An Activity-Based Introduction to Topology

Steven Schlicker

Grand Valley State University

Contents

Preface

Over many years I have taught topology I developed pre- and in-class activities that I used to supplement the texts I had adopted. Eventually I had enough material to eliminate the reliance on outside texts and could rely on the activities. This book is built on those activities. The emphasis for this book is to have students be active learners and to help develop their intuition through working activities and examples. Although it is difficult to capture the essence of active learning in a textbook, this book is an attempt to do just that.

The goals for these materials are several.

- To carefully introduce the ideas behind the definitions and theorems to help students develop intuition and understand the logic behind them.

- To help students understand that mathematics is not done as it is often presented. I expect students to experiment through examples, make conjectures, and then refine or prove their conjectures. I believe it is important for students to learn that definitions and theorems don't pop up completely formed in the minds of most mathematicians, but are the result of much thought and work.

- To help students develop their communication skills in mathematics. Students are expected to read and complete activities before class and come prepared with questions. Students regularly work in groups and present their work in class. Outside of class students work pre-class activities that are designed to help review previous material and to prepare them for the discussion of new material. In addition, students also individually write solutions to exercises on which they receive significant feedback. Communication skills are essential in any discipline and a heavy focus is placed on their development.

Layout

Each section of the book contains preview activities, in-class activities, and exercises. The various types of activities serve different purposes.

- Preview activities are designed for students to complete before class to motivate the upcoming topic and prepare them with the background and information they need for the class activities and discussion.

- The in-class activities engage students in common intellectual experiences. These activities are used to provide motivation for the material, opportunities for students to prove substantial course material on their own, or examples to help reinforce the meanings of definitions or theorems and their proofs. The ultimate goal is to help students develop their intuition for and understanding of abstract concepts. Students often complete these in-class activities, then present their results to the entire class.

Each section contains a collection of exercises. The exercises occur at a variety of levels of difficulty and most force students to extend their knowledge in different ways. While there are some standard, classic problems that are included in the exercises, many problems are open ended and expect a student to develop and then verify conjectures.

Organization

This text begins by formally introducing sets and functions. Although these topics are familiar to students, it is my experience that the level of understanding of sets and functions for most students is not yet sufficiently deep enough to ensure success with functions throughout the course. Chapter 2 focuses on metric spaces. It is my belief and my experience that students better understand the abstract concepts of neighborhoods, open sets, continuity, etc., if they are first experienced in a more familiar, concrete context like metric spaces. Metric spaces are easier for students to grasp than general topological spaces as they provide a notion of distance that is comforting to students. Metric spaces also allow one to introduce and motivate important topological concepts in a more familiar context. For example, by first encountering continuity of functions in a metric space setting, and revisiting the idea from different perspectives, the definition of continuity in the more abstract setting of topological spaces seems more accessible and natural. This perspective was also noted by Felix Hausdorff, who is considered as one of the founders of modern topology. His text *Grundzüge der Mengenlehre* (Fundamentals of Set Theory) (Felix Hausdorff, Leipzig, Von Veit, 1914. Translated to English as *Set Theory* by John R. Aumann et al, New York, Chelsea PublishingCompany, 1957) provided one of the first systematic treatments of topological spaces. As Hausdorff writes (p. 210) concerning the introduction of the concepts of topology following topics of basic set theory,

> "A quite generally worded theory of this nature would of course cause considerable complications, and deliver few positive results. But among the special examples that occupy a heightened interest belongs, apart from the theory of a [totally] ordered sets, especially the theory of point-sets in space, in fact here the foundational relationship is again a function of pairs of elements, namely the distance between two points: a function which however now is capable of infinitely many values."

To the Student

The objectives of this book and its inquiry-based format place the responsibility of learning the material where it belongs – on your shoulders. It is imperative that you engage in the material by completing the preview activities and the in-class activities in order to develop your intuition and understanding of the material. If you do this, and ask questions when you have them, your probability of success will be greatly enhanced. Good luck!

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.

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.

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.

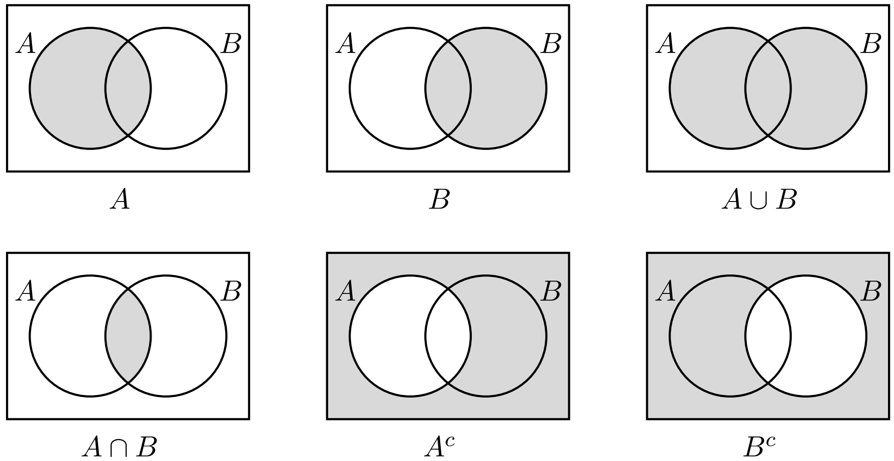

Figure 1.1: Vennn diagrams

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 $(A \cap B)^c$.

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

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 for the exercises.

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 $I = \mathbb{Z}^+$.

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.

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.

To conclude this section we summarize some properties of sets. Many of these properties can be extended to arbitrary collections of sets. Most of the proofs are straightforward. A few are left as exercises.

Theorem 1.3. *Let A, B, and C 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 $i \in I$.

Exercises

(1) Let $X \subset Y \subset Z$. Prove or disprove.

  (a) $C_Y(X) \subset C_Z(X)$.

  (b) $Z \setminus (Y \setminus X) = X \cup (Z \setminus Y)$.

(2) Let A and B be subsets of a universal set U. Prove the associative and distributive laws. That is, prove each of the following.

  (a) $(A \cap B) \cap C = A \cap (B \cap C)$

  (b) $(A \cup B) \cup C = A \cup (B \cup C)$

  (c) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

  (d) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(3) Prove DeMorgan's Laws. That is, let $\{A_\alpha\}$ is a collection of sets indexed by a set I in some universal set U. Prove that

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

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

(4) What familiar set is $\emptyset \times A$ for any set A? Explain.

(5) Let $X \subset Y \subset Z$. Prove the following.

  (a) $C_Y(X) \subset C_Z(X)$.

  (b) $Z \setminus (Y \setminus X) = X \cup (Z \setminus Y)$.

(6) If A is a set, the power set of A, denoted 2^A is the collection of all subsets of A.

  (a) List the elements of $2^{\{1,2\}}$.

  (b) If A is a set with three elements, how many elements are in 2^A?

  (c) If A is a set with n elements, make a conjecture about the number of elements in 2^A. Prove your conjecture?

(7) If A is a set, the power set of A, denoted 2^A is the collection of all subsets of A. Prove or disprove each of the following.

  (a) If A is a set, then $A \in 2^A$.

  (b) If A is a set, then $A \subset 2^A$.

  (c) If A is a set, then $\{A\} \subset 2^A$.

  (d) If A is a set, then $\emptyset \in 2^A$.

  (e) If A is a set, then $\emptyset \subset 2^A$.

  (f) If A and B are sets and $A \subseteq B$, then $2^A \subseteq 2^B$.

(8) Let A and B be sets, both of which have at least two distinct members. Prove that there is a subset $W \subset A \times B$ that is not the Cartesian product of a subset of A with a subset of B. [Thus, not every subset of a Cartesian product is the Cartesian product of a pair of subsets.]

(9) Let A, B, and C be subsets of a set X. Express each of the following sets in mathematical notation using the symbols $\cup$, $\cap$, and $\setminus$.

  (a) The elements of X that belong to A and B, but not C.

  (b) The elements of X that belong to C and either A or B.

  (c) The elements of X that belong to A but not to both B and C.

  (d) The elements of X that belong to none of the sets A, B, and C.

  (e) The elements of X that fail to belong to at least two of the sets A, B, and C.

  (f) The elements of X that fail to belong to at most one of the sets A, B, and C.

(10) Let I be the set of real numbers that are greater than 0. For each $x \in I$, let A_x be the open interval $(0, x)$. Prove that $\bigcap_{x\in I} A_x = \emptyset$, $\bigcup_{x\in I} A_x = I$. For each $x \in I$, let B_x be the closed interval $[0, x]$. Prove that $\bigcap_{x\in I} B_x = \{0\}$, $\bigcup_{x\in I} B_x = I \cup \{0\}$.

(11) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

  (a) If A, B, and C are sets and $A \subseteq B$ and $A \subseteq C$, then $A \subseteq (B \cap C)$.

  (b) If A, B, and C are sets and $A \subseteq C$ and $B \subseteq C$, then $(A \cup B) \subseteq C$.

  (c) If A and B are subsets of a set X and $A \subseteq B$, then $(X \setminus A) \subseteq (X \setminus B)$.

  (d) If A and B are subsets of a set X and $A \subseteq B$, then $(X \setminus B) \subseteq (X \setminus A)$.

(e) If A and B are sets, then $(A \cup B) \setminus B = A$.

(f) If A and B are sets, then $A \setminus (A \setminus B) = B$.

(g) If A, B, and C are sets, then $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$.

(h) If A and C are subsets of a set X, then $(A \setminus C) = A \cap (X \setminus C)$.

(i) There are no elements of the set $\{\emptyset\}$.

(j) There are two distinct objects that belong to the set $\{\emptyset, \{\emptyset\}\}$.

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.

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.

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.

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)?

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:

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

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.

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 exercises.

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.

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 as an exercise.

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.

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

Exercises

(1) Let A and B be sets, both of which have at least two distinct members. Prove that there is a subset $W \subset A \times B$ that is not the Cartesian product of a subset of A with a subset of B. [Thus, not every subset of a Cartesian product is the Cartesian product of a pair of subsets.]

(2) The cardinality of a finite set is defined to be the number of elements of that set. We denote the cardinality of a set A as $|A|$. Let A and B be sets with $|A| = n$ and $|B| = m$ for some positive integers m and n. Prove that there is a bijection $f : A \to B$ if and only if $n = m$.

(3) Let X and Y be sets and let $f : X \to Y$ be a function.

  (a) Let A be a subset of X. Show that $A \subseteq f^{-1}(f(A))$. Show that in general, $A \neq f^{-1}(f(A))$. (Hint: To show that the sets are not equal, consider sets X and Y with two elements.)

  (b) Let B be a subset of Y. Show that $f(f^{-1}(B)) \subseteq B$. Show that in general, $f(f^{-1}(B)) \subseteq B$. (Hint: To show that the sets are not equal, consider sets X and Y with two elements.)

  (c) Prove that f is a surjection if and only if $f(f^{-1}(B)) = B$ for every subset B of Y.

  (d) Prove that f is an injection if and only if $f^{-1}(f(A)) = A$ for every subset A of X.

(4) 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. Prove or disprove each of the following. If a statement is not true, is either containment true? Prove your answers.

  (a) $f\left(\bigcap_{\alpha \in I} A_\alpha\right) = \bigcap_{\alpha \in I} f(A_\alpha)$

(b) $f^{-1}\left(\bigcap_{\beta\in J} B_\beta\right) = \bigcap_{\beta\in J} f^{-1}(B_\beta)$

(5) Let X and Y be sets and let $f : X \to Y$ be a function.

(a) Let A be a subset of X. Show that $A \subseteq f^{-1}(f(A))$.

(b) Let B be a subset of Y. Show that $f(f^{-1}(B)) \subseteq B$.

(c) Prove that f is a surjection if and only if $f(f^{-1}(B)) = B$ for every subset B of Y.

(d) Prove that f is an injection if and only if $f^{-1}(f(A)) = A$ for every subset A of X.

(6) Let A and B be nonempty sets, and let $f : A \to B$ be a bijection. Prove that

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

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

(7) Let X and Y be sets and let $f : X \to Y$ be a function.

(a) Let A be a subset of X. Show that $A \subseteq f^{-1}(f(A))$. Show that in general, $A \neq f^{-1}(f(A))$. (Hint: To show that the sets are not equal, consider sets X and Y with two elements.)

(b) Let B be a subset of Y. Show that $f(f^{-1}(B)) \subseteq B$. Show that in general, $f(f^{-1}(B)) \neq B$. (Hint: To show that the sets are not equal, consider sets X and Y with two elements.)

(c) Prove that f is a surjection if and only if $f(f^{-1}(B)) = B$ for every subset B of Y.

(d) Prove that f is an injection if and only if $f^{-1}(f(A)) = A$ for every subset A of X.

(8) Prove that the inverse of a function f from a nonempty set A to a set B is a function if and only if f is a bijection.

(9) Let X_1 and X_2 be nonempty sets, and let $X = X_1 \times X_2$. Define $\pi_i : X \to X_i$ by $\pi_i(x) = x_i$, where $x = (x_1, x_2)$. We call π_i the *projection* of X onto X_i. Let Y_1 and Y_2 be nonempty sets, and let $Y = Y_1 \times Y_2$. Assume that for each i there is a function $f_i : X_i \to Y_i$. For example, let $X_i = \{i, i+1\}$ and $Y_i = \{-i, -i-1\}$. We could then define f_i by $f_i(x) = -x$ for i either 1 or 2.

(a) Prove that π_i is a surjection for each i.

(b) Prove that there is a unique function $f : X \to Y$ such that $\pi_i \circ f = f_i \circ \pi_i$ for each i. (Note that one of the π_i maps X to X_i and the other maps Y to Y_i.)

(c) The function f from part (b) is denoted as $f = f_1 \times f_2$. Let Z_1 and Z_2 be two nonempty sets, and let $Z = Z_1 \times Z_2$. Assume that there are functions $g_i : Y_i \to Z_i$ for each i. Show that

$$(g_1 \times g_2) \circ (f_1 \times f_2) = (g_1 \circ f_1) \times (g_2 \circ f_2).$$

(d) Suppose that each f_i has an inverse h_i. Show that $(f_1 \times f_2)^{-1} = h_1 \times h_1$.

(10) Prove that the inverse of a function f from a nonempty set A to a set B is a function if and only if f is a bijection.

(11) Let A be the set of all functions $f : [a, b] \to \mathbb{R}$ that are continuous on $[a, b]$ (use your memory of continuous functions from calculus for this problem). Let B be the subset of A consisting of all functions possessing a continuous derivative on $[a, b]$. Let C be the subset of B consisting of all functions whose value at a is 0.

(a) Let $d : B \to A$ be defined by

$$d(f) = f'.$$

Is the function d invertible? Justify your response.

(b) To each function $f \in A$, let $h(f)$ be the function defined by

$$(h(f))(x) = \int_a^x f(t)\,dt$$

for $x \in [a, b]$.

i. Verify that h maps A to C.

ii. Show that h is invertible by finding a function $g : C \to A$ such that g and h are inverse functions.

(12) For each of the following functions, determine if the function is an injection, a surjection, a bijection, or none of these. Justify all of your conclusions.

(a) $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = 5x + 3$, for all $x \in \mathbb{R}$

(b) $G : \mathbb{Z} \to \mathbb{Z}$ defined by $G(x) = 5x + 3$, for all $x \in \mathbb{Z}$

(c) $f : (\mathbb{R} \setminus \{4\}) \to \mathbb{R}$ defined by $f(x) = \dfrac{3x}{x-4}$, for all $x \in (\mathbb{R} \setminus \{4\})$

(d) $g : (\mathbb{R} \setminus \{4\}) \to (\mathbb{R} \setminus \{3\})$ defined by $g(x) = \dfrac{3x}{x-4}$, for all $x \in (\mathbb{R} \setminus \{4\})$

(e) $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ defined by $h(x) = x^2$ for every $x \in \mathbb{R}$, where $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$

(f) $k : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $k(x) = x^2$ for every $x \in \mathbb{R}_{\geq 0}$

(13) Let $\mathbb{N}$ be the set of positive integers. Define $f : \mathbb{N} \to \mathbb{Z}$ as follows: For each $n \in \mathbb{N}$, let

$$f(n) = \frac{1 + (-1)^n(2n-1)}{4}.$$

Is the function f an injection? Is the function f a surjection? Justify your conclusions. (Hint: Start by calculating several outputs for the function before you attempt to write a proof. In exploring whether or not the function is an injection, it might be a good idea to use cases based on whether the inputs are even or odd. In exploring whether f is a surjection, consider using cases based on whether the output is positive or less than or equal to zero.)

(14) An operation $*$ on a set S is a function from $S \times S$ to S that assigns to the pair $(x, y) \in S \times S$ the element $x * y$ in S. For example, addition of integers can be defined as a function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ that maps the pair $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ to the integer $f(a, b) = a + b$.

(a) Is the function f an injection? Justify your conclusion.

(b) Is the function f a surjection? Justify your conclusion.

(15) Let A, B, and C be sets and let $f : A \to B$ and $g : B \to C$ be functions.

(a) Is it true that if $g \circ f$ is an injection, then both f and g are injections? If the answer is no, are there any conditions that f or g must satisfy if $g \circ f$ an injection? Prove your answers.

(b) Is it true that if $g \circ f$ is a surjection, then both f and g are surjections? If the answer is no, are there any conditions that f or g must satisfy if $f \circ g$ a surjection? Prove your answers.

(16) (a) Is composition of functions a commutative operation? Prove your answer.

(b) Is composition of functions an associative operation? Prove your answer.

(17) (a) Define $f : \mathbb{Z}_5 \to \mathbb{Z}_5$ by $f([x]) = [x^2 + 4]$ for all $[x] \in \mathbb{Z}_5$. Write the inverse of f as a set of ordered pairs, and explain why f^{-1} is not a function.

(b) Define $g : \mathbb{Z}_5 \to \mathbb{Z}_5$ by $g([x]) = [x^3 + 4]$ for all $[x] \in \mathbb{Z}_5$. Write the inverse of g as a set of ordered pairs, and explain why g^{-1} is a function.

(c) Is it possible to write a formula for $g^{-1}([y])$, where $[y] \in \mathbb{Z}_5$? The answer to this question depends on whether or not it is possible to define a cube root of elements of $\mathbb{Z}_5$. Recall that for a real number x, we define the cube root of x to be the real number y such that $y^3 = x$. That is,

$$y = \sqrt[3]{x} \text{ if and only if } y^3 = x.$$

Using this idea, is it possible to define the cube root of each element of $\mathbb{Z}_5$? If so, what is $\sqrt[3]{[0]}$, $\sqrt[3]{[1]}$, $\sqrt[3]{[2]}$, $\sqrt[3]{[3]}$, and $\sqrt[3]{[4]}$.

(d) Now answer the question posed at the beginning of part (c). If possible, determine a formula for $g^{-1}([y])$ where $g^{-1} : \mathbb{Z}_5 \to \mathbb{Z}_5$.

(18) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If A is a subset of X, then $A \subseteq f^{-1}(f(A))$.

(b) If A is a subset of X, then $f^{-1}(f(A)) \subseteq A$.

(c) If B is a subset of Y, then $B \subseteq f(f^{-1}(B))$.

(d) If B is a subset of Y, then $f(f^{-1}(B)) \subseteq B$.

(e) If A_1 and A_2 are subsets of X with $A_1 \subseteq A_2$, then $f(A_1) \subseteq f(A_2)$.

(f) If B_1 and B_2 are subsets of Y with $B_1 \subseteq B_2$, then $f^{-1}(B_1) \subseteq f^{-1}(B_2)$.

(g) If B_1 and B_2 are subsets of Y with $B_1 \subseteq B_2$, then $f^{-1}(B_2) \subseteq f^{-1}(B_1)$.

(h) If A_1 and A_2 are subsets of X, then $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.

(i) If B_1 and B_2 are subsets of Y, then $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.

(j) If A_1 and A_2 are subsets of X, then $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$.

(k) If B_1 and B_2 are subsets of Y, then $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

(l) If A_1 and A_2 are subsets of X, then $f(A_1 \setminus A_2) = f(A_1) \setminus f(A_2)$.

(m) If B_1 and B_2 are subsets of Y, then $f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$.

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

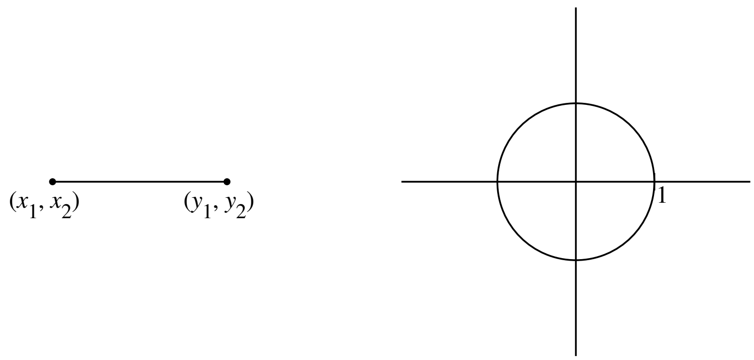

Figure 3.1: The Euclidean distance between (x_1, x_2) and (y_1, y_2) and the Euclidean unit circle in $\mathbb{R}^2$.

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.

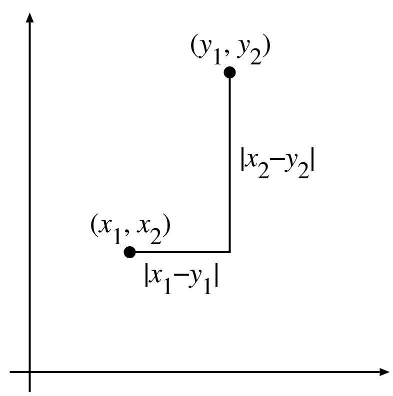

Figure 3.2: The taxicab distance between (x_1, x_2) and (y_1, y_2) in $\mathbb{R}^2$.

The taxicab metric can be extended to $\mathbb{R}^n$ for any $n \geq 1$ as follows. If $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are in $\mathbb{R}^n$, then the taxicab distance $d_T(x, y)$ from x to y is defined as

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

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*.

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 $[0, 1]$,

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

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 3.3. 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.

Just as with the Euclidean and taxicab metrics, item (c) in Activity 3.1 can be extended to $\mathbb{R}^n$ as follows. If $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are in $\mathbb{R}^n$, then the maximum distance $d_M(x, y)$ from x to y is defined as

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

The metric d_M is called the *max* metric. In the following section we prove that the Euclidean metric is in fact a metric. Proofs that d_T and d_M are metrics are left to the exercises.

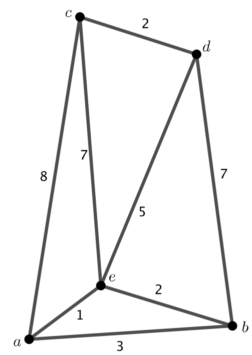

Figure 3.3: A graph to define a metric.

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 $x = y$.

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.

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

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

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

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

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

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

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

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

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.

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

$$\begin{aligned}
\sum (x_i + y_i)^2 &= \sum \left(x_i^2 + 2x_iy_i + y_i^2\right) \\
&= \sum x_i^2 + 2\sum x_iy_i + \sum y_i^2 \\
&\leq \sum x_i^2 + 2\left(\sqrt{\sum x_i^2}\right)\left(\sqrt{\sum y_i^2}\right) + \sum y_i^2 \\
&= \left(\sqrt{\sum x_i^2} + \sqrt{\sum y_i^2}\right)^2.
\end{aligned}$$

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

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

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 exercises. 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. The proof is left to an exercise.

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.

Exercises

(1) Let $X = \{1, 3, 5\}$ and define $d : X \times X \to \mathbb{R}$ by $d(x, y) = xy - 1 \pmod{n}$. That is, $d(x, y)$ is the remainder when $xy - 1$ is divided by n.

- For each value of n, determine if d defines a metric on X. Prove your answers.
- For any positive real number r, we define an open ball centered at b of radius r in a metric space (Y, d_Y) to be the set

$$B(b, r) = \{y \in Y \mid d_Y(y, b) < r\}.$$

If (X, d) is a metric space for a given value of n, determine all of the open balls in X centered at 1.

(a) $n = 4$

(b) $n = 8$

(2) Let X be a set. Show that the function d (the discrete metric) defined by

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

is a metric.

