitrary topological spaces. As we will see, the answer is no. To find a counterexample, we need to look in an infinite topological space. There are many examples, but a standard example to consider is the *topologist's sine curve*. This curve S is defined as the union of the sets

$$S_1 = \{(0, y) \mid -1 \leq y \leq 1\} \text{ and } S_2 = \left\{\left(x, \sin\left(\frac{1}{x}\right)\right) \,\middle|\, 0 < x \leq 1\right\}.$$

A picture of S is shown in Figure 18.3.

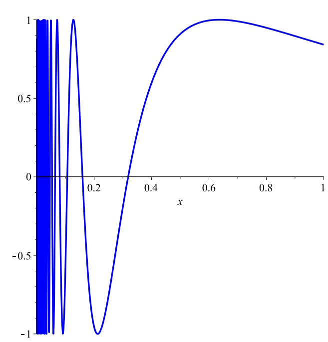


Figure 18.3: The topologist's sine curve.

To understand if S is connected, let us consider the relationship between S and S_2. Figure 18.3 seems to indicate that $S = \overline{S_2}$. To see if this is true, let $q = (0, y) \in S_1$, and let N be a neighborhood of q. Then there is an $\epsilon > 0$ such that $B = B(q, \epsilon) \subseteq N$. Choose $K \in \mathbb{Z}^+$ such that $\frac{1}{\arcsin(y)+2\pi K} < \epsilon$, and let $z = \frac{1}{\arcsin(y)+2\pi K}$. Then

$$\begin{aligned} d_E\left(q, \left(z, \sin\left(\frac{1}{z}\right)\right)\right) &= d_E\left((0,y),(z,\sin(\arcsin(y)+2\pi K)\right) \\ &= d_E((0,y),(z,\sin(\arcsin(y)))) \\ &= d_E((0,y),(z,y)) \\ &= |z| \\ &< \epsilon, \end{aligned}$$

and so $(z, \arcsin(z)) \in B(q, \epsilon)$ and every neighborhood of q contains a point in S_2. Therefore, $S_1 \subseteq S_2' \subseteq \overline{S_2}$ and $\overline{S_2} = S$ in S. The fact that S is connected follows from Theorem 17.8.

Now that we know that S is connected, the following theorem demonstrates that S is a connected space that is not path connected.

Theorem 18.14. *The topologist's sine curve is connected but not path connected.*

Proof. We know that S is connected, so it remains to show that S is not path connected. The sets S_1 and S_2 are connected (as continuous images of the interval $[0,1]$ and $(0,1]$, respectively). We will prove that there is no path p in S from $p(0) = (0,0)$ to $p(1) = b$ for any point $b \in S_2$ by contradiction. Assume the existence of such a path p. Let $U = p^{-1}(S_1)$ and $V = p^{-1}(S_2)$. Then

$$[0,1] = p^{-1}(S) = p^{-1}(S_1 \cup S_2) = p^{-1}(S_1) \cup p^{-1}(S_2) = U \cup V. \tag{18.2}$$

Note that S_2 is an open subset of S, since $S_2 = \left(\bigcup_{z=(x,y)\in S_2} B\left(z, \frac{x}{2}\right)\right) \cap S$. So the continuity of p implies that V is an open subset of $[0,1]$. Also, the fact that $p(0) \in S_1$ means that $U \neq \emptyset$, and the fact that $p(1) \in S_2$ means that $V \neq \emptyset$. If we demonstrate that U is an open subset of $[0,1]$, then Equation (18.2) will imply that $[0,1]$ is not connected, a contradiction. So we proceed to prove that U is open in $[0,1]$.

Let $x \in U$, and so $p(x)$ in S_1. The set $O = B_S\left(p(x), \frac{1}{2}\right) \cap S$ is open in S. The continuity of p then tells us that $p^{-1}(O)$ is open in $[0,1]$. So there is a $\delta > 0$ such that the open ball $B = B_{[0,1]}(x, \delta)$ is a subset of $p^{-1}(O)$. We will prove that $p(B) \subseteq S_1$. This will imply that $B \subseteq U$ and so U is a neighborhood of each of its points, and U is therefore an open set.