(3) Let Q be the set of all rational numbers in reduced form. A rational number $\frac{r}{s}$ is in reduced form if $s > 0$ and r and s have no common factors. Define $d : Q \times Q \to \mathbb{R}$ by

$$d\left(\frac{a}{b}, \frac{r}{s}\right) = \max\{|a - r|, |b - s|\}.$$

(a) Prove that d is a metric.

(b) A metric allows us to determine which elements in our metric space are "close" together. Describe the set of elements in Q that are a distance no more than 1 from $\frac{2}{3}$ using this metric d.

(4) Prove that the taxicab metric d_T is a metric on $\mathbb{R}^n$.

(5) Let A and B be nonempty finite subsets of $\mathbb{R}^n$, and let $A + B = \{a + b \mid a \in A, b \in B\}$.

(a) Prove that $\max(A + B) \leq \max A + \max B$.

(b) Prove that the max metric d_M is a metric on $\mathbb{R}^n$.

(6) If $x = (x_1, x_2, \ldots, x_n)$, we let $|x| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$. For $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots y_n)$, define $d_H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$d_H(x, y) = \begin{cases} 0 & \text{if } x = y \\ |x| + |y| & \text{otherwise.} \end{cases}$$

(a) Show that d_H is a metric (called the *hub* metric).

(b)

i. Let $a = \left(\frac{1}{2}, 0\right)$. Explicitly describe which points are in the set $B(a, 1)$ in $(\mathbb{R}^2, d_H)$. (See Exercise 1 for the definition of an open ball.)

ii. Let $a = (3, 4)$. Explicitly describe which points are in the set $B(a, 1)$ in $(\mathbb{R}^2, d_H)$.

iii. Now explicitly describe all open balls in $(\mathbb{R}^2, d_H)$.

(7) Let $\mathbb{Z}$ be the set of integers and let p be a prime. For each pair of distinct integers m and n there is an integer $t = t(m, n)$ such that $m - n = k \times p^t$, where p does not divide k. For example, if $p = 5$, $m = 34$, and $n = 7$, then $m - n = 27 = 27 \times 5^0$. So $t(43, 7) = 27$. However, if $m = 54$ and $n = 4$, then $m - n = 50 = 2 \times 5^2$. So $t(54, 4) = 2$.

Define $d : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ by

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

(a) Prove that if a, b, and c are in $\mathbb{Z}$, then

$$t(a, c) \geq \min\{t(a, b), t(b, c)\}.$$

(b) Prove that $(\mathbb{Z}, d)$ is a metric space.

(c) Let $p = 3$. Describe the set of all elements x in $(\mathbb{Z}, d)$ such that $d(x, 0) = 1$.

(d) Continue with $p = 3$. Describe the set of all elements x in $(\mathbb{Z}, d)$ such that $d(x, 0) < \frac{1}{2}$.

(8) Let (X, d) be a metric space, and let $a \in X$. Prove each of the following.

(a) There is a neighborhood that contains a.

(b) If N is a neighborhood of a and $N \subseteq N'$, then N' is a neighborhood of a.

(c) If M and N are neighborhoods of a, then so is $M \cap N$.

(9) Let (X, d) be a metric space and let x_1 and x_2 be distinct points in X. Prove that there are open sets O_1 containing x_1 and O_2 containing x_2 such that $O_1 \cap O_2 = \emptyset$. (This shows that we can separate points in metric spaces with open sets. Separation properties are important in topology.)

(10) Let (X, d_X) and (Y, d_Y) be metric spaces. We can make the Cartesian product $X \times Y$ into a metric space by defining a metric d' on $X \times Y$ as follows. If (x_1, y_2) and (x_2, y_2) are in $X \times Y$, then

$$d'((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}.$$

You may assume without proof that d' is a metric on $X \times Y$.

(a) Let $(X, d_X) = (\mathbb{R}^2, d_M)$ and $(Y, d_Y) = (\mathbb{R}^2, d_T)$. What is

$$d'(((1,2),(1,-1)),((0,5),(2,-2)))?$$

Recall that

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

and

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

(b) Let $(X, d_X) = (\mathbb{R}, d_E)$ and $(Y, d_Y) = (\mathbb{R}, d)$, where d is the discrete metric. Let

$$A = \{(x,y) \in \mathbb{R}\times\mathbb{R} \mid -1 \le x \le 1 \text{ and } -1 \le y \le 1\}$$

in $X \times Y$. Let $a = (0,1)$ in $X \times Y$. Describe, geometrically, what the open ball $B(a,1)$ looks like in the product space $X \times Y$. Draw a picture of this open ball. Is a an interior point of A? Explain.

(11) Let $X = \mathbb{R}^+$, the set of positive reals, and define $d : X \times X \to \mathbb{R}$ by

$$d(x,y) = |\ln(y/x)|.$$

Prove or disprove: d is a metric on X.

(12) Let $d : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be defined by

$$d(x,y) = \frac{|x-y|}{|x-y|+1}.$$

Show that d is a metric on $\mathbb{R}$. (Hint: For the triangle inequality, note that $d(x,y) = f(|x-y|)$ where $f(t) = \frac{t}{t+1}$. Note that f is an increasing function.)

(13) Let (X, d) be a metric space and k be a constant. Define $kd : X \times X \to \mathbb{R}$ by

$$(kd)(x,y) = kd(x,y).$$

Under what, if any, conditions is kd a metric on X. Justify your answer.

Assume $k > 0$. Let $x, y, z \in X$. Then

$$(kd)(x,y)+(kd)(y,z) = kd(x,y)+kd(y,z) = k(d(x,y)+d(y,z)) \ge k(d(x,z)) = (kd(x,z).$$

(14) A real valued function f on an interval is *concave* if

$$f((1-\alpha)x+\alpha y) \ge (1-\alpha)f(x)+\alpha f(y) \tag{3.3}$$

for all $\alpha \in [0,1]$ and all x and y in the interval with $x \ne y$. As Figure 3.4 indicates, (3.3) implies that the graph of a concave function f on any interval $[x,y]$ lies above the secant line joining the points $(x, f(x))$ and $(y, f(y))$.

(a) Show that if f is a concave function on $[0,\infty)$ and $f(0) \ge 0$, an interval and a and b are in the interval, then

$$f(a)+f(b) \ge f(a+b).$$

(Hint: Consider (3.3) with $y = 0$. Then use the fact that $\frac{a}{a+b}$ is in $[0,1]$.)

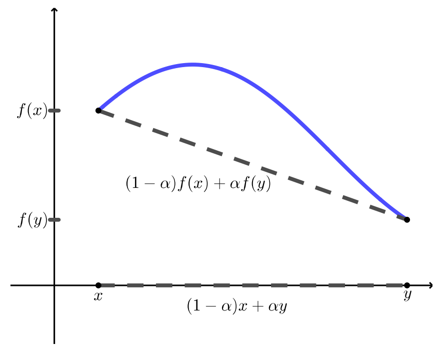

Figure 3.4: A concave function.

(b) Suppose (X, d) is a metric space and $f : [0, \infty) \to [0, \infty)$ is an increasing, concave function such that $f(x) = 0$ if and only if $x = 0$. Prove that $f \circ d$ is a metric on X.

(15) In our society, a great deal of information is communicated electronically. Bank transactions, television programs, military communications, cell phone calls, digital images, and almost any interchange one can think of either can be or is digitized and transmitted electronically. In many situations we need to compare one set of data to another (e.g., Internet searches for text strings or image matches, DNA strands), and metrics are often used for this purpose. Computers work in a binary system, that is they recognize only zeros and ones. So a digital text message is a string of zeros and ones. That is, a digital message is a collection of elements in the space X^n for some positive integer n, where $X = \{0, 1\}$. Each element in X^n is called a *word* - that is, a word is an element in X^n denoted in the form $(x_1, x_1, \ldots, x_n)$. Just like in the English language, where not every combination of letters corresponds to words that make sense, not every word is recognizable as part of an intelligible message. We might, for example, code the letters of the alphabet by assigning numbers 1-26 to the letters, then make them elements of X^n by converting to binary. The collection of all intelligible words is called a *code*. So a code is just some subset of X^n that all parties agree are sensible words. The words in a code are called *code words*. To deal with problems that occur in transmitting digital messages, like scrambling (*encoding*) messages, unscrambling (*decoding*) messages, and detecting and correcting errors in messages, it is useful to have a way to measure distance between words. One way is to use the Hamming metric.

Definition 3.6. Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be words in X^n. The **Hamming distance** d_H between x and y is

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

Recall that for each i both x_i and y_i are either 0 or 1. So

$$|x_i - y_i| = \begin{cases} 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i. \end{cases}$$

In other words, $d_H(x, y)$ counts the number of components at which x and y are different.

(a) Explain why d_H is a metric.

(b) Suppose we create a code

$$C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$$

in X^6, where

$$\begin{array}{lll} c_1 = (0,0,0,0,0,0), & c_2 = (0,0,0,0,1,1), & c_3 = (0,0,0,1,0,1), \\ c_4 = (0,0,1,0,0,1), & c_5 = (0,0,0,1,1,0), & c_6 = (0,0,1,0,1,0), \\ c_7 = (0,0,1,1,0,0), & c_8 = (0,0,1,1,1,1). & \end{array}$$

That is, the words c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8 are the only words that can comprise a message. Suppose we are on the receiving end of the message

$$(0,0,0,1,1,1)\ (0,0,1,1,0,0)\ (1,0,0,0,0,0)\ (0,0,0,0,1,1)\ (0,0,1,0,0,1).$$

i. Find $d_H(c_2, c_8)$.

ii. How do we know that an error has occurred in transmission of the message we received?

iii. To correct the errors in this received message, we replace the incorrect words with the code word(s) in C closest to them. Correct this message. (Note that there may be more than one possible substitution. Find all of the possibilities.)

(16) The Levenshtein metric is one measure of distance that researchers use to understand DNA. DNA is composed of double chains of nucleotides, which wind together to form a double helix. The nucleotides come four types: adenine (A), cytosine (C), guanine (G), and thymine (T). The nucleotides in the two chains of a DNA strand pair together, (A with T, and C with G), so the nucleotides in one chain determine the nucleotides in the other. Therefore, we can represent a DNA strand with a string of letters from the alphabet $\{A, C, G, T\}$. One problem DNA researchers have is how to compare two strands of DNA, and Levenshtein metric is one way that the distance between strands can be measured. Other metrics could be used, but the Levenshtein metric is appropriate to the task for several reasons. During evolution, changes in DNA sequences arise due to nucleotide substitution, or the insertion or deletion of nucleotides. These evolutionary changes are well modeled by the operations that determine the Levenshtein distance than other metrics. In addition, the Levenshtein metric can be used to calculate distances between strings of different lengths. The Levenshtein metric also has applications in spell checkers, speech recognition, and automated plagiarism detection. To understand how the Levenshtein metric is calculated, consider the question of how far apart the words “green” and “grease” are.

To compare these words, we have to be able to change letters, and add or delete letters. If $x = x_1x_2\cdots x_n$ is a string of letters, we allow the following operations:

a deletion: replace x with $x_1 \cdots x_{i-1}x_{i+1} \cdots x_n$ for some i,

an insertion: replace x with $x_1 \cdots x_i y x_{i+1} \cdots x_n$, where y is an allowable letter and $0 \leq i \leq n$,

a substitution: replace x with $x_1 \cdots x_{i-1} y x_{i+1} \cdots x_n$, where y is an allowable letter and $1 \leq i \leq n$.

(a) Using the allowable operations, change the word "green" into the word "grease". Specifically identify each operation you use. (Note: the intermediate strings of letters do not have to form recognizable words.) How many operations did you use?

(b) If it took three operations to transform "green" into "grease", we could say that the distance between "green" and "grease" is at most 3. However, it may be possible to transform "green" into "grease" in fewer than 3 operations, which might change our opinion of the distance between these words.

In general, to define the Levenshtein distance d_L between a sting x and a string y, let m_d denote the number of deletions, m_i the number of insertions, and m_s the number of substitutions we use to get from x to y. There may be many different combinations of m_d, m_i, and m_s that get us from x to y, so we want the smallest number.

Definition 3.7. The **Levenshtein distance** $d_L(x, y)$ between strings x and y is

$$d_L(x, y) = \min\{m_d + m_i + m_s\}.$$

Prove that the Levenshtein distance function is really a metric on the set of all possible words (sensical or nonsensical).

(c) A spell checker corrects the misspelled word "tupotagry". Using the Levenshtein metric, which word would the spell checker use as the closest to "tupotagry"? Why?

"topography" "topology" "tautology"

(17) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) The function $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $d(x, y) = (x - y)^2$ is a metric on $\mathbb{R}$.

(b) Every nonempty set can be made into a metric space.

(c) It is possible to define an infinite number of metrics on every set containing at least two elements.

(d) Let (X, d_X) and (Y, d_Y) be metric spaces with $|X| \geq 2$. Then the function $d : X \times Y \to \mathbb{R}$ defined by $d((a, b), (c, d)) = d_X(a, c)d_Y(b, d)$ is a metric on $X \times Y$.

(e) Let (X, d) be a metric space. If X is infinite, then the range of d is also an infinite set.

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 $s \in S$.

If a subset S of $\mathbb{R}$ has a lower bound, we say that S is *bounded below*. So the 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 $\mathbb{R}$?

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?

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 4.3. Let (X, d) be a metric space, let $x \in X$, and let A be a nonempty 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 nonempty subset of X

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

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

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.

Exercises

(1) Let S be a nonempty subset of $\mathbb{R}$ that is bounded below. Let $a \in \mathbb{R}$, and define $a + S$ to be $a + S = \{a + s \mid s \in S\}$.

(a) Explain why $a + \inf(S)$ is a lower bound for $a + S$. Explain why $a + S$ has an infimum.

(b) Let b be a lower bound for $a + S$. Show that $a + \inf(S) \geq b$. Then explain why $a + \inf(S) = \inf(a + S)$.

(2) Let S be a nonempty subset of $\mathbb{R}$.

(a) Assume that S is bounded above, and let $t = \sup(S)$. Show that for every $r < t$, there is a number $s \in S$ such that $r < s \leq t$.

(b) Assume that S is bounded below, and let $t = \inf(S)$. Show that for every $r > t$, there is a number $s \in S$ such that $t \leq s < r$.

(3) Let A and B be nonempty subsets of $\mathbb{R}$ that are bounded above and below. Let $A + B = \{a + b \mid a \in A, b \in B\}$.

(a) Follow the steps below to show that $\sup(A + B) = \sup(A) + \sup(B)$.

i. Let $x = \sup(A)$ and $y = \sup(B)$. Show that $x + y$ is an upper bound for $A + B$.

ii. The previous part shows that $A + B$ is bounded above and so has a supremum. Let $z = \sup(A + B)$. Explain why $z \leq x + y$.

iii. To show that $z = x+y$ we have to prove that z cannot be strictly less than $x+y$. Suppose to the contrary that $z < x + y$. Let $\epsilon = x + y - z$. Use the result of Exercise 2 to arrive at a contradiction.

(b) Prove that $\inf(A + B) = \inf(A) + \inf(B)$.

(c) Prove or disprove: $\sup(A \cup B) = \max\{\sup(A), \sup(B)\}$

(d) Prove or disprove: $\inf(A \cup B) = \min\{\inf(A), \inf(B)\}$

(4) Let $X = C[a, b]$, the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ on an interval $[a, b]$. Define $d : X \times X \to \mathbb{R}$ by

$$d(f, g) = \sup\{|f(x) - g(x)| \mid x \in [a, b]\}.$$

(5) In this exercise we prove the *Archimedean property* of the natural numbers. Note that the set of natural numbers, denoted $\mathbb{N}$ or $\mathbb{Z}^+$, is the set of all positive integers.

Theorem 4.4 (The Archimedean Property.)**.** *Given any real number x, there exists a natural number N such that $N > x$.*

Let x be a real number.

(a) Suppose that there is no positive integer N such that $N > x$. Explain how we can conclude that $\mathbb{Z}^+$ is bounded above.

(b) Assuming that $\mathbb{Z}^+$ is bounded above, explain why $\mathbb{Z}^+$ must have a least upper bound M.

(c) Explain why M cannot be a least upper bound for $\mathbb{Z}^+$. Explain why this proves the Archimedean property.

(6) In this exercise we prove two consequences of the Archimedean property (see Exercise 5). One of the consequences is the following theorem:

Theorem 4.5. *Given real numbers x and y with $x > 0$, there exists a natural number N such that $Nx > y$.*

(a) Let x and y be real numbers with $x > 0$.

i. Show that if the Archimedean property is true, then so is Theorem 4.5.

ii. Show that if Theorem 4.5 is true, then so is the Archimedean property. Conclude that Theorem 4.5 is equivalent to the Archimedean property.

(b) Show that if x is a positive number, then there exists a positive integer N such that $\frac{1}{N} < x$.

(7) We can use greatest lower bounds to prove the following theorem.

Theorem 4.6. *Given any two distinct real numbers x and y, there is a rational number that lies between them.*

This theorem tells us an important fact – that the rational numbers are what is called *dense* in the set of real numbers. We prove this theorem in this exercise. Let x and y be real numbers and assume $x < y$. By the Archimedean property of the natural numbers (see Exercises 5 and 6), there is a positive integer n such that $n(y - x) > 1$. Let $S = \{k \in \mathbb{Z} \mid k > nx\}$.

(a) Show that S is bounded below in $\mathbb{R}$.

(b) Explain why S contains an integer m such that if $q \in \mathbb{Z}$ with $q < m$, then $q \leq nx$. If may be helpful to use the Well-Ordering Principle that states

Every subset of the integers that is bounded below contains its infimum.

(The Well-Ordering Principle is one of many axioms that are equivalent to the Principle of Mathematical Induction. These principles are taken as axioms and are assumed to be true.)

(c) Explain why $m > nx$ and $m - 1 \leq nx$. Use these inequalities, along with $n(y - x) > 1$, to show that $nx < m < ny$. Then find a rational number that is strictly between x and y.

(8) Show that every open ball in $(\mathbb{R}^2, d_E)$ contains a point $x = (x_1, x_2)$ with both x_1 and x_2 rational.

(9) We are familiar with solving the quadratic equation $x^2 - 2 = 0$ to obtain the solutions $\pm\sqrt{2}$. But do we really know that the number $\sqrt{2}$ exists? We address that question in this exercise and demonstrate the existence of the number $\sqrt{2}$ using the greatest lower bound.

(a) To begin, let $S = \{x \in \mathbb{R}^+ \mid x^2 > 2\}$. Explain why S must have a greatest lower bound m.

(b) In what follows we demonstrate that $m^2 = 2$, which makes $m = \sqrt{2}$. We consider the cases $m^2 < 2$ and $m^2 > 2$.

i. Suppose $m^2 < 2$. Show that there is a positive integer n such that

$$\left(m + \frac{1}{n}\right)^2 < 2.$$

Explain why this also cannot happen.

ii. Suppose $m^2 > 2$. Show that there is a positive integer n such that

$$\left(m - \frac{1}{n}\right)^2 > 2.$$

Explain why this also cannot happen.

(c) Explain how we have demonstrated the existence of $\sqrt{2}$.

(10) Similar to Exercise 7 we can prove the following theorem.

Theorem 4.7. *Given any two distinct real numbers x and y, there is an irrational number that lies between them.*

(a) The first step is to demonstrate the existence of an irrational number. We will do that by proving that $\sqrt{2}$ is irrational. Proceed by contradiction and assume that $\sqrt{2}$ is a rational number. That is, $\sqrt{2} = \frac{r}{s}$ for some positive integers r and s such that r and s have no positive common factors other than 1.

i. Explain why $r^2 = 2s^2$. Since 2 is prime, it follows that 2 divides r.

ii. Show that 2 divides s. Explain how this proves that $\sqrt{2}$ is an irrational number.

(b) Let x and y be distinct real numbers. Show that there exists an integer q and a positive integer N such that $z = \frac{q\sqrt{2}}{2^N}$ is an irrational number between x and y. (Hint: Consider the approach in Exercise 7 .)

(11) Let (X, d) be a metric space and A a nonempty subset of X. For $x, y \in X$, prove that $d(x, A) \leq d(x, y) + d(y, A)$.

(12) Prove that if (X, d) is a metric space and B and C are nonempty subsets of X, then

$$d(a, B \cup C) = \min\{d(a, B), d(a, C)\}$$

for every $a \in X$.

(13) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses. Throughout, let S and T be bounded subsets of $\mathbb{R}$ (a subset of $\mathbb{R}$ is bounded if it is both bounded above and bounded below).

(a) Any nonempty subset of S is bounded.

(b) If $S + T = \{s + t \mid s \in S, t \in T\}$, then $\sup(S \cap T) = \max\{\sup(S), \sup(T)\}$.

(c) Let $S + T = \{s + t \mid s \in S, t \in T\}$, then $\inf(S \cap T) = \min\{\inf(S), \int(T)\}$.

(d) If U is a nonempty subset of S, then $\sup(U) \leq \sup(S)$.

(e) If U is a nonempty subset of S, then $\inf(S) \leq \inf(U)$.

(f) If A is a subset of $\mathbb{R}$ and $x \in \mathbb{R}$ with $d(x, A) = 0$, then $x \in A$.

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".

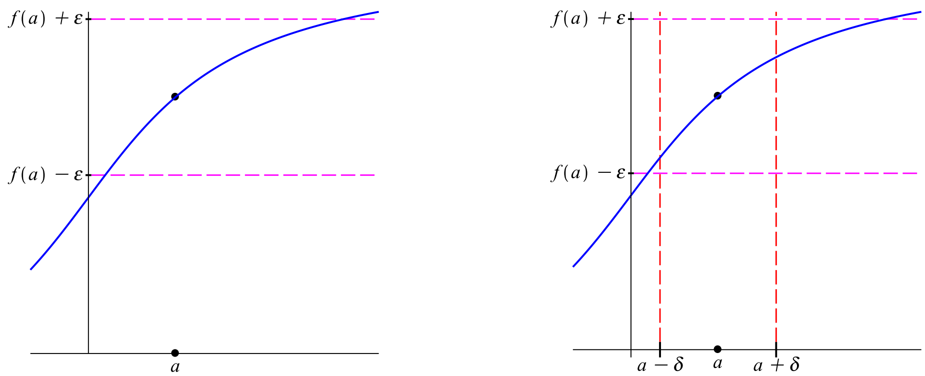

Figure 5.1: Demonstrating the definition of continuity at a point.

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 $x = 1$.

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 as an exercise. 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.)

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.

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.

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.

Exercises

(1) Let $(Y, d_Y) = (\mathbb{R}, d_E)$, where d_E is the Euclidean metric. On $\mathbb{R}$ we have $d_E(a, b) = |a - b|$.

  (a) Let $(X, d_X) = (\mathbb{R}^2, d_E)$. Prove or disprove: the function $f : X \to Y$ defined by $f((x_1, x_2)) = x_1 + x_2$ is continuous.

  (b) Let $(X, d_X) = (\mathbb{R}^2, d_M)$ where d_M is the max metric. Prove or disprove: the function $f : X \to Y$ defined by $f((x_1, x_2)) = x_1 + x_2$ is continuous.

(2) Let X be the set of continuous functions $f : [a, b] \to \mathbb{R}$. Let d^* be the distance function on X defined by

$$d^*(f, g) = \int_a^b |f(t) - g(t)| \, dt,$$

for $f, g \in X$. For each $f \in X$, set

$$I(f) = \int_a^b f(t)\,dt.$$

Prove that the function $I : (X, d^*) \to (\mathbb{R}, d)$ is continuous.

(3) Let X be any set and define $d : X \times X \to \mathbb{R}$ by

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

Exercise 2 in Section 3 asks us to show that d is a metric (the discrete metric) on $\mathbb{R}$.

Let (X, d_X) and (Y, d_Y) be metric spaces with d_X the discrete metric. Determine all of the continuous functions f from X to Y.

(4) Let f and g be continuous functions from $(\mathbb{R}, d_E)$ to $(\mathbb{R}, d_E)$.

(a) Let $k \in \mathbb{R}$ with $k \neq 0$ and define $kf : \mathbb{R} \to \mathbb{R}$ by $(kf)(x) = kf(x)$ for all $x \in \mathbb{R}$. Show that kf is a continuous function.

(b) Define $f + g : \mathbb{R} \to \mathbb{R}$ by $(f + g)(x) = f(x) + g(x)$ for all $x \in \mathbb{R}$. Show that $f + g$ is a continuous function.

(5) Let (X, d) be a metric space, and let A be a nonempty subset of X. Exercise 11 of Section 4 tells us that

$$d(b, A) \leq d(b, c) + d(c, A)$$

for all $b, c \in X$.

Define $f : X \to \mathbb{R}$ by $f(x) = d(x, A)$. Let $b \in X$. Given $\epsilon > 0$, show that there is a neighborhood N of b such that $x \in N$ implies $f(x) \in B(f(b), \epsilon)$. Conclude that f is a continuous function. (Assume the metric on $\mathbb{R}$ is the Euclidean metric.)

(6) Define $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as

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

Prove that d is a metric.

(7) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, with both copies of $\mathbb{R}$ having the Euclidean metric. Assume that $f(x) = 0$ whenever x is rational. Prove that $f(x) = 0$ for every $x \in \mathbb{R}$. (Hint: Use Exercise 7 in Section 4.)

(8) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 0$ if x is irrational and $f(x) = 1$ if x is rational. Assume the Euclidean metric on both copies of $\mathbb{R}$. Show that f is not continuous at any point in $\mathbb{R}$. (Hint: Use Exercises 7 and 10 in Section 4.)

(9) Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = 0$ if x is irrational and $g(x) = x$ if x is rational. Assume the Euclidean metric on both copies of $\mathbb{R}$. Show that g is continuous only at 0.

(10) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) Let $f : X \to Y$ be a function, where (X, d_X) and (Y, d_Y) are metric spaces. If d_X is the discrete metric and d_Y is any metric, then f is continuous.

(b) Let $f : X \to Y$ be a function, where (X, d_X) and (Y, d_Y) are metric spaces. If d_Y is the discrete metric and d_X is any metric, then f is continuous.

(c) Let d_1 and d_2 be two metrics on a set X. The identity function $i_X : (X, d_1) \to (X, d_2)$ defined by $i_X(x) = x$ for every $x \in X$ is continuous.

(d) Let f and g be continuous functions from $(\mathbb{R}^2, d_T)$ to $(\mathbb{R}, d_E)$. Then the function $f + g$ from $(\mathbb{R}^2, d_T)$ to $(\mathbb{R}, d_E)$ defined by $(f + g)(x) = f(x) + g(x)$ for every $x \in \mathbb{R}^2$ is a continuous function.

(e) If (X, d_X) and (Y, d_Y) are metric spaces with $y \in Y$, then the function $f : X \to Y$ defined by $f(x) = y$ for every $x \in X$ is a continuous function.

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.

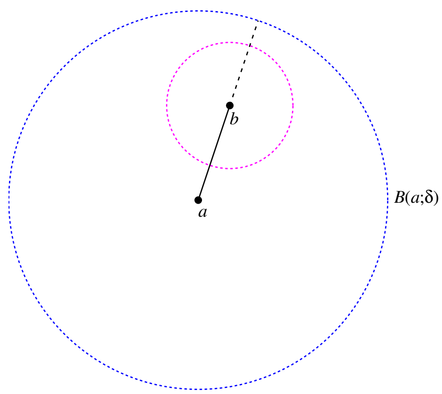

Figure 6.1: $B(a, \delta)$ as a neighborhood of b.

(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 $N \subseteq M$, then M 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 for the exercises.

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.

Exercises

(1) Determine, with proof, which of the following sets A is a neighborhood of a in the indicated metric space.

(a) $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ in $(\mathbb{R}^2, d_E)$ with $a = (0.5, 0.5)$

(b) A is the x-axis in $(\mathbb{R}^2, d_T)$ with $a = (0, 0)$, where d_T is the taxicab metric

(c) A is the set of rational numbers in $(\mathbb{R}, d_E)$ with $a = 0$

(d) A is the set of positive integers in (Q, d) and $a = 1$, where Q is the set of all rational numbers in reduced form with metric $d : Q \times Q \to \mathbb{R}$ defined by

$$d\left(\frac{a}{b}, \frac{c}{d}\right) = \max\{|a - c|, |b - d|\}$$

(The fact that d is a metric is the topic of Exercise 3 in Section 3.)

(2) Let $X = \{1, 3, 5\}$ and define $d_X : X \times X \to \mathbb{R}$ by $d_X(x, y) = xy - 1 \pmod 8$. That is, $d_X(x, y)$ is the remainder when $xy - 1$ is divided by 8. That d_X is a metric on X is examined in Exercise 1 of Section 3. Let (Y, d_Y) be a metric space. Is it possible to define a function $f : X \to Y$ that is not continuous? Explain.

(3) If $x = (x_1, x_2, \ldots, x_n)$, we let $|x| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$. For $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots y_n)$, define $d_H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$d_H(x, y) = \begin{cases} 0 & \text{if } x = y \\ |x| + |y| & \text{otherwise.} \end{cases}$$

The fact that d_H is a metric is examined in Exercise 6 in Section 3.

Let $(X, d_X) = (\mathbb{R}^2, d_H)$ and let $(Y, d_Y) = (\mathbb{R}, d_E)$. Define $f : X \to Y$ and $g : X \to Y$ by

$$f(x) = \begin{cases} 0 & \text{if } x = (0, 0) \\ 1 & \text{otherwise} \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 0 & \text{if } |x| < 1 \\ 1 & \text{otherwise} \end{cases}.$$

One of f, g is continuous and the other is not. Determine which is which, with proof.

(4) (a) Let $f : (\mathbb{R}, d_E) \to (\mathbb{R}, d_E)$ be defined by $f(x) = ax + b$ for some real numbers a and b with $a \neq 0$. Let $p \in \mathbb{R}$ and let $r > 0$. Show that $f^{-1}(B(f(p), r))$ contains an open ball centered at p. Conclude that every linear function from $(\mathbb{R}, d_E)$ to $(\mathbb{R}, d_E)$ is continuous. (Hint: By Exercise 4 in Section 5 we can assume $a > 0$ to simplify the problem.)

(b) Let $f : (\mathbb{R}, d_E) \to (\mathbb{R}, d_E)$ be defined by $f(x) = ax^2 + bx + c$ for some real numbers a, b, and c with $a \neq 0$. Let $p \in \mathbb{R}$ and let $r > 0$. $f^{-1}(B(f(p), r))$ contains an open ball centered at p. Conclude that every quadratic function from $(\mathbb{R}, d_E)$ to $(\mathbb{R}, d_E)$ is continuous. (Hint: Consider cases.)

(5) Let a and b be distinct points of a metric space X. Prove that there are neighborhoods N_a and N_b of a and b respectively such that $N_a \cap N_b = \emptyset$.

(6) Let (X, d) be a metric space and let $a \in X$. Prove each of the following.

(a) There is a neighborhood that contains a.

(b) If N is a neighborhood of a and $N \subseteq M$, then M is a neighborhood of a.

(c) If M and N are neighborhoods of a, then so is $M \cap N$.

(7) Let $f : (\mathbb{R}, d_E) \to (\mathbb{R}, d_E)$ be a continuous function. Show that if $f(a) > 0$ for some $a \in \mathbb{R}$, then there is a neighborhood N of a such that $f(x) > 0$ for all $x \in N$.

(8) Let (X, d) be a metric space where d is the discrete metric. Show that every subset of X is a neighborhood of each of its points.

(9) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If N is a neighborhood of a point a in a metric space X, then any open ball contained in N is also a neighborhood of a.

(b) If N is a neighborhood of a point a in a metric space X, then N is a neighborhood of each of its points.

(c) If X and Y are metric spaces and $f : X \to Y$ is a continuous function, then $f(N)$ is a neighborhood of $f(a)$ in Y whenever N is a neighborhood of a in X.

(d) If X and Y are metric spaces and $f : X \to Y$ is continuous at $a \in X$, and N is a neighborhood of $f(a)$ in Y, then $f^{-1}(N)$ is a neighborhood of a in X.

(e) If a is a point in a metric space X and if δ is a positive real number, then the open ball $B(a, \delta)$ contains infinitely many points of X.

(f) If N_1, N_2, ..., N_k are neighborhoods of a point a in a metric space X for some positive integer k, then $\bigcap_{i=1}^{k} N_i$ is a neighborhood of a.

(g) If N_α is a neighborhood of a point a in a metric space X for all α in some indexing set I, then $\bigcap_{\alpha \in I} N_\alpha$ is a neighborhood of a.

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.

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.

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.

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 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?

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 $(\mathbb{R}, d_E)$

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 $O \subseteq \text{Int}(A)$.

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 for the exercises.

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.

Exercises

(1) Prove that a subset O of a metric space X is open if and only if $O = \text{Int}(O)$.

(2) Let A and B be subsets of a metric space X with $A \subseteq B$. Prove or disprove the following.

(a) $\text{Int}(A) \subseteq \text{Int}(B)$

(b) $\text{Int}(\text{Int}(A)) = \text{Int}(A)$

(3) Let (X, d_X) and (Y, d_Y) be metric spaces and let $f : X \to Y$ be a function. Prove that f is continuous if and only if $f^{-1}(\text{Int}(B)) \subseteq \text{Int}(f^{-1}(B))$ for every subset B of Y.

(4) Let $a, b \in \mathbb{R}$ with $a < b$. Show that the set $(a, b]$ in $(\mathbb{R}, d_E)$ is not an open set.

(5) Let $A = \{(x, y) \in \mathbb{R}^2 \mid 1 < x < 3, 0 < y < 1\}$.

(a) Is A an open set in $(\mathbb{R}^2, d_E)$? Prove your answer.

(b) Is A an open set in $(\mathbb{R}^2, d_T)$? Prove your answer.

(c) Is A an open set in $(\mathbb{R}^2, d_M)$? Prove your answer.

(6) Let S be a finite set of points in $\mathbb{R}^2$. Is the set $\mathbb{R}^2 \setminus A$ an open set in $(\mathbb{R}^2, d_E)$? Prove your answer.

(7) Consider the metric space (Q, d), where $d : Q \times Q \to \mathbb{R}$ is defined by

$$d\left(\frac{a}{b}, \frac{u}{v}\right) = \max\{|a-u|, |b-v|\}$$

(The fact that d is a metric is the topic of Exercise 3 in Section 3.) Describe the open ball $B(q, 2)$ in Q if $q = \frac{2}{5}$.

(8) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If A and B are nonempty subsets of a metric space X, then $\text{Int}(A \cup B) \subseteq \text{Int}(A) \cup \text{Int}(B)$.

(b) If A and B are nonempty subsets of a metric space X, then $\text{Int}(A) \cup \text{Int}(B) \subseteq \text{Int}(A \cup B)$.

(c) If A and B are nonempty subsets of a metric space X, then $\text{Int}(A \cap B) \subseteq \text{Int}(A) \cap \text{Int}(B)$.

(d) If A and B are nonempty subsets of a metric space X, then $\text{Int}(A) \cap \text{Int}(B) \subseteq \text{Int}(A \cap B)$.

(e) Every subset of an open set in a metric space (X, d) is open in X.

(f) A subset O of $\mathbb{R}^2$ is open under the Euclidean metric d_E if and only if O is open under the taxicab metric d_T.

(g) Let X = $[0, 1] \cup [2, 3]$ endowed with the Euclidean metric. Then $[0, 1]$ is an open subset of X.

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 $\epsilon > 0$.

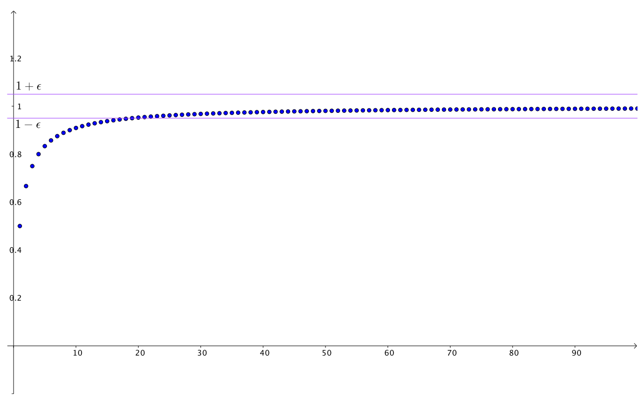

Figure 8.1: The limit of the sequence $\left(\frac{n}{1+n}\right)$.

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.

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?

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?

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.

Exercises

(1) Determine, with proof, the convergence or divergence of each of the following sequences in the indicated metric spaces.

(a) $a_n = 1 + \frac{1}{n}$ in $(\mathbb{R}, d_E)$

(b) $a_n = (2, n)$ in $(\mathbb{R}^2, d_M)$

(c) a_n is the function defined by
$$a_n(x) = \frac{1}{n}x$$
where X is the set of real valued functions on the interval $[0, 1]$ and the metric d is defined by
$$d(f, g) = \sup\{|f(x) - g(x)| \mid x \in [0, 1]\}.$$
(See Exercise 4 in Section 4.)

(2) Let A be a subset of $\mathbb{R}$.

(a) Show that if A is bounded above, then there is a sequence (a_n) in A such that $\lim a_n = \sup(A)$.

(b) Show that if A is bounded below, then there is a sequence (a_n) in A such that $\lim a_n = \inf(A)$.

(3) Let (X, d) be a metric space, let $x \in X$, and let A be a nonempty subset of X. Recall that the distance from x to A is
$$d(x, A) = \inf\{d(x, a) \mid a \in A.$$
In this exercise we see how we can view the distance between a point and a set in terms of sequences. Let $m = d(x, A)$. We will show that there must be a sequence (a_n) in A so that $d(x, A) = \lim d(x, a_n)$.

(a) For each $n \in \mathbb{Z}^+$, let $B_n = B\left(x; m + \frac{1}{n}\right)$. Why must $B_n \cap A \neq \emptyset$ for each $n \in \mathbb{Z}^+$?

(b) Let $a_n \in B_n \cap A$ for each n. What property does this sequence have? Explain how we have just proved the following theorem.

Theorem 8.7. *Let (X, d) be a metric space, let $x \in X$, and let A be a nonempty subset of X. Then there exists a sequence (a_n) in A such that*
$$\lim d(x, a_n) = d(x, A).$$

(4)

(a) Let (Y, d') be a subspace of (X, d). Let $a_1, a_2, \ldots$ be a sequence of points in Y and let $a \in Y$. Prove that if $\lim_n a_n = a$ in (Y, d'), then $\lim_n a_n = a$ in (X, d).

(b) Show that the converse of part (a) is false by considering the subspace $(\mathbb{Q}, d_{\mathbb{Q}})$ (the rational numbers) of (R, d). Let $a_1, a_2, \ldots$ be a sequence of rational numbers such that $\lim_n a_n = \sqrt{2}$. Prove that , given $\epsilon > 0$, there is a positive integer N such that for $n, m > N$, $|a_n - a_m| < \epsilon$. Does the sequence $a_1, a_2, \ldots$ converge when considered to be a sequence of points in $\mathbb{Q}$?

(5) In this exercise we prove some standard results about limits of sequences from calculus. Let (a_n) and (b_n) be convergent sequences in a metric space $(\mathbb{R}, d_E)$.

(a) Show that $\lim ka_n = k \lim a_n$ for any real number k.

(b) Show that $\lim(a_n + b_n) = \lim a_n + \lim b_n$.

(c) Show that the sequence (a_n) is bounded. That is, show that there is a positive real number M such that $|a_n| \leq M$ for all $n \in \mathbb{Z}^+$.

(d) Show that $\lim a_n b_n = \lim a_n \ \lim b_n$.

(e) If $b_n \neq 0$ for every n and $\lim b_n \neq 0$, show that $\lim \frac{a_n}{b_n} = \frac{\lim a_n}{\lim b_n}$.

(6) Let f and g be continuous functions from $\mathbb{R}$ to $\mathbb{R}$, both with the standard Euclidean metric. Define the function fg from $\mathbb{R}$ to $\mathbb{R}$ by

$$(fg)(x) = f(x)g(x) \text{ for every } x \in \mathbb{R}.$$

(a) Prove that fg is a continuous function.

(b) Assume that $g(x) \neq 0$ for every $x \in \mathbb{R}$. Define the function $\frac{f}{g}$ from $\mathbb{R}$ to $\mathbb{R}$ by $\frac{f}{g}(x) = \frac{f(x)}{g(x)}$ for every $x \in \mathbb{R}$. Use the definition of continuity to prove that $\frac{f}{g}$ is a continuous function.

(7) Let $(c_n) = (a_n, b_n)$ be a sequence in $(\mathbb{R}^2, d_E)$. Show that the sequence (c_n) converges to a point (a, b) if and only if (a_n) converges to a and (b_n) converges to b in $(\mathbb{R}, d_E)$.

(8) Define $f : (\mathbb{R}, d_E) \to (\mathbb{R}, d_E)$ by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ x & \text{if } x \text{ is rational.} \end{cases}$$

(a) Show that f is continuous at exactly one point. Assume that both copies of $\mathbb{R}$ are given the Euclidean topology.

(b) Modify the function f to construct a new function $g : \mathbb{R} \to \mathbb{R}$ such that g is continuous at exactly the numbers 0 and 1. Prove your result. Can you see how to extend this to construct a function $h : \mathbb{R} \to \mathbb{R}$ that is continuous at any given finite number of points?

(9) Let X be the set of real valued functions on the interval $[0,1]$ and let d be the metric on X defined by

$$d(f,g) = \sup\{|f(x) - g(x)| \mid x \in [0,1]\}.$$

(See Exercise 4 in Section 4.)

There is a difference between the point-wise convergence of a sequence of functions and convergence in the metric space (X,d) that we explore in this exercise. For each $n \in \mathbb{Z}^+$, define $f_n : [0,1] \to \mathbb{R}$ by $f_n(x) = x^n$.

(a) Let $0 \leq a < 1$. Show that the sequence (a_n) where $a_n = a^n$ converges to 0 in $(\mathbb{R}, d_E)$.

(b) Since the sequence (1) converges to 1, if we look at the behavior at each point, we might think that the sequence (f_n) converges to the function f defined by

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

Determine if the sequence (f_n) converges to (f) in the metric space (X,d).

(10) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If (a_n) is a sequence in $(\mathbb{R}, d_E)$ with $a_{n+1} < a_n$ for each $n \in \mathbb{Z}^+$ and the set $\{a_n\}$ is bounded below, then $\inf\{a_n \mid n \in \mathbb{Z}^+\}$ is the limit of the sequence (a_n).

(b) Let X be a metric space and A a nonempty subset of X. If $a \in X$ and if $B(a,r)$ in X contains a point of A for every $r > 0$, then there is a sequence in A that converges to a.

(c) Let R be a nonempty subset of $\mathbb{R}$ that is bounded above and below. If S is a nonempty subset of $\mathbb{R}$ and $x \leq y$ for all $x \in S$ and for all $y \in R$, then $\sup(S) \leq \inf(R)$.

(d) The sequence $\left(\frac{1}{n}\right)$ converges to 0 in the metric space Q of all rational numbers in reduced form with metric d defined by

$$d\left(\frac{a}{b}, \frac{r}{s}\right) = \max\{|a-r|, |b-s|\}.$$

(See Exercise 3 in Section 3.)

(e) The only convergent sequences in a metric space (X,d) with discrete metric d are the sequences that are eventually constant. (A sequence (a_n) in a metric space X is eventually constant if there is an element $a \in X$ and an $N \in \mathbb{Z}^+$ such that $a_n = a$ for all $n \geq N$.)

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 as an exercise.

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.

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.

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 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, ..., 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 $f^{-1}(Y \setminus B)$?

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?

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?

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.

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?

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 $\overline{A} \subseteq C$.

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** 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.*

The proof of Theorem 9.13 is left to the exercises.

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 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*

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

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.?

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.

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.

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.

Exercises

(1) Let (X, d) be a metric space. Let $a \in X$, and let $r > 0$. We know that the open ball $B(a, r) = \{x \in X \mid d(a, x) < r\}$ is an open set. Let

$$B[a, r] = \{x \in X \mid d(a, x) \leq r\}.$$

Prove or disprove: $B[a, r]$ is a closed set in X.

(2) Let (X, d) be a metric space. We have seen that it is possible for a subset of X to be both open and closed. There is a characterization of sets that are both open and closed in terms of their boundaries. Find and prove such a characterization. (Your statement should have the form: A subset A of a metric space X is both open and closed if and only if the boundary of A is _______.)

(3) Let A be a subset of a metric space. Let A' be the set of limit points of A and A^i the set of isolated points of A. Prove the following.

(a) $A \cup A^i = A \cup A'$

(b) $A' \cap A^i = \emptyset$

(c) $A \subseteq A' \cup A^i$

(d) $x \in \overline{A}$ if and only if there is a sequence of points of A which converges to x

(e) $\overline{A}$ is the intersection of all closed sets that contain A

(f) Int(A) is the union of all open sets contained in A

(g) $\overline{A}$ is the disjoint union of Int(A) and Bdry(A)

(h) $\overline{X \setminus A} = X \setminus \text{Int}(A)$

(i) $\text{Int}(X \setminus A) = X \setminus \overline{A}$

(4) Let (X, d) be a metric space and A a subset of X. Prove that a point $x \in X$ is a limit point of A if and only if every open ball centered at x contains a point in A different from x.

(5) Let A be a subset of a metric space. Let A' be the set of limit points of A and A^i the set of isolated points of A.

  (a) Prove that $A' \cap A^i = \emptyset$ and $A \subseteq A' \cup A^i$.

  (b) Prove that $x \in \overline{A}$ if and only if there is a sequence of points of A which converges to x.

  (c) Prove that if F is a closed set such that $A \subseteq F$, then $\overline{A} \subseteq F$. Then prove that $\overline{A}$ is the intersection of all such closed sets F and hence is closed.

(6) Recall that the distance from a point x in a metric space X to a nonempty subset A of X is

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

Prove that a subset C of a metric space X is closed if and only if whenever $x \in X$ and $d(x, C) = 0$, then $x \in C$.

(7) Let (X, d) be a metric space. In this exercise we show that some subsets of X, other than $\emptyset$ and X must be closed. Show that any finite subset of X is closed. (Hint: What are the limits points of a finite subset?)

(8) Prove that a subset C of a metric space X is closed if and only if C contains its boundary.

(9) Let (X, d_X) and (Y, d_Y) be metric space and let $f : X \to Y$ be a function.

  (a) Prove that f is continuous if and only if $f^{-1}(\text{Int}(B)) \subseteq \text{Int}(f^{-1}(B))$ for any subset B of Y.

  (b) Must the statement in part (a) be a containment instead of an equality? Give an example to verify that the containment is the best we can do, or provide a proof that we can actually use set equality instead of the containment.

(10) Let (X, d) be a metric space and let A be a subset of X. Prove that every boundary point of A is either a limit point or an isolated point of A.

(11) Let (X, d) be a metric space, and let A and B be subsets of X.

  (a) Is it the case that $\overline{A \cup B} = \overline{A} \cup \overline{B}$? If true, prove it. If false, show why and prove any containment that is true.

  (b) Is it the case that $\overline{A \cap B} = \overline{A} \cap \overline{B}$? If true, prove it. If false, show why and prove any containment that is true.

(12) Recall that an infinite union of closed sets in a metric space may not be closed, and that an infinite intersection of open sets in a metric space may not be open. In this exercise we explore situations in which we can conclude that an infinite union of closed sets is closed and an infinite intersection of open sets is open. Let (X, d) be a metric space.

  (a) We first establish a preliminary result. Let C be a closed subset of X and $x \in X$. Prove that if $x \notin C$, then $d(x, C) > 0$.

(b) Let $\{C_\alpha\}$ be a collection of closed subsets of X for α in some indexing set I with the property that given any $x \in X$, there exists an $\epsilon_x > 0$ such that $B(x, \epsilon_x)$ intersects at most finitely many of the sets C_α. Prove that $\bigcup_{\alpha \in I} C_\alpha$ is closed.

(c) Determine and prove an analogous statement for open sets in X.

(13) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If x is a point in a metric space X, then the singleton set $\{x\}$ is closed.

(b) The only subsets of $\mathbb{R}$ that are both open and closed under the standard metric are $\emptyset$ and $\mathbb{R}$.