Every element in B is mapped into O by the path p. The set O is complicated, consisting of infinitely many sub-curves of the curve S_2, along with points in S_1, as illustrated in Figure 18.4. To simplify our analysis, let us consider the projection onto the x-axis. The function $P_x : \mathbb{R}^2 \to \mathbb{R}$ defined by $P_x(x,y) = x$ is a continuous function. Let $I = P_x(p(B))$. Since $p(B) \subseteq O$, we know that $I \subseteq P_x(O)$. Let $Z = P_x(O)$. So $I \subseteq Z$. Since B is a connected set (B is an interval), we know that $p(B)$ is a connected set. The fact that P_x is continuous means that $I = P_x(p(B))$ is connected as well. Now I is a bounded subset of $\mathbb{R}$, so I must be a bounded interval. Recall that $x \in B$ and so $p(x) \in p(B)$. The fact that $p(x) \in S_1$ tells us that $0 = P_x(p(x)) \in P_x(p(B)) = I$. So $I \neq \emptyset$. There are two possibilities for I: either $I = \{0\}$, or I is an interval of positive length. We consider the cases.

Suppose $I = \{0\}$. Then the projection of $p(B)$ onto the x-axis is the single point 0 and $p(B) \subseteq S_1$ as desired. Suppose that I is an interval of the form $[0,d]$ or $[0,d)$ for some positive number d. The structure of O would indicate that there must be some gaps in the set Z, the projection of

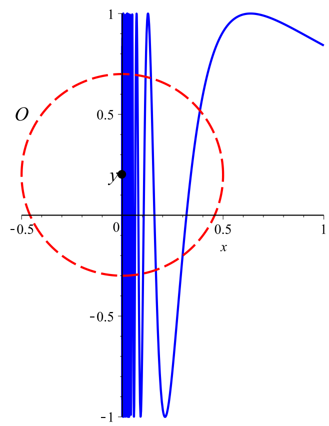


Figure 18.4: The set O.

O onto the x-axis. This implies that I cannot be a connected interval. We proceed to show this. In other words, we will prove that $I \setminus Z \neq \emptyset$ (which is impossible since $I \subseteq Z$). Remember that $p(x) \in S_1$, so let $p(x) = (0, q)$. We consider what happens if $q < \frac{1}{2}$ and when $q \geq \frac{1}{2}$.

Suppose $q < \frac{1}{2}$. Then the ball $B_S\left(p(x), \frac{1}{2}\right)$ contains only points with y value less than 1. Let $N \in \mathbb{Z}^+$ so that $t = \frac{1}{\pi/2 + 2N\pi} < d$. Then $t \in I$. But $\sin\left(\frac{1}{t}\right) = \sin(\pi/2 + 2N\pi) = \sin(\pi/2) = 1$, and so $\left(t, \sin\left(\frac{1}{t}\right)\right)$ is not in O. Thus, $t \notin Z$. Thus we have found a point in $I \setminus Z$.

Finally, suppose $q \geq \frac{1}{2}$. Then the ball $B_S\left(p(x), \frac{1}{2}\right)$ contains only points with y value greater than -1. Let $N \in \mathbb{Z}^+$ so that $t = \frac{1}{3\pi/2 + 2N\pi} < d$. Then $t \in I$. But $\sin\left(\frac{1}{t}\right) = \sin(3\pi/2 + 2N\pi) = \sin(3\pi/2) = -1$, and so $t \notin Z$. Thus we have found a point in $I \setminus Z$.

We conclude that there can be no path in S from $(0, 0)$ to any point in S_2, completing our proof that S is not path connected. (In fact, the argument given shows that there is no path in S from any point in S_1 to any point in S_2. ■

Summary

Important ideas that we discussed in this section include the following.

- A path in a topological space X is a continuous function p from the interval $[0, 1]$ to X. If $p(0) = a$ and $p(1) = b$, then p is a path from a to b.
- A subspace A of a topological space X is path connected if, given any $a, b \in A$ there is a path in A from a to b.
- The path-component of an element a in a topological space (X, τ) is the largest path connected subset of X that contains a.
- A topological space (X, τ) is locally path connected at x if every neighborhood of x contains a path connected subset with x as an element. The space (X, τ) is locally path connected if X is locally path connected at every point.

- Connectedness and path connectedness are equivalent in finite topological spaces, and path connectedness implies connectedness in general. However, there are topological spaces that are connected but not path connected. One example is the topologist's sine curve.

Index