(c) If (X, d) is the metric space with $X = \{1, 3, 5\}$ and $d(x, y) = xy - 1 \pmod 8$, then the set $\{1, 3\}$ is both open and closed in X.

(d) If X is a metric space and $A \subseteq X$, then $\text{Int}(\overline{A}) = A$.

(e) The boundary of any subset of a metric space X is a closed set.

(f) If A is a subset of a metric space X, then $A \subseteq A' \cup A^i$ where A' is the set of limit points of A and A^i is the set of isolated points of 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

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?

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 exercises.

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

$$d((x_1, x_2), (y_1, y_2)) = \sqrt{d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2}.$$

We show that d is a metric 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

$$\begin{aligned} d(x,z)^2 &= d_1(x_1,z_1)^2 + d_2(x_2,z_2)^2 \\ &\leq (d_1(x_1,y_1) + d_1(y_1,z_1))^2 + (d_2(x_2,y_2) + d_2(y_2,z_2))^2 \\ &= \left(d_1(x_1,y_1)^2 + d_2(x_2,y_2)^2\right) + 2\left(d_1(x_1,y_1)d_1(y_1,z_1) + d_2(x_2,y_2)d_2(y_2,z_2)\right) \\ &\qquad + \left(d_1(y_1,z_1)^2 + d_2(y_2,z_2)^2\right) \\ &= d(x,y)^2 + 2\left(d_1(x_1,y_1)d_1(y_1,z_1) + d_2(x_2,y_2)d_2(y_2,z_2)\right) + d(y,z)^2 \\ &\leq d(x,y)^2 + d(y,z)^2. \end{aligned}$$

Since all terms are non-negative we conclude that

$$d(x,z) \leq \sqrt{d(x,y)^2 + d(y,z)^2} \leq d(x,y) + d(y,z).$$

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 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 exercises.

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.

Exercises

(1) Determine if the following sets S are open in the subspace A of the topological space $(\mathbb{R}, d_E)$.

  (a) $S = [1, 2)$ in $A = [1, 3]$

  (b) $S = \{1, 2\}$ in $A = \mathbb{Q}$

  (c) $S = \{1, 2\}$ in $A = \mathbb{Z}$

(2) Let O be an open set in a metric space (X, d). Show that a subset U of O is open in $(O, d|_O)$ if and only if U is open in (X, d).

(3) Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a continuous function. If A is a subspace of X, must the restriction $f|_A$ of f to A mapping A to Y be continuous? Give a proof that the restriction is continuous, or an example to show that the restriction need not be continuous.

(4) Prove Theorem 10.3. That is, let (X, d) be a metric space and A a subset of X. Prove that 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 $C_A = C_X \cap A$.

(5) Let (X, d_X) and (Y, d_Y) be metric spaces. Prove or disprove: the function $d : X \times Y \to \mathbb{R}$ defined by

$$d((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2)$$

is a metric on $X \times Y$.

(6) Let (X_i, d_i) be metric spaces for i from 1 to some positive integer n. Let $d : \prod_{i=1}^n X_i \to \mathbb{R}$ be defined

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

when $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are in X. Show that d is a metric on $\prod_{i=1}^n X_i$.

(7) Let (x_n) be a non-decreasing sequence of real numbers that is bounded above. That is, $x_n \leq x_{n+1}$ for every n and there is a positive real number K such that $x_n \leq K$ for every n. Show that the sequence (x_n) converges.

(8) It is possible to consider infinite products as metric spaces. One important example is Hilbert space H which consists of all infinite sequences (x_n) where $x_n \in \mathbb{R}$ for every n and $\sum_{k=1}^{\infty} x_k^2$ is finite. Hilbert space has important applications in physics, particularly in quantum mechanics.

(a) Give two distinct elements in H and one infinite sequence that is not in H. Explain your examples.

(b) We define the norm of an element $x = (x_n)$ in H as

$$\|x\| = \sqrt{\sum_{k=1}^{\infty} x_k^2}.$$

From this norm we can define a distance between elements $x = (x_n)$ and $y = (y_n)$ in H as follows:

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

where $x - y = (x_n - y_n)$. Another way to write d is

$$d(x,y) = \sqrt{\sum_{k=1}^{\infty} (x_k - y_k)^2}.$$

One potential problem with this function d is that we need to know that if x and y are in H, then $x - y \in H$. That is, show that if $\sum_{k=1}^{\infty} x_k^2$ and $\sum_{k=1}^{\infty} y_k^2$ are finite, then $\sum_{k=1}^{\infty} (x_k - y_k)^2$ is also finite. (Hint: Consider a finite sum and use Exercise 7.)

(c) Show that d is a metric on H.

(d) Let $E^m = \{(x_n) \in H \mid x_k = 0 \text{ for } k > m\}$. Let $f : E^m \to \mathbb{R}^m$ be defined by $f((x_n)_{n=1}^{\infty}) = (x_n)_{n=1}^{m}$. Show that f is a bijection such that $d((x_n), (y_n)) = d_E(f((x_n)), f((y_n)))$ for any elements (x_n), (y_n) in H. So E^m is essentially the same as $\mathbb{R}^m$ and so we can consider the space $\mathbb{R}^m$ as embedded in H as a subspace of H for every $m \in \mathbb{Z}^+$.

(9) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If d is the discrete metric on a metric space X, then for any subspace A of X, the restriction of d to A is the discrete metric.

(b) If d is a metric on a space X that is not the discrete metric, and if A is a subset of X, then $d|_A$ cannot be the discrete metric.

(c) Let A be a subspace of a metric space (X, d). If a sequence (a_n) is in A and $\lim a_n = a$ for some $a \in A$, then $\lim a_n = a$ in X.

(d) Let A be a subspace of a metric space (X, d). If a sequence (a_n) is in A and $\lim a_n = a$ for some $a \in X$, then $\lim a_n = a$ in A.

(e) If (X, d_X) and (Y, d_Y) are metric spaces, then the function $d : X \times Y \to \mathbb{R}$ defined by

$$d((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}$$

is a metric on $X \times Y$.

(f) If (X, d_X) and (Y, d_Y) are metric spaces, then the function $d : X \times Y \to \mathbb{R}$ defined by

$$d((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) d_Y(y_1, y_2)$$

is a metric on $X \times Y$.

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 nonempty 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?

Examples of Topologies

In our preview activity we saw several examples of topologies. Suppose X is a nonempty 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*.

[1]The symbol τ is the Greek lowercase letter tau.

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* or the *cofinite topology*. Note that if X is finite, then τ_{FC} is just the discrete 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. As a non-trivial example, an interesting topology defined on the positive integers is due to S.W. Golomb. One can use this topology to prove that there are infinitely many primes. This topology also makes the positive integers into a connected Hausdorff space (more on these concepts later). The Golomb topology is defined as follows. If a and b are coprime integers in $\mathbb{Z}^+$ (that is, a and b have no common positive factors other than 1 (so the greatest common divisor of a and b is 1)), let

$$B_{a,b} = \{a + bn \mid n \geq 0\}.$$

The collection of sets $B_{a,b}$ is a basis for the Golomb topology, and the topological space $(\mathbb{Z}^+, \tau)$ is called the *Golomb space*. It is an exercise in number theory to prove that the sets $B_{a,b}$ for a basis for a topology, so we will not go into the details.

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 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, $\ldots$, 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, $\ldots$, 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 $B \subseteq \bigcap_{1 \leq k \leq n} B_k$.

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.

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.)

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.

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.

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 $\text{Int}(A)$? 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 $\text{Int}(A)$? 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 $\text{Int}(A)$? 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 $\text{Int}(A)$? What is 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 $\text{Int}(A)$ 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 $\text{Int}(A)$ is an open set in X, and that $\text{Int}(A)$ is the largest open subset of X contained in A. First we demonstrate that $\text{Int}(A)$ is an open set. Let $a \in \text{Int}(A)$. 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 $O \subseteq \text{Int}(A)$. Thus, $\text{Int}(A)$ is a neighborhood of each of its points and, consequently, $\text{Int}(A)$ is an open set.

The proof that $\text{Int}(A)$ 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 $\text{Int}(A)$ 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 $O \subseteq \text{Int}(A)$.

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 $O = \text{Int}(O)$.*

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.

Exercises

(1) For each integer a, let $a\mathbb{Z} = \{ka \mid k \in \mathbb{Z}\}$. That is, $a\mathbb{Z}$ is the set of all multiples of a. Show that $\{a\mathbb{Z} \mid a \in \mathbb{Z}\}$ is a basis for a topology on $\mathbb{Z}$.

(2) Let a and b be integers with $a \neq 0$. Let $a\mathbb{Z} + b = \{ak + b \mid k \in \mathbb{Z}\}$.

  (a) Show that $\{a\mathbb{Z} + b \mid a, b \in \mathbb{Z}, a \neq 0\}$ is a basis for a topology τ on $\mathbb{Z}$.

  (b) Let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by $f(n) = n + (-1)^n$. Show that f is a homeomorphism from $\mathbb{Z}$ to Z with the topology τ.

(3) Find all of the topologies on X if

  (a) X is a single point set

  (b) X is a two point set

(c) X is a three point set (Hint: there are 29 distinct topologies).

(4) For each $n \in \mathbb{Z}^+$, let $O_n = \{n, n+1, n+2, \ldots\}$. Let $\tau = \{\emptyset, O_1, O_2, O_3, \ldots\}$. Show that $(\mathbb{Z}^+, \tau)$ is a topological space.

(5) Let A, B be two subsets in a topological space X. What can you say about the relationships between $\text{Int}(A \cap B), \text{Int}(A \cup B)$ and $\text{Int}(A) \cap \text{Int}(B), \text{Int}(A) \cup \text{Int}(B)$, respectively? Verify your results.

(6) Let X be a nonempty set and let p be an element in X. Let τ_p be the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p. Show that τ_p is a topology on X. (This topology is called the *particular point topology*.

(7) Let X be a nonempty set and let p be an element in X. Let $\tau_{\overline{p}}$ be the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do not contain p. Show that $\tau_{\overline{p}}$ is a topology on X. (This topology is called the *excluded point topology*.

(8) Let n be a positive integer and let $\mathcal{P}_n$ be the collection of all polynomials in n real variables x_1, x_2, ..., x_n. As a specific example, the polynomial

$$f(x_1, x_2, x_3) = 2x_1x_3 + 5x_1x_2^2x_3^4 - x_2 + 10x_1^5x_3$$

is in $\mathcal{P}_3$. If $f(x_1, x_2, \ldots, x_n)$ is in $\mathcal{P}_n$, let $Z(f)$ be the set of zeros of the polynomial f. That is,

$$Z(f) = \{(x_1, x_2, \ldots, x_n) \mid f(x_1, x_2, \ldots, x_n) = 0\}.$$

Note that $Z(f)$ is a subset of $\mathbb{R}^n$. For example, if $n = 2$ and $f(x_1, x_2) = x_1^2 - x_2$ then $Z(f)$ is the set of ordered pairs in $\mathbb{R}^2$ satisfying $x_1^2 - x_2 = 0$, or $x_2 = x_1^2$. This is the graph of the parabola $y = x^2$ in the plane.

(a) Describe $Z(f)$ in $\mathbb{R}^2$ if $f(x_1, x_2) = x_1^2 - 1$.

(b) If E is a set of polynomials in $\mathcal{P}_n$, we let $Z(E) = \bigcap_{f \in E} Z(f)$ be the set of common zeros of all of the polynomials in E. Describe $Z(E)$ if $E = \{x_1 + x_2 + x_3, x_1 - x_2 - x_3, 3x_1 + x_2 + x_3\}$ in $\mathbb{R}^3$.

(c) Let E be a set of polynomials in $\mathcal{P}_n$. Prove or disprove $Z(E) = \bigcap_{f \in E} Z(f)$.

(d) Let $\mathcal{B}$ be the set of complements of the sets $Z(f)$ for $f \in \mathcal{P}_n$. Show that $\mathcal{B}$ is a basis for a topology on $\mathbb{R}^n$. The resulting topology is called the *Zariski* topology.

(e) Is the set $S = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0 \text{ or } x_2 = 0\}$ an open set in $\mathbb{R}^2$ with the Zariski topology? Explain.

(f) Explain why the Zariski topology when $n = 1$ is just the cofinite topology on $\mathbb{R}$. That is, show that every set that is open in the cofinite topology is open in the Zariski topology and that every set that is open in the Zaariski topology is open in the cofinite topology.

(9) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) The set $\{\emptyset, \{a,b\}, \{a,b,d,f\}, \{d,f\}, X\}$ is a topology on the set $X = \{a,b,c,d,e,f\}$.

(b) The set $\mathbb{Z}$ is an open subset of $\mathbb{R}$ using the finite complement topology τ_{FC} on $\mathbb{R}$.

(c) The set $\mathcal{B} = \{\{b\}, \{c\}, \{a,b\}, \{b,c,d\}\}$ is a basis for the topology τ on the set $X = \{a,b,c,d\}$, where

$$\tau = \{\emptyset, \{b\}, \{c\}, \{a,b\}, \{b,c\}, \{a,b,c\}, \{b,c,d\}, X\}.$$

(d) Let X be a nonempty set. If τ is the discrete topology, then the topological set (X,τ) is metrizable.

(e) The point b is an interior point of the subset $A = \{a,b,d\}$ in the topological space (X,τ), where $X = \{a,b,c,d\}$ and

$$\tau = \{\emptyset, \{a\}, \{a,b\}, \{c\}, \{d\}, \{c,d\}, \{a,c,d\}, \{a,c\}, \{a,d\}, \{a,b,c,\}, \{a,b,d\}, X\}.$$

(f) If τ_1 and τ_2 are topologies on a space X, then $\tau_1 \cup \tau_2$ is also a topology on X.

(g) If τ_1 and τ_2 are topologies on a space X, then $\tau_1 \cap \tau_2$ is also a topology on X.

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?
- What are the separation axioms T_1, T_2, T_3, and T_4. What is the underlying idea behind these properties?

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.

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. See Exercise 13 in Section 19 for additional information.)

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.

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 $\overline{A} \subseteq C$.

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}.

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.*

The proof of Theorem 12.11 is left to the exercises.

Separation Axioms

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 topology

$$\tau = \{\emptyset, \{c\}, \{a, c\}, \{b, c\}, X\}.$$

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.

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 $\{x_1, x_2, \ldots, x_n\}$.)

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. ■

In most cases we work in Hausdorff spaces, so Hausdorff spaces are important because we can separate distinct points with open sets. It is also of interest to consider what other types of objects we can separate with open sets. For example, the indiscrete topology is quite bad in the sense that its open sets can't distinguish between distinct points. That is, if x and y are distinct points in a space with the indiscrete topology, then every open set that contains x also contains y. By contrast, in a Hausdorff space we can separate distinct points with open sets. This is an example of what is called a "separation" property. Other types of separation properties describe different types of topological spaces. These separation properties determine what kind of objects we can separation with disjoint open sets – e.g., points, points and closed sets, closed sets and closed sets. The following are the most widely used separation properties. These properties rule out kinds of unwelcome properties that a topological space might have. For example, recall that limits of sequences are unique in Hausdorff spaces. (We traditionally call these separation properties “axioms” because we generally assume that our topological spaces have these properties. However, these are not axioms in the usual sense of the word, but rather properties.)

Definition 12.16. Let X be a topological space.

(1) The space X is a $\boldsymbol{T_1}$**-space** or **Frechet space** if for every $x \neq y$ in X, there exist an open set U containing y such that $x \notin U$.

(2) The space X is a $\boldsymbol{T_2}$**-space** or a **Hausdorff space** if for every $x \neq y$ in X, there exist disjoint open sets U and V such that $x \in U$ and $y \in V$.

(3) The space X is **regular** if for each closed set C of X and each point $x \in X \setminus C$, there exists disjoint open sets U and V in X such that $C \subseteq U$ and $x \in V$. The space X is a $\boldsymbol{T_3}$**-space** or a **regular Hausdorff space** if X is a regular T_1 space.

(4) The space X is a **normal** space if for each pair C and D of disjoint closed subsets of X there exist disjoint open sets U and V such that $C \subseteq U$ and $D \subseteq V$. The space X is a $\boldsymbol{T_4}$**-space** or a **normal Hausdorff space** if X is a normal T_1 space.

The use of the variable T comes from the German “Trennungsaxiome” for separation axioms. Note that these are not technically axioms, but rather properties. An interesting question is why we insist that T_3 and T_4-spaces also be T_1. We want these axioms to provide more separation at the index increases. Consider a space X with the indiscrete topology. In this space, nothing is separated. However, this space is vacuously regular and normal. To avoid this seeming incongruity, we insist on working only with T_1 spaces. Note that a space with the indiscrete topology is not T_1.

It is the case that every T_4-space is T_3, every T_3-space is T_2, and every T_2-space is T_1. Verification of these statements are left to the exercises. These properties are also all different. That is, there are T_1-spaces that are not T_2 and T_2-spaces that are not T_3. These problems are also in the exercises. The fact that there are T_3-spaces that are not T_4 is a bit difficult. An example is the *Niemytzki plane*. The Niemytzki plane is the upper half plane $X = \{(x, y) \in \mathbb{R}^2 \mid y \geq 0\}$. Let L be the boundary of X. That is, $L = \{(x, y) \in \mathbb{R}^2 \mid y = 0\}$. A basis for the topology on X consists of the standard open disks centered at points with $y > 0$ along with the open disks in $X \setminus L$ that are tangent to L together with their points of tangency. We won’t verify that the Niemytzki plane is T_3 but not T_4. The interested reader can find an accessible proof in the article “Another Proof that the Niemytzki Plane is not Normal” by David H. Vetterlein in the *Pi Mu Epsilon Journal*, Vol. 10, No. 2 (SPRING 1995), pp. 119-121.

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.
- Separation axioms tell us that kinds of objects can be separated by open sets.
  - In a T_1-space, we can separate two distinct points with one open set. That is, given distinct points x and y in a T_1 topological space X, there is an open set U that separates y from x in the sense that $y \in U$ but $x \notin U$.
  - In a T_2-space X we can separate points more distinctly. That is, if x and y are different points in X, there exist disjoint open sets U and V such that $x \in U$ and $y \in V$.
  - In a T_3-space X we can separate a point from a closed set that does not contain that point. That is, if C is a closed subset of X and x is a point not in C, there exists disjoint open sets U and V in X such that $C \subseteq U$ and $x \in V$.
  - In a T_4-space X we can separate disjoint closed sets. That is, if C and D are disjoint closed subsets of X, there exist disjoint open sets U and V such that $C \subseteq U$ and $D \subseteq V$.

Exercises

(1) Determine exactly which finite topological spaces are Hausdorff. Prove your result.

(2) Let (X, τ) be a topological space and let A be a subset of X. Prove that $\overline{A} = A \cup \text{Bdry}(A)$.

(3) Let A a subset of a topological space. Prove that $\text{Bdry}(A) = \emptyset$ if and only if A is open and closed.

(4) Let X be a nonempty set with at least two elements and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

Let $A = (0, 1]$ be a subset of $\mathbb{R}$. Find, with proof, $\overline{A}$, Int(A), and Bdry(A) when

(a) $\mathbb{R}$ has the topology τ_p with $p = 0$

(b) $\mathbb{R}$ has the topology $\tau_{\overline{p}}$ with $p = 0$.

(5) Let $\mathcal{B} = \{[a, b) \mid a < b \text{ in } \mathbb{R}\}$.

(a) Show that $\mathcal{B}$ is a basis for a topology $\tau_{\ell\ell}$ on $\mathbb{R}$. This topology is called the *lower limit* topology on $\mathbb{R}$. The line $\mathbb{R}$ with the topology $\tau_{\ell\ell}$ is sometimes called the *Sorgenfrey line* (after the mathematician Robert Sorgenfrey).

(b) Show that every open interval (a, b) is also an open set in the lower limit topology.

(c) If τ_1 and τ_2 are topologies on a set X such that $\tau_1 \subseteq \tau_2$, then τ_1 is said to be a *coarser* topology that τ_2, or τ_2 is a *finer* topology that τ_1. Part (b) shows that the lower limit topology may be a finer topology than the Euclidean metric topology. However, the two topologies could be the same – that is they could define the same open sets. Determine if the lower limit topology is actually a finer topology than the Euclidean metric topology on $\mathbb{R}$? Justify your answer.

(d) Let $a < b$ be in $\mathbb{R}$. Is the set $[a, b)$ clopen in $(\mathbb{R}, \tau_{\ell\ell})$? Prove your answer.

(6) A subset A of a topological space X is said to be *dense* in X if $\overline{A} = X$.

(a) Show that $\mathbb{Q}$ is dense in $\mathbb{R}$ using the Euclidean metric topology.

(b) Is $\mathbb{Z}$ dense in $\mathbb{R}$ using the Euclidean metric topology? Prove your answer.

(c) Let A be a subset of a topological space A. Prove that A is dense in X if and only if $A \cap O \neq \emptyset$ for every open set O.

(7) Let X be a topological space and let A be a subset of X.

(a) Show that the sets Int(A), Bdry(A), and Int(A^c) are mutually disjoint (that is, the intersection of any two of these sets is empty).

(b) Prove that $X =$ Int$(A) \cup$ Bdry$(A) \cup$ Int(A^c).

(8) Prove that a subspace of a Hausdorff space is a Hausdorff space.

(9) Let X be a nonempty set with at least two elements and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

Determine, with proof, if X is a Hausdorff space when

(a) X has the topology τ_p

(b) X has the topology $\tau_{\overline{p}}$.

(10) For each integer a, let $a\mathbb{Z} = \{ka \mid k \in \mathbb{Z}\}$. That is, $a\mathbb{Z}$ is the set of all integer multiples of a.

(a) Show that $\{a\mathbb{Z} \mid a \in \mathbb{Z}\}$ is a basis for a topology τ on $\mathbb{Z}$. (Hint: What set is $m\mathbb{Z} \cap n\mathbb{Z}$?)

(b) Work in the topological space $(\mathbb{Z}, \tau)$. Let $A = \mathbb{E}$, the set of even integers.

i. Find, with justification, Int(A).

ii. Find, with justification, $\overline{A}$.

(c) Recall that a point a in a subset A of a metric space X is an isolated point of A if there is a neighborhood N of a in X such that $N \cap A = \{a\}$. We can make the same definition in any topological space.

Definition 12.17. A point a in a subset A of a topological space X is an isolated point of A if there is a neighborhood N of a such that $N \cap A = \{a\}$.

(a) If A is a subset of a topological space X, prove that a point $a \in A$ is an isolated point of A if and only if $\{a\}$ is an open set in A.

(b) We proved that in a metric space every boundary point of a set A is either a limit point or an isolated point of A. (See Exercise 10 in Section 9.) Is the same statement true in a topological space? Prove your answer.

(c) Prove that a subset C of a topological space X is closed if and only if C contains its boundary.

(d) Work in the topological space $(\mathbb{Z}, \tau)$. Let $B = \mathbb{N} = \{n \in \mathbb{Z} \mid n \geq 1\}$. That is, $\mathbb{N}$ is the set of natural numbers.

i. Find, with justification, Int(B).

ii. Find, with justification, $\overline{B}$.

(11) Consider the Double Origin topology defined as follows. Let $X = \mathbb{R}^2 \cup \{0^*\}$, where 0^* is considered as a point that is not in $\mathbb{R}^2$ (0^* is our double origin). As a basis for the open sets, we use the standard open balls for every point except 0 and 0^*. For the point 0, we define open sets to be

$$N(0, r) = \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < \frac{1}{r^2}, y > 0\right\} \cup \{0\}$$

and for 0^* we define open sets to be

$$N(0^*, r) = \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < \frac{1}{r^2}, y < 0\right\} \cup \{0^*\}.$$

So $N(0, r)$ is the top half of a disk of radius $\frac{1}{r}$ centered at the origin, excluding the y-axis but including the origin, and $N(0^*, r)$ is the bottom half of a disk of radius $\frac{1}{r}$ centered at the origin, excluding the y-axis and including the point 0^*.

(a) Show that the collection of sets described as a basis for the Double Origin topology is actually a basis for a topology.

(b) Is X with the Double Origin topology Hausdorff? Prove your answer.

(12) (a) Show that finite sets are closed in $\mathbb{R}^n$ with the Zariski topology.

(b) Show that $\mathbb{R}^n$ with the Zariski topology is not Hausdorff. (Exercise 8 in Section 11 shows that a basis for the Zariski topology on $\mathbb{R}^n$ is the collection of sets of the form $\mathbb{R}^n \setminus Z(f)$, where $Z(f)$ is the set of zeros of the polynomial f in n variables.)

(13) Consider the digital line topology τ_{dl} on $\mathbb{Z}$ with basis $\{B(n)\}$, where

$$B(n) = \begin{cases} \{n\} & \text{if } n \text{ is an odd integer,} \\ \{n-1, n, n+1\} & \text{if } n \text{ is an even integer.} \end{cases}$$

.

(a) Let $A = \{-1, 0, 1\}$ of $(\mathbb{Z}, \tau_{dl})$.

i. Find the limit points and boundary points of A. Prove your conjectures. Is every limit point of A a boundary point of A? Is every boundary point of A a limit point of A?

ii. Find $\overline{A}$ and write $X \setminus \overline{A}$ as a union of open sets.

(b) Now consider the subset $B = \{0\}$ of $(\mathbb{Z}, \tau_{dl})$.

i. Find the limit points and boundary points of B. Prove your conjectures. Is every limit point of B a boundary point of B? Is every boundary point of B a limit point of B?

ii. Find $\overline{B}$ and write $X \setminus \overline{B}$ as a union of open sets.

(14)

(a) Prove that a topological space X is T_1 if and only if each singleton set is closed.

(b) Show that every T_2-space is T_1, that every T_3-space is T_2, and that every T_4-space is T_3.

(15) In this exercise we illustrate spaces that are T_1 but not T_2 and T_2 but not T_3.

(a) Show that $\mathbb{R}$ with the finite complement topology is T_1 but not T_2.

(b) Define the space $\mathbb{R}_K$ to be the set of reals with topology τ with a basis that consists of the standard open intervals in $\mathbb{R}$ along with all sets of the form $(a, b) \setminus K$, where (a, b) is any open interval and $K = \left\{\frac{1}{k} \mid k \in \mathbb{Z}^+\right\}$. Show that $\mathbb{R}_K$ is T_2 but not T_3.

(16) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) Every limit point of a subset A of a topological space X is also a boundary point of A.

(b) Every boundary point of a subset A of a topological space X is also a limit point of A.

(c) If X is a topological space and $A \subseteq X$ such that $\text{Int}(A) = \overline{A}$, then A is both open and closed.

(d) If X is a topological space and A and B are subsets of X with $\overline{A} = \overline{B}$ and $\text{Int}(A) = \text{Int}(B)$, then $A = B$.

(e) If A and B are subsets of a topological space X, then $\overline{A \cap B} = \overline{A} \cap \overline{B}$.

(f) If A and B are subsets of a topological space X, then $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

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

$$(X, \tau_X) = (\{1, 2, 3, 4\}, \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4\}, X\})$$

and let

$$(Y, \tau_Y) = (\{a, b, c\}, \{\emptyset, \{a\}, \{a, c\}, Y\}).$$

Define $f : X \to Y$ by $f(1) = a$, $f(2) = c$, $f(3) = f(4) = b$.

(a) Show that f is a continuous function.

(b) Even though f is continuous, it is possible that $f(O)$ may not be open for every open set in X. Find such an example for this function f.

Functions f that have the property that $f(O)$ is open whenever O is open in X are called *open* functions.

Definition 13.3. Let $f : X \to Y$ be a function from a topological space X to a topological space Y. Then f is an *open* function if $f(U)$ is open in Y whenever U is open in X.

There is a similar definition of a *closed* function.

(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.

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.

If there is a bijection between metric spaces that preserves distances, we say that the metric spaces are *metrically equivalent*.

Definition 13.4. Two metric spaces (X, d_X) and (Y, d_Y) are **metrically equivalent** if there is a bijection $f : X \to Y$ such that

$$\begin{aligned} d_X(x, y) &= d_Y(f(x), f(y)) \\ d_Y(u, v) &= d_X(f^{-1}(u), f^{-1}(v)) \end{aligned}$$

for all $x, y \in X$ and $u, v \in Y$.

Any function f that preserves distances (like the one in Definition 13.4) is called an *isometry*.

Definition 13.5. 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*.

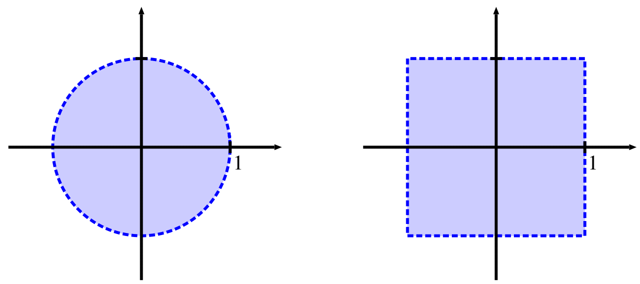

Figure 13.1: The open unit balls in $(\mathbb{R}^2, d_E)$ and $(\mathbb{R}^2, d_M)$.

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 13.6. 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 13.6 is called a *homeomorphism*.

Definition 13.7. 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 13.8. *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

$$d'(x, y) \leq Kd(x, y)$$
$$d(x, y) \leq K'd'(x, y)$$

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

$$d'(i_X(x), i_X(a)) = d'(x, a) \leq Kd(x, a) < K\delta = K\left(\frac{\epsilon}{K}\right) = \epsilon.$$

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.

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 13.9. 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 13.10. 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.

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 13.11. 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

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.

Exercises

(1) Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be a homeomorphism. Let A be a subset of X.

(a) If x is a limit point of A, must $f(x)$ be a limit point of $f(A)$? Prove your answer.

(b) If x is an interior point of A, must $f(x)$ be an interior point of $f(A)$? Prove your answer.

(c) If x is a boundary point of A, must $f(x)$ be a boundary point of $f(A)$? Prove your answer.

(2) Let (X, τ_X), (Y, τ_Y), and (Z, τ_Z) be topological spaces.

(a) Let $f : X \to Y$ and $g : Y \to Z$ be continuous functions. Prove that $g \circ f : X \to Z$ is a continuous function.

(b) Let $f : (X, \tau_X) \to (Y, \tau_Y)$ be a homeomorphism. Let h be a function from (Y, τ_Y) to (Z, τ_Z) and let k be a function from (Z, τ_Z) to (X, τ_X) .

i. Prove that h is continuous if and only if $h \circ f$ is continuous.

ii. Prove that k is continuous if and only if $f \circ k$ is continuous.

(3) Let $X = \{a, b, c, d\}$ with topology $\tau = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, d\}, \{a, b, d\}, X\}$.

(a) Find a function $f : X \to X$ that is continuous at exactly one point, or show that no such function exists.

(b) Find a function $f : X \to X$ that is continuous at exactly two points, or show that no such function exists.

(c) Find a function $f : X \to X$ that is continuous at exactly three points, or show that no such function exists.

(4) Consider $\mathbb{R}$ and $\mathbb{R}^2$ equipped with the Euclidean topology. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let

$$\Gamma_f = \{(x, f(x)) \mid x \in \mathbb{R}\}$$

be the graph of f. Note that Γ_f is a subspace of $\mathbb{R}^2$ and is a topological space using the subspace topology.

(a) Show that if f is a continuous function, then Γ_f is homeomorphic to $\mathbb{R}$.

(b) If we remove the condition that f is continuous, must it still be the case that Γ_f is homeomorphic to $\mathbb{R}$? Prove your conjecture.

(5) Let X be a nonempty set and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

(a) Let p be a fixed point in $\mathbb{R}$. Is the identity function $i : \mathbb{R} \to \mathbb{R}$ defined by $i(x) = x$ for all $x \in \mathbb{R}$ a homeomorphism from $(\mathbb{R}, \tau_p)$ to $(\mathbb{R}, \tau_e)$? Prove your answer.

(b) Is $(\mathbb{R}, \tau_p)$ homeomorphic to $(\mathbb{R}, \tau_e)$ with the specific point $p = 0$? Prove your answer.

(6) A topological space X is *embedded* in a topological space Y if there is a homeomorphism from X to some subspace of Y. The homeomorphism is called an *embedding*.

(a) Show that if X is the open interval $(0, 1)$ with the Euclidean metric topology, then X can be embedded in the topological space $\mathbb{R}$ with the Euclidean metric topology.

(b) Show that there exist non-homeomorphic topological spaces A and B for which A can be embedded in B and B can be embedded in A.

(7) Let $X = \{a, b, c\}$. There are 29 distinct topologies on X, shown below. Determine the number of distinct homeomorphism classes for these 29 topologies and identify the elements of each homeomorphism class. Justify your answers.

1. $\{\emptyset, X\}$
2. $\{\emptyset, \{a, b\}, X\}$
3. $\{\emptyset, \{a, c\}, X\}$
4. $\{\emptyset, \{b, c\}, X\}$
5. $\{\emptyset, \{a\}, X\}$
6. $\{\emptyset, \{b\}, X\}$
7. $\{\emptyset, \{c\}, X\}$
8. $\{\emptyset, \{a\}, \{a, b\}, X\}$
9. $\{\emptyset, \{a\}, \{a, c\}, X\}$
10. $\{\emptyset, \{a\}, \{b, c\}, X\}$
11. $\{\emptyset, \{b\}, \{a, b\}, X\}$
12. $\{\emptyset, \{b\}, \{a, c\}, X\}$
13. $\{\emptyset, \{b\}, \{b, c\}, X\}$
14. $\{\emptyset, \{c\}, \{a, b\}, X\}$
15. $\{\emptyset, \{c\}, \{a, c\}, X\}$
16. $\{\emptyset, \{c\}, \{b, c\}, X\}$
17. $\{\emptyset, \{a\}, \{a, b\}, \{a, c\}, X\}$
18. $\{\emptyset, \{b\}, \{a, b\}, \{b, c\}, X\}$
19. $\{\emptyset, \{c\}, \{a, c\}, \{b, c\}, X\}$
20. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$
21. $\{\emptyset, \{a\}, \{c\}, \{a, c\}, X\}$
22. $\{\emptyset, \{b\}, \{c\}, \{b, c\}, X\}$
23. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, X\}$
24. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c\}, X\}$
25. $\{\emptyset, \{a\}, \{c\}, \{a, c\}, \{a, b\}, X\}$
26. $\{\emptyset, \{a\}, \{c\}, \{a, c\}, \{b, c\}, X\}$
27. $\{\emptyset, \{b\}, \{c\}, \{b, c\}, \{a, b\}, X\}$
28. $\{\emptyset, \{b\}, \{c\}, \{b, c\}, \{a, c\}, X\}$
29. the discrete topology

(8) Show that property T_i is a topological property for each i. (See Section 12 for definitions of the separation axioms.)

(9) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If $f : X \to Y$ is a continuous function between topological spaces X and Y, then for every open subset U of X, $f(U)$ is open in Y.

(b) If τ_{FC} is the finite complement topology, then $f(x) = x^2$ mapping $(\mathbb{R}, \tau_{FC})$ to $(\mathbb{R}, \tau_{FC})$ is continuous.

(c) If $f : X \to Y$ is a bijective function between topological spaces X and Y, and for every open subset U of X, $f(U)$ is open in Y, then f is a homeomorphism.

(d) If X and Y are topological space with the discrete topologies, and if $f : X \to Y$ is a bijection, then the spaces X and Y are homeomorphic.

(e) Let S be a set and let R_1 and R_2 be equivalence relations on S. Then $R = R_1 \cap R_2$ is also an equivalence relation on S.

(f) Let S be a set and let R_1 and R_2 be equivalence relations on S. Then $R = R_1 \cup R_2$ is also an equivalence relation on S.

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 nonempty 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

$$\tau_A = \{O \cap A \mid O \in \tau\}.$$

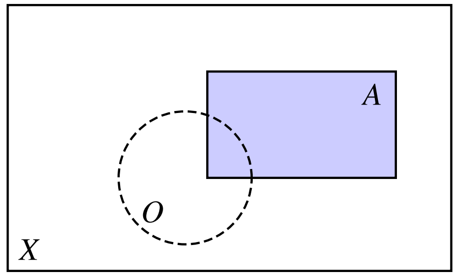

Figure 14.1: A potentially open subset in a subspace.

(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 nonempty 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 nonempty 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 nonempty 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 $\mathbb{Z}$?

The Subspace Topology

In our preview activity, we saw that the intersection of the open sets in a topological space X with any nonempty 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) Recall that a subset of a topological space is closed if its complement is open. Given that (A, τ_A) is a topological space, how is a closed set in A defined? Such a set will be called *relatively closed*.

(b) Recall that a subset U of A is relatively open if and only if $U = A \cap O$ for some open subset of X. With this in mind, 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.

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.

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.

It is left to the exercises 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.

Exercises

(1) Let X and Y be topological spaces and $f : X \to Y$ be a continuous function. If A is a subspace of X, prove that $f|_A : A \to Y$ is also continuous.

(2) Let X be a topological space, let A be a subspace of X, and let B be a subspace of A. Show that the subspace topology that B inherits from A is the same as the subspace topology that B inherits from X.

(3) Let A be a subspace of a topological space X and let B be a subset of A.

(a) Prove that a point x in A is a limit point of B in the subspace topology for A if and only if x is a limit point of B in the topology on X.

(b) Prove that the closure of B in the subspace topology for A is equal to $\overline{B} \cap A$, where $\overline{B}$ is the closure of B in X.

(4) Show that $\mathbb{R}$ is homeomorphic to any interval of the form (a, ∞) or $(-\infty, b)$.

(5) Let X be a topological space.

(a) Let O be an open subset of X. Prove that a subset A of O is open in O if and only if A is open in X.

(b) Let C be a closed subset of X. Prove that a subset B of C is closed in C if and only if B is closed in X.

(6) A property of a topological space is said to be hereditary if that property is inherited by every subspace. We state this more formally in the following definition.

Definition 14.2. A property P of a topological space X is **hereditary** if every subspace of X also has property P.

Show that properties T_1, T_2, and T_3 are hereditary. (The separation axioms T_i are found in Section 12.) The fact that T_4 is not hereditary is somewhat difficult. One example is the Tychonoff plank (which is normal) with the Deleted Tychonoff plank (which is not normal) as subspace. An interested reader can consult *Counterexamples in Topology (2nd ed.)*, Lynn Arthur Steen and J. Arthur Seebach, Jr., Dover Publications, 1978.

(7) Suppose that $f : X \to Y$ is a homeomorphism from a topological space X to a topological space Y. Let $a \in X$. Must the subspace $X' = X \setminus \{a\}$ of X be homeomorphic to the subspace $Y' = Y \setminus \{f(a)\}$ of Y? Prove your conjecture.

(8) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If X has the discrete topology, then every subspace of X has the discrete topology.

(b) If X is a topological space that does not have the discrete topology, then no subspace of X has the discrete topology.

(c) If $f : X \to Y$ is a continuous function between topological spaces X and Y, and X is Hausdorff, then the subspace $f(X)$ of Y is Hausdorff.

(d) If A is a subspace of a topological space X and B is a subset of X, then the closure of $B \cap A$ in the subspace topology for A equals $\overline{B} \cap A$, where $\overline{B}$ is the closure of B in X.

(e) If A is a subspace of a topological space X and C is a subset of X, then the interior of $C \cap A$ in the subspace topology for A equals $\text{Int}(C) \cap A$, where $\text{Int}(C)$ is the interior of C in X.

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.

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 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.

(1) The **quotient topology** on Y is the set

$$\{U \subseteq Y \mid p^{-1}(U) \in \tau_X\}.$$

(2) Any function $p : X \to Y$ is a **quotient map** if p is surjective and for $U \subseteq Y$, U is open in Y if and only if $p^{-1}(U)$ is open in X.

(3) If $p : X \to Y$ is a quotient map, then the space Y is 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.)

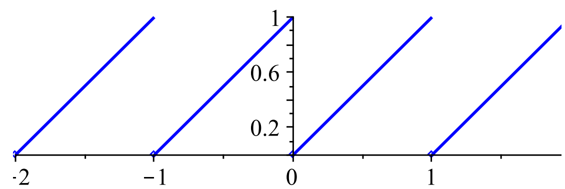

Figure 15.1: The graph of $p(x) = x - \lfloor x \rfloor$.

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^*.

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.

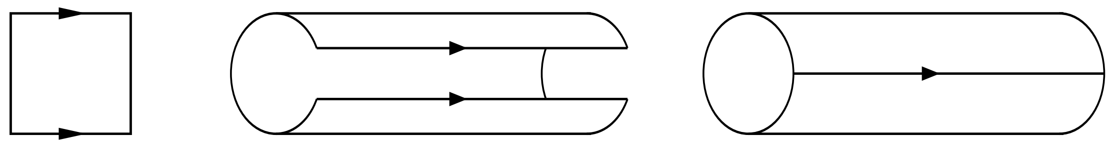

Figure 15.2: A tube as the identification space $X/\!\sim$.

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 $0 < y < 1$ and $0 \le x \le 1$, $(x, 0) \sim (x, 1)$ if $0 \le x \le 1$. 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.

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 $0 < y < 1$ and $0 \le x \le 1$ and $(x, 0) \sim (1 - x, 0)$ when $0 \le x \le 1$.

ii. $(x, y) \sim (x, y)$ if $0 < x < 1$ and $0 < y < 1$, $(x, 0) \sim (x, 1)$ for $0 < x < 1$, $(0, y) \sim (1, y)$ for $0 < y < 1$, 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.

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 $0 < y < 1$, $(x, 0) \sim (1 - x, 1)$ 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 $f = \overline{f} \circ p$.

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

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 $Y = S^1$.

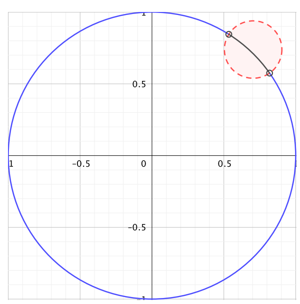

Figure 15.3: A basis element for S^1.

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 15.3. 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 $\overline{f}([t]) = f(t)$ 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.

Exercises

(1) Let X be the real numbers with the standard topology and let $p : X \to \{a, b, c\}$ be defined by

$$p(x) = \begin{cases} a & \text{if } x < 0 \\ b & \text{if } x = 0 \\ c & \text{if } x > 0. \end{cases}$$

What is the quotient topology?

(2) Define an equivalence relation $\sim$ on $\mathbb{R}^2$ by $(x_1, y_1) \sim (x_2, y_2)$ whenever $x_2 - x_1 \in \mathbb{Z}$ and $y_2 - y_1 \in \mathbb{Z}$.

(a) Prove that $\sim$ is an equivalence relation on $\mathbb{R}^2$.

(b) The quotient space is a familiar space. Find that space and explain why it is the quotient space.

(3) Find an example of a continuous surjection that is not a quotient map

(4) Let X be a topological space and let A be a subspace of X. Define a relation $\sim$ on X whose equivalence classes are A and $\{x\}$ if $x \notin A$. In this case the quotient space is denoted as X/A (think of this space as obtained by crushing A to a point and leaving everything else alone). Describe each of the following quotient spaces.

(a) X is the closed interval $[0, 1]$ in $\mathbb{R}$ and $A = \{0, 1\}$

(b) $X = \{(x, y) \mid x^2 + y^2 = 1\}$, $A = \{(-1, 0), (1, 0)\}$

(c) If $X = \{(x, y) \mid x^2 + y^2 \leq 1\}$ and $A = \{(x, y) \mid x^2 + y^2 = 1\}$

(5) Let (X, τ) be the topological space where $X = \{1, 2, 3, 4\}$ and

$$\tau = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4\}, \{2, 3, 4\}, \{1, 3, 4\}, \{1, 2, 3, 4\}\}.$$

Let $Y = \{a, b, c\}$.

(a) Let $p : X \to Y$ be defined by $p(1) = a$, $p(2) = a$, $p(3) = b$, and $p(4) = c$. Find the quotient topology τ_p on Y defined by the function p.

(b) Let $q : X \to Y$ be defined by $q(1) = c$, $q(2) = c$, $q(3) = b$, and $q(4) = a$. Find the quotient topology τ_q on Y defined by the function q.

(c) Are the spaces (Y, τ_p) and (Y, τ_q) homeomorphic? If yes, write down a specific homeomorphism and explain why your mapping is a homeomorphism. If not, explain why not.

(6) Let $D^2 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ be the unit disk in $\mathbb{R}^2$ with the standard topology, and let $S^1 = \{(x, y) \mid x^2 + y^2 = 1\}$ be the boundary of D^2. Describe the quotient spaces $D^2/\sim$ for each equivalence relation (assume that points are similar to themselves). Let $x = (s_1, t_1)$ and $y = (s_2, t_2)$.

(a) $x \sim y$ if $s_1 = s_2$ for x, y in D^2

(b) $x \sim y$ if $s_1 = s_2$ for x, y in S^1

(c) $x \sim y$ if x and y are diagonally opposite each other for x, y in D^2

(7) Let $X = I \times I$ where I is the interval $[0, 1]$ with the standard metric topology, and define an equivalence relation on X by $x \sim x$ for all $x \in I$ and $(s_1, t_1) \sim (s_2, t_2)$ when $t_1 = t_2 > 0$.

(a) Describe the quotient space $X/\sim$, and describe the quotient topology.

(b) Show that $X/\sim$ is not Hausdorff.

(8) Let X be a nonempty set and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

Let $\sim$ be the equivalence relation on $\mathbb{Z}$ defined by $x \sim y$ if $x \equiv y \pmod 3$. Describe the quotient space $\mathbb{Z}/\sim$ and then determine, with justification, the quotient topology on $\mathbb{Z}/\sim$ when

(a) $\mathbb{Z}$ has the topology τ_p with $p = 1$

(b) $\mathbb{Z}$ has the topology τ_e with $p = 1$.

(9) In the process of developing techniques of drawing in perspective, renaissance artists found it necessary to consider a point at infinity where all lines intersect. This creates a geometry that extends the concept of the real plane. This new geometry is the real projective plane $\mathbb{R}P^2$, which can be thought of as the quotient space of $\mathbb{R}^3 \setminus \{(0,0,0)\}$ with the relation $\sim_P$ such that $(x_1, x_2, x_3) \sim_P (y_1, y_2, y_3)$ in $\mathbb{R}^3 \setminus \{(0,0,0)\}$ if and only if there is a nonzero real number k such that $y_1 = kx_1$, $y_2 = kx_2$, and $y_3 = kx_3$. In the projective plane, parallel lines intersect at a point at infinity, just as they seem to with our human vision.

(a) Show that $\sim_P$ is an equivalence relation.

(b) Give a geometric description of the elements in the quotient space $\mathbb{R}P^2$.

(c) There are other ways to visualize $\mathbb{R}P^2$. For example, explain why the real projective plane $\mathbb{R}P^2$ is homeomorphic to the quotient space $S^2/\sim$ of $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mod x^2+y^2+z^2 = 1\}$, where $\sim$ identifies antipodal points on S^2. No formal proofs are necessary, but a convincing explanation is in order.

(d) Since we identify antipodal points on S^2 in the space $S^2/\sim$, we can think of this space in another way. If P is a point on S^2 not on the equator, then its antipodal point is also not on the equator. So we can think of $S^2/\sim$ as the top hemisphere, along with the equator on which antipodal points are identified, as illustrated at left in Figure 15.4.

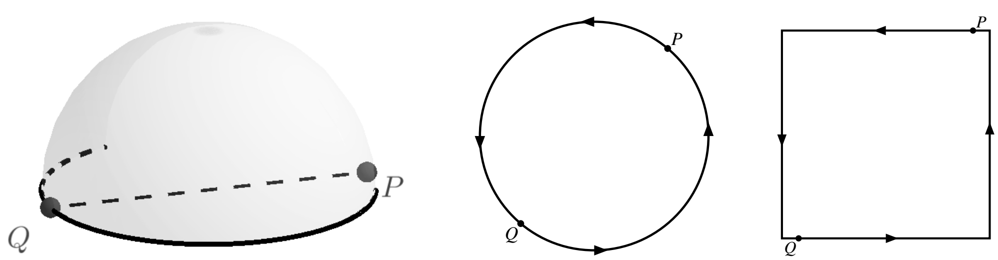

Figure 15.4: Three perspectives of $\mathbb{R}P^2$.

By projecting the points on the hemisphere down to the xy-plane, we can represent $S^2/\sim$ as a disk whose antipodal points are identified, as seen in the middle in Figure

15.4. Use this last perspective to explain why $\mathbb{R}P^2$ can be realized as a square where opposite sides are identified in opposite directions as shown at right in Figure 15.4.

(e) The projective plane $\mathbb{R}P^2$ is a complicated object – it cannot be embedded in $\mathbb{R}^3$ and so it is not something that can be easily visualized. The projective plane is a non-orientable surface and is also important in classifying surfaces – basically, every closed surface is made up of spheres, tori, and/or projective planes. In this part of the exercise we see how the projective plane itself is made by adjoining a Möbius strip to a disk (think of sewing the boundary of Möbius strip to the perimeter of a disk).

i. Start with the model of $\mathbb{R}P^2$ shown at left in Figure 15.5. Partition this object into three pieces as shown at right in Figure 15.5. Explain why the shaded region in the middle figure, separated out at right in Figure 15.5, is a Möbius strip.

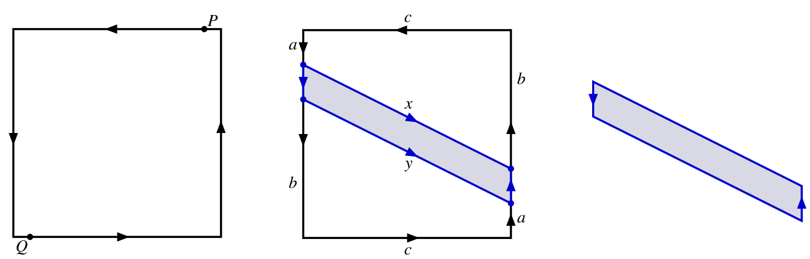

Figure 15.5: Splitting the real projective plane.

ii. The space S that remains after we remove the Möbius strip from $\mathbb{R}P^2$ are shown at left in Figure 15.6. The two spaces that follow are homeomorphic to S. Describe the homeomorphisms that produce the spaces from S. Then explain how $\mathbf{R}P^2$ is obtained by attaching a Möbius strip to a disk along its boundary.

(10) Let $X = \mathbb{R}^2$ with the standard topology, and let $Y = \{x \in \mathbb{R} \mid x \geq 0\}$ with the standard topology. Let $f : X \to Y$ be defined by $f((x, y)) = x^2 + y^2$.

(a) Show that f is a continuous surjective function.

(b) Prove that the quotient space X^* of X defined by f is homeomorphic to Y.

(11) Let (X, τ) be a topological space. A subspace A of X is a *retract* of X (or that X retracts onto A) if there is a continuous function $r : X \to A$ such that $r(a) = a$ for all $a \in A$. Such a map r is called a *retraction*. Intuitively, a subspace A of X is a retract of X if we can continually collapse (or retract) X onto A without moving any of the points in A. Certain types of retracts, namely deformation retracts, are important in algebraic topology.

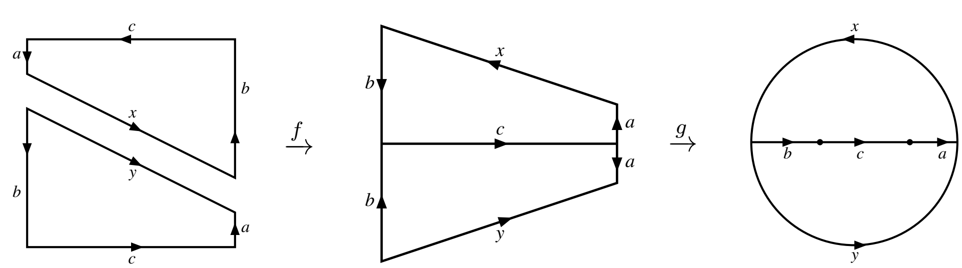

Figure 15.6: Recognizing the space S.

(a) Show that every nonempty space can retract to a point.

(b) Show that $\{-1, 1\}$ is a retract of $\mathbb{R} \setminus \{0\}$.

(c) Show that every retraction is a quotient map.

(d) Show that if X is Hausdorff and A is a retract of X, then A must be a closed subset of X.

(12) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) Let p be a surjection from a topological space X to a nonempty set Y. The quotient topology on Y is the largest topology on Y such that p is continuous.

(b) Let $X = [0, 1]$ with the Euclidean metric topology and define the relation $\sim$ on X as $x \sim y$ if x and y are either both rational or both irrational. Then the quotient space $X/\sim$ is a two point space with the indiscrete topology.

(c) Let p be a surjection from a topological space X to a nonempty set Y. The quotient topology on Y is the largest topology on Y such that p is continuous.

(d) Let $X = [0, 1]$ with the Euclidean metric topology and define the relation $\sim$ on X as $x \sim y$ if x and y are either both rational or both irrational. Then the quotient space $X/\sim$ is a two point space with the indiscrete topology.

(e) If X is a topological space, Y is a set, and $p : X \to Y$ is a surjection, then $p(U)$ is open in the quotient topology.

(f) If $\sim$ is an equivalence relation on a topological space X, then the quotient space $X/\sim$ is the set of all equivalence classes of X with the quotient topology.

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.

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.

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.*

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 nonempty 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, ..., a_n in A such that the sets O_{a_1}, O_{a_2}, ..., 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?

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 $M > 0$.*

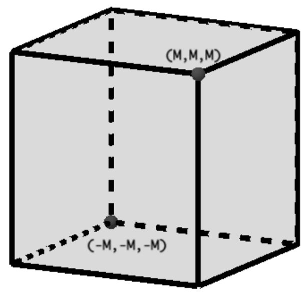

Figure 16.1: A standard 3-cube Q_M^3.

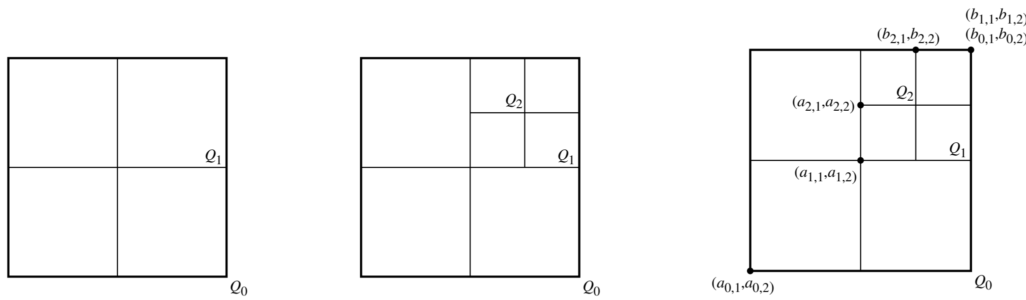

Figure 16.2: Left : Q_1. Middle: Q_2. Right: Labeling the corners.

Proof. We proceed by contradiction and assume that there is an $n \in \mathbb{Z}^+$ and a positive real number M such that Q_M^n is not compact. So there exists an open cover $\{O_\alpha\}$ with α in some indexing set I of Q_M^n that has no finite sub-cover. Let $Q_0 = Q_M^n$ 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 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.*

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}, ..., 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.

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.

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.

Exercises

(1) Let τ_1 and τ_2 be two topologies on a set X. If $\tau_1 \subseteq \tau_2$, then τ_1 is said to be a *coarser* (or *weaker*) topology than τ_2, or τ_2 is a *finer* (or *stronger*) topology than τ_1. The finest topology on any set is the topology on which every subset is open. This is the discrete topology. The coarsest topology on any set is the one that has as open sets the minimal collection of sets that must be open. This is the indiscrete topology. It is a reasonable question to ask what properties of a topological space, if any, are passed from weaker to stronger topologies or from stronger to weaker. We investigate two properties in this exercise. First we explore the notation of weaker and stronger topologies.

(a) Let $X = \{a, b, c\}$. Are there any topologies σ on X such that σ is not the discrete topology but there are no stronger topologies on X other than the discrete topology? Explain.

(b) Let τ_1 and τ_2 be two topologies on a set X. If $\tau_1 \subseteq \tau_2$, what does compactness under τ_1 or τ_2 imply, if anything, about compactness under the other topology? Justify your answers.

(2) Let (X, τ) be a topological space

(a) Prove that the union of any finite number of compact subsets of X is a compact subset of X.

(b) If X is Hausdorff, prove that the intersection of any finite number of compact subsets of X is a compact subset of X.

(3) Let $\mathbb{E}$ be the set of even integers, and let $\tau = \{\mathbb{Z}\} \cup \{O \subseteq \mathbb{E}\}$. That is, τ is the collection of all subsets of $\mathbb{E}$ along with $\mathbb{Z}$.

(a) Prove that τ is a topology on $\mathbb{Z}$.

(b) Find all compact subsets of $(\mathbb{Z}, \tau)$. Verify your answer.

(c) Prove or disprove: If A and B are compact subsets of a topological space X, then $A \cap B$ is also a compact subset of X.

(4) Let X be a topological space.

(a) Prove that if X is Hausdorff and C is a compact subset of X, then for each $x \in X \setminus C$ there exist disjoint open sets U and V such that $x \in U$ and $C \subseteq V$.

(b) Prove that if X is a compact Hausdorff space, then X is normal.

(5) Let X be a nonempty set and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

Determine, with proof, the compact subsets of X when

(a) X has the particular point topology τ_p

(b) X has the excluded point topology $\tau_{\overline{p}}$.

(6) Let X be a topological space. A family $\{F_\alpha\}_{\alpha \in I}$ of subsets of X is said to have the *finite intersection property* if for each finite subset J of I, $\bigcap_{\alpha \in J} F_\alpha \neq \emptyset$. Prove that X is compact if and only if for each family $\{F_\alpha\}_{\alpha \in I}$ of closed subsets of X that has the finite intersection property, we have $\bigcap_{\alpha \in I} F_\alpha \neq \emptyset$.

(7) Even though $\mathbb{R}$ is not a compact space, if $x \in \mathbb{R}$, then $x \in [x-1, x+1]$ and so every point in $\mathbb{R}$ is contained in a compact subset of $\mathbb{R}$. So if we view $\mathbb{R}$ from the perspective of a point in $\mathbb{R}$, the space $\mathbb{R}$ looks compact around that point. This is the idea of local compactness. Locally compact spaces are important in the general topology of function spaces.

Definition 16.13. A topological space X is **locally compact** if for each $x \in X$ there is an open set O such that $p \in O$ and $\overline{O}$ is compact.

(a) Explain why $\mathbb{R}^n$ is locally compact for each $n \in \mathbb{Z}^+$.

(b) Show that any compact space is locally compact.

(c) Consider the Sorgenfrey line from Exercise 5 in Section 12. Recall that the Sorgenfrey line is the space $\mathbb{R}$ with a basis $\mathcal{B} = \{[a, b) \mid a < b \text{ in } \mathbb{R}\}$ for its topology. Show that the Sorgenfrey line is Hausdorff but not locally compact.

(8) Introduced by Felix Hausdorff in the early 20th century as a way to measure the distance between sets, the Hausdorff metric (also called the Pompeiu-Hausdorff metric) has since been widely studied and has many applications. For example, the United States military has used the Hausdorff distance in target recognition procedures. In addition, the Hausdorff metric has

been used in image matching and visual recognition by robots, medicine, image analysis, and astronomy.

The basic idea in these applications is that we need a way to compare two shapes. For example, if a manufacturer needs to mill a specific product based on a template, there is usually some tolerance that is allowed. So the manufacturer needs a way to compare the milled parts to the template to determine if the tolerance has been met or exceeded.

The Hausdorff metric is also familiar to fractal aficionados for describing the convergence of sequences of compact sets to their attractors in iterated function systems. The variety of applications of this metric make it one that is worth studying.

To define the Hausdorff metric, we begin with the distance from a point x in a metric space X to a subset A of X as

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

Since images will be represented as compact sets, we restrict ourselves to compact subsets of a metric space. In this case the infimum becomes a minimum and we have

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

We now extend that idea to define the distance from one subset of X to another. Let A and B be nonempty compact subsets of X. To find the distance from the set A to the set B, it seems reasonable to consider how far each point in A is from the set B. Then the distance from A to B should measure how far we have to travel to get from *any* point in A to B.

Definition 16.14. Let (X, d) be a metric space and let A and B be nonempty compact subsets of X. Then **distance** $d(A, B)$ **from** A **to** B is

$$d(A, B) = \max_{a \in A} \left\{ \min_{b \in B} \{d(a, b)\} \right\}.$$

Note: since A and B are compact, $d(A, B)$ is guaranteed to exist.

(a) A problem with d as in Definition 16.14 is that d is not symmetric. Find examples of compact subsets A and B of $\mathbb{R}^n$ with the Euclidean metric such that $d(A, B) \neq d(B, A)$.

(b) Even though the function d in Definition 16.14 is not a metric, we can define the Hausdorff distance in terms of d as follows.

Definition 16.15. Let (X, d) be a metric space and A and B nonempty compact subsets of X. Then **Hausdorff distance between** A **and** B is

$$h(A, B) = \max\{d(A, B), d(B, A)\}.$$

Let A be the circle in $\mathbb{R}^2$ centered at the origin with radius 1, let B be the inscribed square, and let $C = \{(1, 0), (-1, 0)\}$ as shown in Figure 16.3.

Determine $h(A, B)$, $h(A, C)$, and $h(B, C)$, and verify that $h(A, C) \leq h(A, B) + h(B, C)$.

(c) It may be surprising that h as in Definition 16.15 is actually a metric, but it is. The underlying space is the collection of nonempty compact subsets of X which we denote at $\mathcal{H}(X)$. Prove the following theorem.

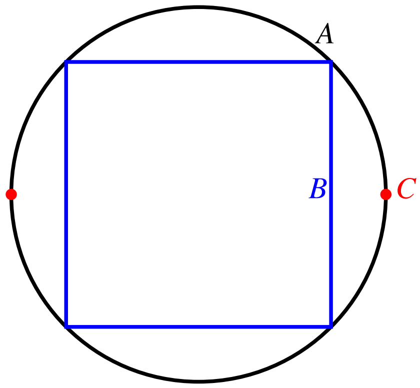

Figure 16.3: Sets A, B, and C.

Theorem 16.16. *Let X be a metric space. The Hausdorff distance function is a metric on $\mathcal{H}(X)$.*

(d) One fun application of the Hausdorff metric is in fractal geometry. You may be familiar with objects like the Sierpinski triangle or the Koch curve. These objects are limits of sequences of sets in $\mathcal{H}(\mathbb{R}^2)$. We illustrate with the Sierpinski triangle. Start with three points v_1, v_2, and v_3 that form the vertices of an equilateral triangle S_0. For i=1,2, or 3, let $v_i = \begin{bmatrix} a_i \\ b_i \end{bmatrix}$. For i=1,2, or 3, we define $\omega_i : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\omega_i\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_i \\ b_i \end{bmatrix}.$$

Then ω_i, when applied to S_0, contracts S_0 by a factor of 2 and then translates the image of S_0 so that the i^{th} vertices of S_0 and the image of S_0 coincide. Such a map is called a *contraction mapping* with *contraction factor* equal to $\frac{1}{2}$. Define $S_{1,i}$ to be $\omega_i(S_0)$. Then $S_{1,i}$ is the set of all points half way between any point in S_0 and v_i, or $S_{1,i}$ is a triangle half the size of the original translated to the i^{th} vertex of the original. Let $S_1 = \bigcup_{i=1}^{3} S_{1,i}$. Both S_0 and S_1 are shown in figure 16.4. We can continue this procedure replacing S_0 with S_1. In other words, for $i = 1$, 2, and 3, let $S_{2,i} = \omega_i(S_1)$. Then let $S_2 = \bigcup_{i=1}^{3} S_{2,i}$. A picture of S_2 is shown in figure 16.4. We can continue this procedure, each time replacing S_{j-1} with S_j. A picture of S_8 is shown in figure 16.4.

To continue this process, we need to take a limit. But the S_i are sets in $\mathcal{H}(\mathbb{R}^2)$, so the limit is taken with respect to the Hausdorff metric.

i. Assume that the length of a side of S_0 is 1. Determine $h(S_0, S_1)$. Then find $h(S_k, S_{k+1})$ for an arbitrary positive integer k.

ii. The Sierpinski triangle will exist if the sequence (S_n) converges to a set S (which would be the Sierpinski triangle). The question of convergence is not a trivial one.

A. Consider the sequence (a_n), where $a_n = \left(1 + \frac{1}{n}\right)^n$ for $n \in \mathbb{Z}^+$. Note that each a_n is a rational number. Explain why the terms in this sequence get

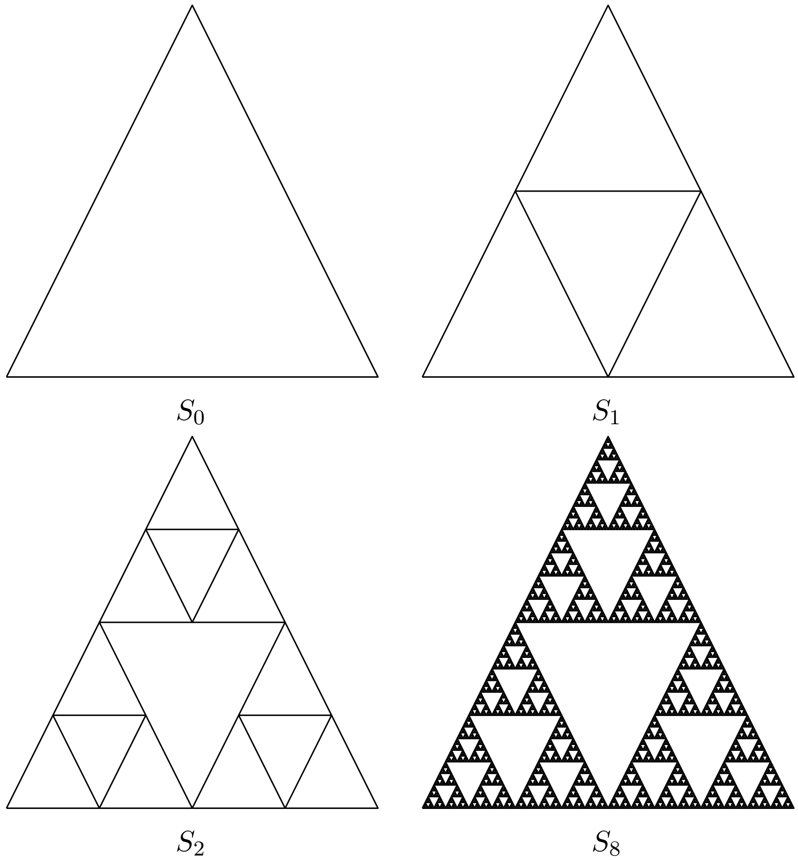

Figure 16.4: S_i for i equal to 0, 1, 2, and 8.

arbitrarily close together, but the sequence does not converge in $\mathbb{Q}$. Explain why the sequence (a_n) converges in $\mathbb{R}$.

B. A sequence (x_n) in a metric space (X, d) is a *Cauchy sequence* if given $\epsilon > 0$ there exists $N \in \mathbb{Z}^+$ such that $d(x_n, x_m) < \epsilon$ whenever $n, m \geq N$. Every convergent sequence is a Cauchy sequence. A metric space X is said to be *complete* if every Cauchy sequence in X converges to an element in X. For example, $(\mathbb{R}, d_E)$ is complete while $(\mathbb{Q}, d_E)$ is not. Although we will not prove it, the metric space $(\mathcal{H}(\mathbb{R}^2), h)$ is complete. Show that the sequence (S_n) is a Cauchy sequence in $\mathcal{H}(\mathbb{R}^2)$. The limit of this sequence is the famous Sierpinski triangle. The picture of S_8 in figure 16.4 is a close approximation of the Sierpinski triangle.

(9) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If X and Y are compact topological spaces and $f : X \to Y$ is a continuous bijection, then f is a homeomorphism.

(b) If X is a compact topological space, then any closed subspace of X is compact.

(c) If X is a Hausdorff space, Y is a compact space, and $f : X \to Y$ is a continuous and bijective function, then f is a homeomorphism.

(d) If X is a compact space, Y is a Hausdorff space, and $f : X \to Y$ is a continuous bijection, then f is a homeomorphism.

(e) Let C be a closed subset of a metric space (X, d) with the metric topology. Then C is compact.

(f) If A is a compact subset of a topological space X, then A is a closed subset of X.

(g) Let (X, τ) be a topological space with τ the discrete topology. Then X is compact if and only if X is finite.

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 nonempty 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

$$\tau = \{U \cap X \mid U \text{ is open in the Euclidean Topology on } \mathbb{R}^2\}.$$

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 ... ".

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 nonempty 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.*

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 nonempty 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 nonempty 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.

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 nonempty 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 $C \subseteq V$.*

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 nonempty 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 $(\mathbb{R}, d_E)$.

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 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?

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.

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 a homeomorphism. 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 $(Y \setminus f(S)) \subseteq (f(U) \cup f(V))$. 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 $x \notin S$. 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) 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.)

(b) 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.)

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 than not homeomorphic.

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.

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 nonempty 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.

Exercises

(1) Let τ_1 and τ_2 be two topologies on a set X. If $\tau_1 \subseteq \tau_2$, then τ_1 is said to be a *coarser* (or *weaker*) topology than τ_2, or τ_2 is a *finer* (or *stronger*) topology than τ_1. The finest topology on any set is the topology on which every subset is open. This is the discrete topology. The coarsest topology on any set is the one that has as open sets the minimal collection of sets that must be open. This is the indiscrete topology. It is a reasonable question to ask what properties of a topological space, if any, are passed from weaker to stronger topologies or from stronger to weaker. We investigate two properties in this exercise. First we explore the notation of weaker and stronger topologies.

(a) Let $X = \{a, b, c\}$. Are there any topologies γ on X such that γ is not the indiscrete topology but there are no weaker topologies on X other than the indiscrete topology? Explain.

(b) Let τ_1 and τ_2 be two topologies on a set X. If $\tau_1 \subseteq \tau_2$, what does connectedness under τ_1 or τ_2 imply, if anything, about connectedness under the other topology? Justify your answers.

(2) Let $f : [a, b] \to \mathbb{R}$ be a continuous function from a closed interval into the reals. Let $U = f(u)$ and $V = f(v)$ be such that $U \leq f(x) \leq V$ for all $x \in [a, b]$. Prove that there is a w between u and v such that $f(w)(b - a) = \int_a^b f(t)\, dt$.

(3) Let A be a connected subset of a topological space X. Prove or disprove:

(a) $\text{Int}(A)$ is connected

(b) $\overline{A}$ is connected

(c) $\text{Bdry}(A)$ is connected

(4) Let $X = \mathbb{R}$ with the finite complement topology. We have shown that every subset of any topological space with the finite complement topology is compact. Now find all of the connected subsets of X. Prove your result.

(5) Give examples, with justification, of subsets A and B of a topological space to illustrate each of the following, or explain why no such example exists:

(a) A and B are connected, but $A \cap B$ is disconnected

(b) A and B are connected, but $A \setminus B$ is disconnected

(c) A and B are disconnected, but $A \cup B$ is connected

(d) A and B are connected and $A \cap B \neq \emptyset$, but $A \cup B$ is disconnected.

(e) A and B are connected and $\overline{A} \cap \overline{B} \neq \emptyset$, but $A \cup B$ is disconnected.

(6) Let $f : S^1 \to \mathbb{R}$ be a continuous function. Show that there is a point $x \in S^1$ with $f(x) = f(-x)$.

(7) Let $a, b \in \mathbb{R}$ with $a < b$. Explain why no two of the sets (a, b), $(a, b]$, and $[a, b]$ homeomorphic.

(8) Even though $X = (0, 1) \cup (1, 2)$ is not a connected space, if x is any element in X then we can surround x with a connected subset of X. This is the idea of local connectedness.

Definition 17.14. A topological space X is **locally connected at a point** $x \in X$ if every neighborhood U of x contains an open connected neighborhood of x. A topological space X is **locally connected** if X is locally connected at each point in X.

(a) Give an example of a locally connected space that is not connected.

(b) It would be reasonable to believe that a connected space is locally connected. However, that is not the case. Consider the space $X = A \cup B$ as a subspace of $\mathbb{R}^2$ with the standard Euclidean metric topology, where $A = \{(x, y) \mid x \text{ is irrational and } 0 \leq y \leq 1\}$ and $B = \{(x, y) \mid x \text{ is rational and } -1 \leq y \leq 0\}$.

i. Explain why X is connected.

ii. Show that X is not locally connected. (Hint: Let x be a point not on the x-axis and find an open ball around x that doesn't intersect the x-axis.)

(c) Prove that a topological space X is locally connected if and only if for every open set O in X, the connected components of O are open in X.

(9) Let A and B be nonempty subsets of a topological space X.

(a) Prove that $A \cup B$ is disconnected if $(\overline{A} \cap B) \cup (A \cap \overline{B}) = \emptyset$.

(b) Prove that X is connected if and only if for every pair of nonempty subsets A and B of X such that $X = A \cup B$ we have $(\overline{A} \cap B) \cup (A \cap \overline{B}) \neq \emptyset$.

(10) Give examples of the following.

(a) A topological space with exactly one cut point.

(b) A topological space with exactly two cut points.

(c) A topological space with infinitely many cut points.

(d) A topological space with no cut points.

(11) Let $a, b \in \mathbb{R}$ with $a < b$. Prove that a homeomorphism $f : [a, b] \to [a, b]$ carries end points into end points.

(12) Let X and Y be a topological spaces.

(a) Assume that X and Y homeomorphic spaces. Prove that there is a one-to-one correspondence between the connected components of X and the connected components of Y.

(b) Assume that X and Y homeomorphic spaces. Prove that there is a one-to-one correspondence between the set of cut points of X and the set of cut points of Y.

(c) Consider each letter in the statement as a topological space with the standard Euclidean metric topology.

TOPOLOGY IS NEAT

Group the letters in the statement into disjoint homeomorphism classes. Explain in detail the reasons for your groupings.

(13) Let (X, τ) be a topological space.

(a) Prove that X is disconnected if and only if X has a proper subset that is both open and closed.

(b) Prove that X is disconnected if and only if there is a continuous function from X onto a discrete two-point topological space.

(14) Let X be the set of real numbers.

(a) Consider X with the topology $\tau_1 = \{\emptyset, [0, 1], X\}$. Prove or disprove: X is connected.

(b) Consider X with the topology $\tau_2 = \{U \subseteq X \mid 0 \in U\} \cup \{\emptyset\}$.

i. Prove or disprove: X is connected.

ii. Prove or disprove: $X \setminus \{0\}$ is connected.

(15) Let X be a nonempty set and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

Determine, with proof, the connected subsets of X when

(a) X has the particular point topology τ_p

(b) X has the excluded point topology $\tau_{\overline{p}}$.

(16) Let (X, τ) be a topological space and A a connected subset of X.

(a) Show that if X is Hausdorff, then A' is connected.

(b) Let $(X, \tau) = (\mathbb{R}, \tau_0)$, where τ_0 is the particular point topology on X. Explain why $A = \mathbb{Z}$ is a connected subset of X. Find $\mathbb{Z}'$ in $(\mathbb{R}, \tau_0)$. Is it true that in any topological space, if A is connected, then so is A'? Explain. (See Exercise 15.)

(17) Let X be a topological space. Prove each of the following.

(a) Each $a \in X$ is an element of exactly one connected component C_a of X.

(b) 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.

(c) If a and b are in X, then either $C_a = C_b$ or $C_a \cap C_b = \emptyset$.

(d) Every connected subset of X is a subset of a connected component.

(e) Every connected component of X is a closed subset of X.

(f) The space X is connected if and only if X has exactly one connected component.

(18) Let X be a topological space with only a finite number of connected components. Show that each component of X is open.

(19) Let X and Y be connected spaces with $f : X \to Y$ a continuous function. Is it the case that if S is a cut set of X, then $f(S)$ is a cut set of Y? Prove your answer.

(20) Let $X = \{a, b, c, d\}$. There are 355 distinct topologies on X, but they fit into the 33 distinct homeomorphism classes listed below. The list is ordered by decreasing number of singleton sets in the topology, and, when that is fixed, by increasing number of two-point subsets and then by increasing number of three-point subsets. Under which topologies is X connected? Prove your answer.

1. the discrete topology
2. $\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}, X\}$
3. $\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}, \{a, b, d\}, X\}$
4. $\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, d\}, \{a, b, c\}, \{a, b, d\}, \{a, c, d\}, X\}$
5. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, X\}$
6. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, X\}$
7. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c, d\}, X\}$
8. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, \{a, b, d\}, X\}$
9. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, b, c\}, X\}$
10. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \{a, c, d\}, X\}$
11. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \{a, b, d\}, X\}$
12. $\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{c, d\}, \{a, c, d\}, \{b, c, d\}, X\}$

13. $\{\emptyset, \{a\}, \{b\}, \{a,b\}, \{a,c\}, \{a,d\}, \{a,b,c\}, \{a,b,d\}, X\}$
14. $\{\emptyset, \{a\}, \{b\}, \{a,b\}, \{a,c\}, \{a,d\}, \{a,b,c\}, \{a,b,d\}, \{a,c,d\}, X\}$
15. $\{\emptyset, \{a\}, X\}$
16. $\{\emptyset, \{a\}, \{a,b\}, X\}$
17. $\{\emptyset, \{a\}, \{a,b\}, \{a,b,c\}, X\}$
18. $\{\emptyset, \{a\}, \{b,c\}, \{a,b,c\}, X\}$
19. $\{\emptyset, \{a\}, \{a,b\}, \{a,c,d\}, X\}$
20. $\{\emptyset, \{a\}, \{a,b\}, \{a,b,c\}, \{a,b,d\}, X\}$
21. $\{\emptyset, \{a\}, \{b,c\}, \{a,b,c\}, \{b,c,d\}, X\}$
22. $\{\emptyset, \{a\}, \{a,b\}, \{a,c\}, \{a,b,c\}, X\}$
23. $\{\emptyset, \{a\}, \{a,b\}, \{a,c\}, \{a,b,c\}, \{a,b,d\}, X\}$
24. $\{\emptyset, \{a\}, \{c,d\}, \{a,b\}, \{a,c,d\}, X\}$
25. $\{\emptyset, \{a\}, \{a,b\}, \{a,c\}, \{a,d\}, \{a,b,c\}, \{a,b,d\}, \{a,c,d\}, X\}$
26. $\{\emptyset, \{a\}, \{a,b,c\}, X\}$
27. $\{\emptyset, \{a\}, \{b,c,d\}, X\}$
28. $\{\emptyset, \{a,b\}, X\}$
29. $\{\emptyset, \{a,b\}, \{c,d\}, X\}$
30. $\{\emptyset, \{a,b\}, \{a,b,c\}, X\}$
31. $\{\emptyset, \{a,b\}, \{a,b,c\}, \{a,b,d\}, X\}$
32. $\{\emptyset, \{a,b,c\}, X\}$
33. $\{\emptyset, X\}$

(21) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If A is a connected subset of a topological space X with $|A| \geq 2$, then every point of A is a limit point of A.

(b) If A is a compact subspace of a Hausdorff space, then A is connected.

(c) If A is a connected subspace of a Hausdorff space, then A is compact.

(d) Every subset of a topological space with the discrete topology is disconnected.

(e) The set $\{a,b\}$ is a connected component of the topological space $X = \{a,b,c,d\}$ with topology

$\tau = \{\emptyset, \{a\}, \{a,b\}, \{c\}, \{d\}, \{c,d\}, \{a,c,d\}, \{a,c\}, \{a,d\}, \{a,b,c\}, \{a,b,d\}, X\}.$

(f) The sets $U = \{a,c,d\}$ and $V = \{a,b,c\}$ form a separation of the set $A = \{c,d\}$ in the topological space $X = \{a,b,c,d\}$ with topology

$\tau = \{\emptyset, \{a\}, \{a,b\}, \{c\}, \{d\}, \{c,d\}, \{a,c,d\}, \{a,c\}, \{a,d\}, \{a,b,c,\}, \{a,b,d\}, X\}.$

(g) The connected topological space $X = \{a,b,c,d\}$ with topology

$$\tau = \{\emptyset, \{a\}, \{b,c\}, \{a,b,c\}, X\}$$

has no cut points.

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.

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.

A consequence of Theorem 18.3 is the following.

Corollary 18.4. *Path connectedness is a topological property.*

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.

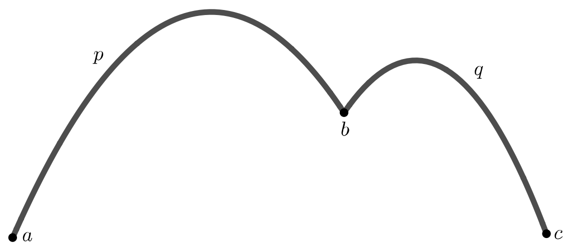

Figure 18.1: A path from a to c.

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*

$$h(x) = \begin{cases} f(x) & \textit{if } x \in A \\ g(x) & \textit{if } x \in B \end{cases}$$

is continuous.

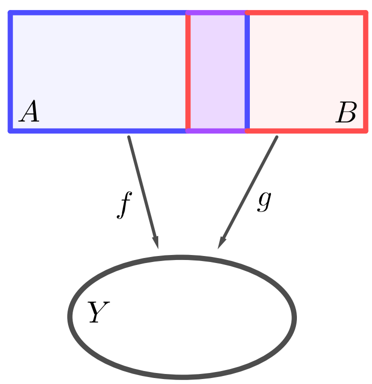

Figure 18.2: The Gluing Lemma.

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

$$(q * p)(x) = \begin{cases} p(2x) & \text{for } 0 \leq x \leq \frac{1}{2} \\ q(2x-1) & \text{for } \frac{1}{2} \leq x \leq 1. \end{cases}$$

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.

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.

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.

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 open 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 an open 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 an open 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?)

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.

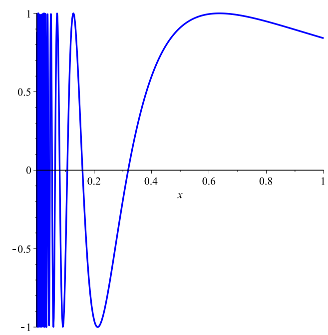

Figure 18.3: The topologist's sine curve.

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.

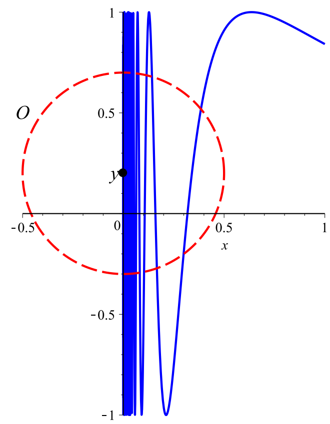

Figure 18.4: The set O.

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 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.

Exercises

(1) Let X be a topological space and for each $x \in X$ let $PC(x)$ denote the path component of x. Prove the following.

  (a) If A is a path connected subset of X, then $A \subseteq PC(x)$ for some $x \in X$.

  (b) The space X is path connected if and only if $X = PC(x)$ for some $x \in X$.

(2) In Activity 17.6 of Section 17 we showed that an arbitrary union of connected sets is connected provided the intersection of those sets is not empty. Is the same result true for path connected sets. That is, if X is a topological space and $\{A_\alpha\}$ for α in some indexing set I is a collection of path connected subsets of X and $\bigcap_{\alpha \in I} A_\alpha \neq \emptyset$, must it be the case that $A = \bigcup_{\alpha \in I} A_\alpha$ is path connected? Prove your answer.

(3) Let X be the subspace of $(\mathbb{R}^2, d_E)$ consisting of the line segments joining the point $(0, 1)$ to every point in the set $\left\{\left(\frac{1}{n}, 0\right) \mid n \in \mathbb{Z}^+\right\}$ as illustrated in Figure 18.5. This space is called the *harmonic broom*.

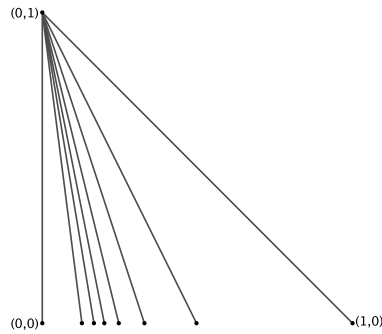

Figure 18.5: The harmonic broom.

(a) Show that the harmonic broom is connected.

(b) Show that the harmonic broom is path connected.

(c) Show that the harmonic broom is not locally connected.

(d) Show that the harmonic broom is not locally path connected. So path connectedness does not imply local path connectedness.

(4) In Exercise 3 we see an example of a space that is path connected but not locally path connected. Is it possible to find a space that is locally path connected but not path connected? Verify your answer.

(5) We know that a space can be connected but not path connected. We also know that local path connectedness does not imply connectedness. However, if we combine these conditions then a space must be path connected. That is, show that if a topological space X is connected and locally path connected, then X is path connected.

(6) Let X be a nonempty set and let p be a fixed element in X. Let τ_p be the particular point topology and $\tau_{\overline{p}}$ the excluded point topology on X. That is

- τ_p is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that contain p.
- $\tau_{\overline{p}}$ is the collection of subsets of X consisting of $\emptyset$, X, and all of the subsets of X that do

Determine, with proof, the path connected subsets of X when

(a) X has the particular point topology τ_p

(b) X has the excluded point topology $\tau_{\overline{p}}$.

(7) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses.

(a) If X is a path connected topological space, then any subspace of X is path connected.

(b) If A and B are path connected subspaces of a topological space X, then $A \cap B$ is path connected.

(c) There is no path from a to b in (X, τ), where τ is the discrete topology.

(d) If X is a compact locally path connected topological space, then X has only finitely many path components.

(e) Every locally path connected space is locally connected.

Section 19

Products of 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 the product of a finite number of topological spaces?
- How do we define a topology on the product of a finite number of topological spaces?
- What is a projection map from a product of a finite number of topological spaces?
- How can we use projection maps to determine the continuity of a function to a product of a finite number of topological spaces?
- What is a subbasis of a topological space?
- What properties do product spaces inherit from their factors?

Introduction

If we have two topological spaces (X, τ_X) and (Y, τ_Y), we might wonder if we can make the set $X \times Y$ into a topological space. A natural approach might be to take as the open sets in $X \times Y$ the sets of the form $U \times V$ where $U \in \tau_X$ and $V \in \tau_Y$.

Preview Activity 19.1. Let $X = \{a, b, c\}$ with $\tau_X = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, X\}$, and let $Y = \{1, 2\}$ with $\tau_Y = \{\emptyset, \{1\}, Y\}$.

(1) Let

$$\mathcal{B} = \{U \times V \mid U \in \tau_X \text{ and } V \in \tau_Y\}. \tag{19.1}$$

List all of the sets in $\mathcal{B}$ along with their elements.

(2) Assume that all of the sets in $\mathcal{B}$ are open sets in $X \times Y$. Should the set $A = \{(a, 1), (a, 2), (b, 1)\}$

be an open set in $X \times Y$? Is the set A of the form $U \times V$ for some open sets U in X and V in Y? Explain. Is $\mathcal{B}$ a topology on $X \times Y$?

(3) If $\mathcal{B}$ is not a topology on $X \times Y$, what is the smallest collection of sets would we need to add to $\mathcal{B}$ to make a topology on $X \times Y$? Explain your process.

The Topology on a Product of Topological Spaces

In our preview activity we learned that we cannot make a topology on a product $X \times Y$ of topological spaces (X, τ_X) and (Y, τ_Y) with just the sets of the form $U \times V$ where $U \in \tau_X$ and $V \in \tau_Y$ as the open sets since the collection of these sets is not closed under arbitrary unions. What we can do instead is consider these unions of all of the sets of the form $U \times V$, where U is open in X and V is open in Y. In other words, consider these sets to be a basis for the topology on $X \times Y$.

Activity 19.1. Let (X, τ) and (Y, τ_Y) be topological spaces, and let $\mathcal{B}$ be as defined in (19.1). Prove that $\mathcal{B}$ is a basis for a topology on $X \times Y$.

The argument from Activity 19.1 can be extended to a product of any finite number of topological spaces. Let n be a positive integer and let (X_i, τ_i) be topological spaces for i from 1 to n. Let

$$\mathcal{B} = \left\{\Pi_{i=1}^n O_i \mid O_i \text{ is open in } X_i\right\}.$$

Since $X_i \in \tau_i$ for every i, every point in $\Pi_{i=1}^n X_i$ is in a set in $\mathcal{B}$. So $\mathcal{B}$ satisfies condition 1 of a basis. Now we show that $\mathcal{B}$ satisfies the second condition of a basis. Let $B_1 = \Pi_{i=1}^n U_i$ and $B_2 = \Pi_{i=1}^n V_i$ for some open sets U_i, V_i in X_i. Suppose $(x_i) \in (B_1 \cap B_2)$. Then for each j we have $x_j \in U_j \cap V_j$ and so

$$(x_i) \in \Pi_{i=1}^n (U_i \cap V_i).$$

Since $U_i \cap V_i$ is an open set in X_i, it follows that $\Pi_{i=1}^n (U_i \cap V_i)$ is in $\mathcal{B}$. Thus, $\mathcal{B}$ is a basis for a topology on $X \times Y$.

This topology generated by products of open sets is called the *box* or *product* topology.

Definition 19.1. Let (X_α, τ_α) be a collection of topological spaces for α in some finite indexing set I. The **box topology** or **product topology** on the product $\Pi_{\alpha \in I} X_\alpha$ is the topology with basis

$$\mathcal{B} = \left\{\Pi_{\alpha \in I} U_\alpha \mid U_\alpha \in \tau_\alpha \text{ for each } \alpha \in I\right\}.$$

So we can always make the product of topological spaces into a topological space using the box topology.

Three Examples

In this section we consider three specific examples of a product of topological spaces.

Activity 19.2. Let $X = [1, 2]$ and $Y = [3, 4]$ as subspaces of $\mathbb{R}^2$.

(a) Explain in detail what the product space $X \times Y$ looks like.

(b) Find, if possible, an open subset of $X \times Y$ that is not of the form $U \times V$ where U is open in X and V is open in Y.

Activity 19.3. Let $X = \mathbb{R}$ and $Y = S^1 = \{(x, y) \mid x^2 + y^2 = 1\}$, the unit circle as a subset of $\mathbb{R}^2$.

(a) Draw a picture of $\mathbb{R}$. For each $x \in \mathbb{R}$, the set $\mathbb{R}_x = \{(x, y) \mid y \in S^1\}$ is a subset of $\mathbb{R} \times S^1$. On your graph of $\mathbb{R}$, draw pictures of $\mathbb{R}_x$ for x equal to -1, 0, and 1. Explain in detail what the product space $\mathbb{R} \times S^1$ looks like.

(b) Consider the sets of the form $B \cap S^1$, where B is an open ball in $\mathbb{R}^2$ (relatively open sets in S^1). What do these sets look like?

(c) Describe the shape of the basis elements for the product topology on $\mathbb{R} \times S^1$ that result from products of the form $U \times V$, where U is an open interval in $\mathbb{R}$ and V is the intersections of S^1 with an open ball in $\mathbb{R}^2$.

Activity 19.4. Let $2S^1 = \{(x, y) \mid x^2 + y^2 = 4\}$ be the circle of radius 2 centered at the origin as a subset of $\mathbb{R}^2$. In this activity we investigate the space $2S^1 \times S^1$.

(a) Draw a picture of $2S^1$ in the xy-plane. For each $p \in S^1$, the set $S^1_p = \{(p, y) \mid y \in S^1\}$ is a subset of $S^1 \times S^1$. On your graph of S^1, draw pictures of S^1_p for p equal to $(1, 0)$, $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$, and $(0, 1)$. Orient the graphs so that the copies of S^1 are perpendicular to $2S^1$. Explain in detail what the product space $2S^1 \times S^1$ looks like.

(b) Consider the sets of the form $B \cap S^1$, where B is an open ball in $\mathbb{R}^2$. What do these sets look like?

(c) Describe the shape of the basis elements for the product topology on $2S^1 \times S^1$ that result from products of the form $U \times V$, where U and V are intersections of S^1 with open balls in $\mathbb{R}^2$.

Projections, and Continuous Functions on Products

Given topological spaces (X_1, τ_1) and (X_2, τ_2), we define $\pi_1 : X_1 \times X_2 \to X_1$ and $\pi_2 : X_1 \times X_2 \to X_2$ by $\pi_1((x, y)) = x$ and $\pi_2((x, y)) = y$. These functions π_1 and π_2 are the *projections* of $X_1 \times X_2$ onto X_1 and X_2, respectively. These projection functions can help us determine when a function f from a topological space Y to $X_1 \times X_2$ is continuous.

Activity 19.5. Let (X_1, τ_1) and (X_2, τ_2) be topological spaces and let O_1 be an open set in X_1.

(a) Determine which set is $\pi_1^{-1}(O_1)$. Verify your conjecture.

(b) Explain why π_1 is continuous.

The same argument as in Activity 19.5 shows that π_2 is also a continuous function. In general, if $X = \Pi_{\alpha \in I} X_\alpha$ is a finite product of topological spaces, then the projection $\pi_\alpha : X \to X_\alpha$ is a continuous function for each $\alpha \in I$. We note here that there is another topology, called the *product topology*) on X with subbasis $S = \bigcup S_{\alpha \in I} S_\alpha$, where

$$S_\alpha = \{\pi_\alpha^{-1}(U_\alpha \mod U_\alpha \text{ is open in } X_\alpha\}.$$

For reasons we won't go into, the product topology is preferred to the box topology for infinite products (many important theorems that hold for finite products will not hold for infinite products using the box topology, but will hold using the product topology). However, the product topology and the box topology are the same for finite products, and since we won't consider infinite products here we will not worry about the distinction. For our purposes we will use the terms "box topology" and "product topology" interchangeably.

Now suppose that X_1, X_2, and Y are topological spaces, and that $f : Y \to X_1 \times X_2$ is a function. Then $\pi_1 \circ f$ maps Y to X_1 and $\pi_2 \circ f$ maps Y to X_2. Since the composition of continuous functions is continuous, we can see that if f is continuous so are $\pi_1 \circ f$ and $\pi_2 \circ f$. To determine if f is a continuous function, it would be useful to know if the converse is true. The next activity provides a first step.

Activity 19.6. Let X_1, X_2, and Y be topological spaces, and let $f : Y \to X_1 \times X_2$ be a function. Assume that both $\pi_1 \circ f$ and $\pi_2 \circ f$ are continuous. Let O_1 be an open set in X_1 and O_2 an open set in X_2.

(a) What set is $(\pi_1 \circ f)^{-1}(O_1)$? Verify your conjecture.

(b) What set is $(\pi_2 \circ f)^{-1}(O_2)$? Verify your conjecture.

(c) Explain why f is continuous.

Let $O = \Pi_{i=1}^n O_i$ be a basic open set in $X = \Pi_{i=1}^n X_i$. We can extend the result of Activity 19.5 to see that

$$\pi_i^{-1}(O_i) = X_1 \times X_2 \times \cdots \times X_{i-1} \times O_i \times X_{i+1} \times \cdots \times X_n.$$

So

$$\Pi_{i=1}^n O_i = \bigcap_{i=1}^n \pi_i^{-1}(O_i).$$

So each basic open set is a finite intersection of sets of the form $\pi_i^{-1}(O_i)$ where O_i is open in X_i. When this happens, we call the collection of sets of the form $\pi_i^{-1}(O_i)$ a *subbasis* of the topology.

Definition 19.2. Let (X, τ) be a topological space. A subset $\mathcal{S}$ of τ is a **subbasis** or **subbase** for τ if the set of all finite intersections of elements of $\mathcal{S}$ is a basis for τ.

As an example, since finite intersections of intervals of the form $(-\infty, b)$ and (a, ∞) give all intervals of the form (a, b), the collection $\mathcal{S} = \{(-\infty, b), (a, \infty) \mid a, b \in \mathbb{R}\}$ is a subbasis for the standard topology on $\mathbb{R}$. Note that this collection itself is not a basis for the standard topology on $\mathbb{R}$. If $X = \Pi_{i=1}^n X_i$ is a product of topological space, then another example of a subbasis is the collection

$$\mathcal{S} = \bigcup_{i=1}^n \{\pi_i^{-1}(O_i) \mid O_i \text{ is open in } X_i\}.$$

This set is a subbasis for the product topology on X (the verification of this is left to the exercises).

As we have discussed before, it can often be easier to define a topology using a basis or subbasis than it is to describe all of the sets in the topology. As we might expect, since the continuity of a function can be determined by the inverse image of basis elements, the continuity of a function can also be determined by the inverse image of subbasis elements.

Activity 19.7. Prove Theorem 19.3. (Hint: Recall that f is continuous if $f^{-1}(B)$ is open in X for each basic open set B.)

Theorem 19.3. *Let (X, τ_X) and (Y, τ_Y) be topological spaces, let $\mathcal{S}$ be a subbasis for τ_Y, and let $f : X \to Y$ be a function. If $f^{-1}(S)$ is open in X for each $S \in \mathcal{S}$, then f is continuous.*

Now we can use projections to determine when functions to product spaces are continuous.

Theorem 19.4. *Let X_i for i from 1 to n and Y be topological spaces, and let $f : Y \to \Pi X_i$ be a function. Then f is continuous if and only if $\pi_i \circ f$ is continuous for each i.*

Proof. Let X_i for i from 1 to n and Y be topological spaces, and let $f : Y \to \Pi X_i$ be a function. If f is continuous, the facts that each π_i is continuous and that composites of continuous functions are continuous show that $\pi_i \circ f$ is continuous for each i.

Now suppose that $\pi_i \circ f$ is continuous for each i. Recall that

$$\mathcal{S} = \{\pi_i^{-1}(O_i) \mid O_i \text{ is open in } X_i\}$$

is a subbasis for the product topology on $\Pi_{i=1}^n X_i$. To prove that f is continuous, Theorem 19.3 tells us that it is enough to show that $f^{-1}(S)$ is open for each S in $\mathcal{S}$. Let O_i be an open set in X_i. Then

$$f^{-1}(\pi_i^{-1}(O_i)) = (\pi_i \circ f)^{-1}(O_i)$$

which is open in Y because $\pi_i \circ f$ is continuous. Therefore, f is continuous. ■

Properties of Products of Topological Spaces

It is natural to ask what topological properties of the topological spaces (X, τ_X) and (Y, τ_Y) are inherited by the product $X \times Y$. We have studied Hausdorff, connected, and compact spaces, and we now consider those properties.

Activity 19.8. Let (X, τ_X) and (Y, τ_Y) be Hausdorff spaces.

(a) What will it take to prove that the space $X \times Y$ with the product topology is Hausdorff?

(b) Suppose that $(x_1, y_1), (x_2, y_2) \in X \times Y$. What does the fact that X is Hausdorff tell us about x_1 and x_2? What can we say about y_1 and y_2?

(c) Complete the proof of the following theorem.

Theorem 19.5. *If (X, τ_X) and (Y, τ_Y) are Hausdorff spaces, then $X \times Y$ with the product topology is a Hausdorff space.*

The proofs that a product of connected spaces is connected, that a product of path connected spaces is path connected, and that a product of compact spaces is compact are a bit more complicated. To prove that a product of two connected spaces is connected, we will use the result of Activity 17.6 in Section 17 that the union of connected subsets is connected if the intersection of the subsets is nonempty. A consequence of this result is the following.

Lemma 19.6. *Let X be a topological space, and let A_α be a connected subset of X for all α in some indexing set I. Let B be a connected subset of X such that $A_\alpha \cap B \neq \emptyset$ for every $\alpha \in I$. Then $B \cup (\bigcup_{\alpha \in I} A_\alpha)$ is connected.*

Proof. Let X be a topological space, and let A_α be a connected subset of X for all α in some indexing set I. Let B be a connected subset of X such that $A_\alpha \cap B \neq \emptyset$ for every $\alpha \in I$. For each $\alpha \in I$ let $B_\alpha = B \cup A_\alpha$. Let $\beta \in I$. Since $B \cap A_\beta \neq \emptyset$, Lemma 19.6 shows that B_β is connected. Given that B is not empty, and $B \subseteq \bigcap_{\alpha \in I} B_\alpha$, we see that $\bigcap_{\alpha \in I} B_\alpha \neq \emptyset$. Lemma 19.6 allows us to conclude that $\bigcup_{\alpha \in I} B_\alpha$ is connected. But

$$\bigcup_{\alpha \in I} B_\alpha = \bigcup_{\alpha \in I} (B \cup A_\alpha) = B \cup \left(\bigcup_{\alpha \in I} A_\alpha \right),$$

and so $B \cup (\bigcup_{\alpha \in I} A_\alpha)$ is connected. ∎

We will use Lemma 19.6 to show that a product of connected spaces is connected.

Theorem 19.7. *If (X, τ_X) and (Y, τ_Y) are connected topological spaces, then $X \times Y$ with the product topology is a connected topological space.*

Proof. Assume (X, τ_X) and (Y, τ_Y) are connected topological spaces. Our approach to proving that $X \times Y$ is connected is to write $X \times Y$ as a union of two connected subspaces whose intersection is not empty. Let $a \in X$. The space $X_a = \{a\} \times Y$ is homeomorphic to Y via the inclusion map i which sends $(a, t) \in \{a\} \times Y$ to the point $t \in Y$. Since Y is connected, so is X'. Let $b \in Y$. The space $Y_b = X \times \{b\}$ is homeomorphic to X via the inclusion map i which sends $(s, b) \in X \times \{b\}$ to the point $s \in X$. Since X is connected, so is Y_b. (The verification of these homeomorphisms is left to the reader.) The point (a, b) is in $X_a \cap Y_b$, so $X_a \cap Y_b \neq \emptyset$ for every $b \in Y$. It follows that $X_a \cup (\bigcup_{t \in Y} Y_t)$ is connected by Lemma 19.6. All that remains is to prove that $X_a \cup (\bigcup_{t \in Y} Y_t) = X \times Y$ and we will have demonstrated that $X \times Y$ is connected. The fact that $X_a \subseteq X \times Y$ and $Y_t \subseteq X \times Y$ for every $t \in Y$ implies that $X_a \cup (\bigcup_{t \in Y} Y_t) \subseteq X \times Y$. It then remains to show that $X \times Y \subseteq X_a \cup (\bigcup_{t \in Y} Y_t)$. Let $(u, v) \in X \times Y$. Then $u \in X$ and $v \in Y$ and $(u, v) \in Y_v$. Thus, $X \times Y \subseteq X_a \cup (\bigcup_{t \in Y} Y_t)$ and so $X \times Y = X_a \cup (\bigcup_{t \in Y} Y_t)$. Therefore, $X \times Y$ is connected. ∎

Once we know that a product of connected topological spaces is connected, we can extend that result to any finite number of connected spaces by induction.

Corollary 19.8. *Let X_k be a connected topological space for k from 1 to n. Then the product $\Pi_{k=1}^n X_k$ is connected.*

The proof is left to the exercises.

We conclude this section by demonstrating that a product of compact topological spaces is compact. It is also true that finite products of path connected and compact spaces are path connected and compact. The proofs are left to the exercises.

Theorem 19.9. *If X and Y are compact topological spaces, then $X \times Y$ is a compact topological space under the product topology.*

Proof. Let (X, τ_X) and (Y, τ_Y) be compact topological spaces. Let $\mathcal{C} = \{O_\alpha\}$ be an open cover of $X \times Y$ for α in some indexing set I. Let $a \in X$ and let $Y_a = \{a\} \times Y$. Since Y_a is homeomorphic to Y, we know that Y_a is compact. The collection $\{O_\alpha \cap Y_a\}$ is an open cover of Y_a, and so has a finite sub-cover $\{O_{\alpha_i}\}_{1 \leq i \leq n}$. The set $N_a = \bigcup_{1 \leq i \leq n} O_{\alpha_i}$ is an open set that contains Y_a. We will show that there is a neighborhood W_a of a that N_a contains the entire set $W_a \times Y$.

Cover the set Y_a with open sets that are contained in N_a (since N_a is open, we can intersect any open set with N_a and still have an open set). Each open set is a union of basis elements, so we can cover Y_a with basis elements $U \times V$ that are contained in N_a. Since Y_a is compact, there is a finite collection $U_1 \times V_1$, $U_2 \times V_2$, $\ldots$, $U_m \times V_m$ of basis elements contained in N_a that cover Y_a. Assume that each $U_i \times V_i$ intersects Y_a (otherwise, we can remove that set and still have a cover). Let $W_a = U_1 \cap U_2 \cap \cdots \cap U_m$. Since $a \in U_i$ for each i, we know that W_a is not empty. Each U_i is open in X and so W_a is open in X. Thus, W_a is a neighborhood of a in X. Now we demonstrate that $W_a \times Y \subseteq \bigcup_{1 \leq i \leq m} U_i \times V_i$. Let $(x, y) \in W_a \times Y$. Since the collection $\{U_i \times V_i\}_{1 \leq i \leq m}$ covers Y_a, the point (a, y) is in $U_k \times V_k$ for some k between 1 and m. So $y \in Y_k$. But $x \in W_a = \bigcap_{1 \leq i \leq m} U_i$, so $x \in U_k$. Thus, $(x, y) \in U_k \times V_k$ and we conclude that $W_a \times Y \subseteq \bigcup_{1 \leq i \leq m} U_i \times V_i$.

So for each $a \in X$, the set N_a contains a set of the form $W_a \times Y$, where W_a is a neighborhood of a in X. So $W_a \times Y$ is covered by a finite sub-cover of our open cover $\mathcal{C}$ of $X \times Y$. The collection $\{W_a \times Y\}_{a \in X}$ is an open cover of $X \times Y$. Since X is compact, the is a finite sub-cover W_1, W_2, $\ldots$, W_r of the open cover $\{W_a\}_{a \in X}$ of X. is an open cover of X. It follows that the sets $W_1 \times Y$, $W_2 \times Y$, $\ldots$, $W_r \times Y$ is a cover of $X \times Y$. For each i, the set $W_i \times Y$ is covered by finitely many of the sets in $\mathcal{C}$, and so the collection of these sets forms a finite sub-cover of $X \times Y$ in $\mathcal{C}$. Therefore, $X \times Y$ is compact. ∎

Summary

Important ideas that we discussed in this section include the following. Throughout, let (X_i, τ_i) be topological spaces for i from 1 to some integer n

- The product of the X_i is the Cartesian product $\Pi_{i=1}^n X_i$.
- The set
$$\mathcal{B} = \{\Pi_{i=1}^n O_i \mid O_i \text{ is open in } X_i\}$$
is a basis for a topology on $\Pi_{i=1}^n X_i$.
- The mapping $\pi_j : \Pi_{i=1}^n X_i \to X_j$ defined by $\pi_j((x_i)) = x_j$ is the projection map onto X_j for j from 1 to n.
- A function f mapping a topological space Y to $\Pi_{i=1}^n X_i$ is continuous if and only if $\pi_j \circ f$ is continuous for every j from 1 to n.
- Let (X, τ) be a topological space. A subset $\mathcal{S}$ of τ is a subbasis for τ if the set $\mathcal{S}$ of all finite intersections of elements of $\mathcal{S}$ is a basis for τ.
- If each X_i is (a) connected, (b) path connected, (c) compact, then $\Pi_{i=1}^n X_i$ is (a) connected, (b) path connected, (c) compact with respect to the product topology.

Exercises

(1)

(a) Let (Y_1, τ_1) and (Y_2, τ_2) be topological spaces, where $Y_1 = \{a, b, c\}$ with $\tau_1 = \{\emptyset, \{a\}, \{b, c\}, Y_1\}$ and $Y_2 = \{1, 2\}$ with $\tau_2 = \{\emptyset, \{1\}, Y_2\}$. Find all of the sets of the form

$$\mathcal{S} = \{\pi_1^{-1}(O_1) \mid O_1 \text{ is open in } Y_1\} \cup \{\pi_2^{-1}(O_2) \mid O_2 \text{ is open in } Y_2\}$$

and verify that these sets generate the product topology on $Y_1 \times Y_2$.

(b) Let X_i for i from 1 to n be topological spaces and let $X = \Pi_{i=1}^n X_i$. Show that the collection

$$\mathcal{S} = \bigcup_{i=1}^{n} \{\pi_i^{-1}(O_i) \mid O_i \text{ is open in } X_i\}$$

is a subbasis for the box topology on X.

(2) Let X be the set of real numbers with the standard Euclidean metric topology and let Y be the real numbers with the discrete topology.

(a) Explain why the set of all "horizontal intervals" of the form

$$I = (a, b) \times \{c\} = \{(x, c) \mid a < x < b\}$$

is a base for the product topology on $X \times Y$.

(b) Find the interior and closure of each of the following subsets of $X \times Y$.

(a) $A = \{(x, 0) \mid 0 \leq x < 1\}$
(b) $B = \{(0, y) \mid 0 \leq y < 1\}$
(c) $C = \{(x, y) \mid 0 \leq x < 1, 0 \leq y < 1\}$

(3) Let X_1 and X_2 be topological space and let A_1 be a subset of X_1 and A_2 a subset of X_2. Assume the product topology on $X_1 \times X_2$. Prove each of the following.

(a) $\overline{A_1 \times A_2} = \overline{A_1} \times \overline{A_2}$

(b) $\text{Int}(A_1 \times A_2) \subseteq \text{Int}A_1 \times \text{Int}A_2$

(4) Let X and Y be topological spaces and let $\{U_\alpha\}_{\alpha \in I}$ and $\{V_\beta\}_{\beta \in J}$ be collections of open sets in X and Y, respectively, for some indexing sets I and J. Show that

$$\bigcup_{\substack{\alpha \in I \\ \beta \in J}} (U_\alpha \times V_\beta) = \left(\bigcup_{\alpha \in I} U_\alpha \right) \times \left(\bigcup_{\beta \in J} V_\beta \right).$$

(5)

(a) If S_1, S_2, T_1, and T_2 are sets, show that

$$(S_1 \times T_1) \cap (S_2 \times T_2) = (S_1 \cap S_2) \times (T_1 \cap T_2).$$

(b) If $\mathcal{B}_X$ is a base for a topology τ_X on a space X and $\mathcal{B}_Y$ is a base for a topology τ_Y on a space Y, show that $\mathcal{B}_X \times \mathcal{B}_Y$ is a base for the product topology on $X \times Y$.

(6) Prove that the product of any finite number of connected spaces is connected.

(7) Prove that the product of any finite number of compact spaces is compact.

(8)

(a) Prove that if (X_1, τ_1) and (X_2, τ_2) are path connected topological spaces, then $X_1 \times X_2$ with the product topology is a path connected topological space.

(b) Prove that the product of any finite number of path connected spaces is path connected.

(9) Let X_1 and X_2 be topological spaces and $X = X_1 \times X_2$. Is it true that if X is compact, then both X_1 and X_2 are compact. Prove your answer.

(10) Let X_1 and X_2 be topological spaces and let $\pi_i : X_1 \times X_2 \to X_i$ be the projection mapping. We have shown that π_i is continuous. Now show that π_i is an open map for i equal 1 and 2. Assume the standard product topology.

(11) Let X_1 and X_2 be topological spaces with cardinalities at least 2. Let $X = X_1 \times X_2$. Prove that the product space topology on $X = X_1 \times X_2$ is the discrete topology if and only if the topologies on X_1 and X_2 are the discrete topologies.

(12) **Digital Topology.** Computers represent information from the real world digitally. That is, a computer screen consists of discrete pixels that are used to mimic the continuous information from the real world. So we exist in $\mathbb{R}^3$, but a computer screen represents information in $\mathbb{Z}^2$ as illustrated in Figure 19.1. It is important to be able to accurately mimic the continuous information from digital data. One of the key ideas is to have a digital version of the Jordan curve theorem which states that a Jordan curve (a continuous loop that does not intersect itself) in the Euclidean plane separates the remainder of the plane into two connected components (the inside and the outside of the curve). Additionally, if a single point is removed from a Jordan curve, the remainder of the plane becomes connected. The reason a digital Jordan curve theorem is important is that it is only necessary to save the Jordan curves which determine regions, along with the associated colors of the regions, rather than having to save the color of every single pixel in an image.

A natural start to building a digital topology might be to identify neighborhoods. The idea of a neighborhood is to consider elements that are close to a point, and in the digital world there are different ways to do this. Given a point (x, y) in $\mathbb{Z}^2$, the 4-neighbors of (x, y) are the points vertically or horizontally adjacent to (x, y): that is, the points $(x \pm 1, y)$ and $(x, y \pm 1)$. The 8-neighbors of (x, y) are the 4-neighbors along with the points diagonally adjacent to (x, y): that is, $(x \pm 1, y)$, $(x, y \pm 1)$, $(x \pm 1, y \pm 1)$. These neighbors are illustrated in Figure 19.1, with the crosses indicating the neighbors of the highlighted point.

In the continuous case, we define a path between points to be a continuous function from $[0, 1]$ to the space. However, we cannot have continuity in the digital world. So we define paths by moving through neighbor points. That is, if k is either 4 or 8, a k-path is a finite sequence p_0, p_1, ..., p_m in $\mathbb{Z}^2$ such that p_1 is a k-neighbor of p_2, p_2 is a k-neighbor of p_3, ..., p_{m-1} is a k-neighbor of p_m.

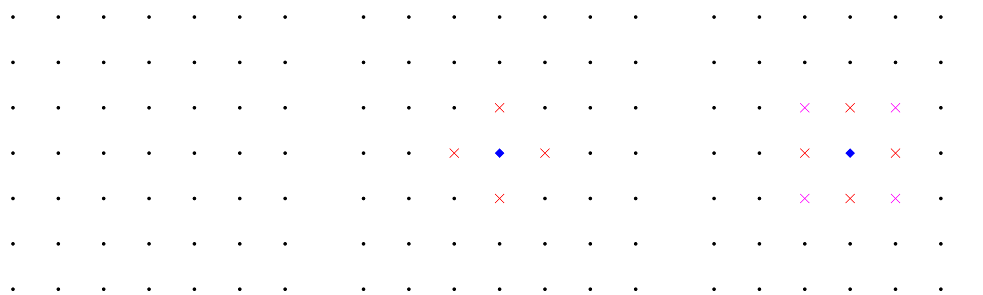

Figure 19.1: Left: The digital plane. Middle: 4-neighbors of a point. Right: 8-neighbors of a point.

(a) Show that there is a 4-path connecting any two points in $\mathbb{Z}^2$. Then explain why there is an 8-path connecting any two points in $\mathbb{Z}^2$.

(b) In the continuous case, every Jordan curve separates $\mathbb{R}^2$ into two connected regions. To have a similar theorem in the discrete case, we need a notion of connectedness in $\mathbb{Z}^2$. Every image is made up of a finite number of pixels, and so we can think of a digital image as existing in a finite subspace of $\mathbb{Z}^2$. Since connectedness and path connectedness are equivalent in finite topological spaces, we us the idea of k-paths to define connectedness in $\mathbb{Z}^2$. We say that a subset S of $\mathbb{Z}^2$ is $k-connected$ if if any two of its points can be joined by a k-path in S.

Figure 19.2 show two sets (curves) in the digital plane indicated by the points that connect the line segments (examples taken from A Topological Approach to Digital Topology, T. Yung Kong, R. Kopperman, and P. Meyer, *American Mathematical Monthly*, 98 (1991), no. 10, 901-917). Let S_1 be the set illustrated at left in Figure 19.2 and S_2 the set at right.

Figure 19.2: Sets S_1 (left) and S_2 (right) in the digital plane.

Is S_1 4 connected? Is S_1 8 connected? Verify your answer. Repeat with S_2.

(c) We can now define a Jordan k-curve to be a finite k-connected set which contains exactly two k-neighbors for each of its points.

Is S_1 a Jordan 4-curve? Is S_1 a Jordan 8-curve? Verify your answer. Repeat with S_2.

(d) As usual, we define a component to be a maximal connected set. Explain why S_1 is a Jordan 8-curve whose complement is connected and why S_2 is a Jordan 4-curve whose complement consists of three connected 4-components. This example shows that there is no Jordan curve theorem in digital topology using the standard notions of k-connectedness with k either 4 or 8. So neither 4-adjacency nor 8-adjacency provides an analogue of the Jordan curve theorem and it is necessary to use a combination of both. That is, a Jordan 4-curve with at least five points separates $\mathbb{Z}^2$ into exactly two 8-components, and a Jordan 8-curve with at least five points separates $\mathbb{Z}^2$ into exactly two 4-components. In Exercise 13 we examine a topology that does admit a digital Jordan curve theorem.

(13) In Exercise 12 we discussed the importance of a digital Jordan curve theorem. In this exercise we describe a topology in which such a theorem exists. Consider Z with the topology τ_1 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* or the *Khalimsky topology* on $\mathbb{Z}$. Notice that all sets of the form $\{n\}$ are open when n is odd.

(a) Show that any set of the form $\{n\}$ where n is even is closed in the digital line topology.

(b) To define a *Khalimsky topology* on $\mathbb{Z}^2$ we use the product topology. Explain why the collection of sets $\{B(m,n)\}$ where

$$B(m,n) = \begin{cases} \{(m,n)\} & m \text{ and } n \text{ odd,} \\ \{(m-i, n-j) \mid -1 \leq i \leq 1, -1 \leq j \leq 1\} & m \text{ and } n \text{ even,} \\ \{(m, n-1), (m,n), (m, n+1)\} & m \text{ odd and } n \text{ even,} \\ \{(m-1, n), (m,n), (m+1, n)\} & m \text{ even and } n \text{ odd} \end{cases}$$

is a basis for the Khalimsky topology τ_2 on $\mathbb{Z}^2$. (This topology was originally published by E. Khalimsky in *Applications of connected ordered topological spaces in topology*, Conference of math. departments of Povolsia, 1970.)

(c) New we want to define a digital Jordan curve. Our first step is to define a *digital path*. Recall that a path in a topological space is a homeomorphism from the interval $[0,1]$ into the space. So we need the concept of a digital interval. If $z_1 < z_2$ in $(\mathbb{Z}, \tau_1)$, the *digital interval* $[z_1, z_2]$ is the set

$$[z_1, z_2] = \{z \in \mathbb{Z} \mid z_1 \leq z \leq z_2\}.$$

The integers z_1 and z_2 are called the *endpoints* of the digital interval $[z_1, z_2]$.

Definition 19.10. Let X be a topological space.

- A **digital path** in X is the range of a continuous function from a digital interval to X.

- A **digital arc** in X is the range of a homeomorphism from a digital interval to X.

Let

$$S_1 = \{(1,-1),(1,1),(-1,1),(-1,-1)\},$$
$$S_2 = \{(0,0),(1,-1),(2,0),(1,1)\}, \text{ and}$$
$$S_3 = \{(1,-1),(1,0),(1,1),(0,1),(-1,1),(-1,0),(-1,-1),(0,-1)\}.$$

Show that S_1 is not a digital path but S_2 and S_3 are digital paths.

(d) To produce a digital Jordan Curve Theorem, we need a definition of a digital Jordan curve.

Definition 19.11. A **digital Jordon curve** is a finite connected set J with $|J| \geq 4$ such that $J \setminus \{j\}$ is a digital arc for each $j \in J$.

So every digital Jordan curve is a connected set. Show that any finite digital path in $\mathbb{Z}^2$ is a connected set. (Hint: Is every digital interval connected?)

(e) The upshot of all of this is the following theorem (a proof can be found in A Topological Approach to Digital Topology, T. Yung Kong, R. Kopperman, and P. Meyer, *American Mathematical Monthly*, 98 (1991), no. 10, 901-917).

Theorem 19.12. *If J is a digital Jordon curve in the digital plane $\mathbb{Z}^2$, then $\mathbb{Z}^2 \setminus J$ has exactly two components.*

The two components in Theorem 19.12 split the digital plane into an infinite region (the outside) and a finite region (the inside).
Show that S_2 is a digital Jordan curve (and thus splits $\mathbb{Z}^2$ into two connected components).

(14) Digital Jordan curves, as described in Exercise 13, are important in order to have a digital Jordan curve theorem. Christer O. Kiselman presents the following theorem to characterize digital Jordan curves in *Discrete Geometry for Computer Imagery*, Springer-Verlag, 2000, p. 46-56.

Theorem 19.13. *A subset J of $\mathbb{Z}^2$ equipped with the Khalimsky topology is a digital Jordan curve if and only if $J = \{P_1, P_2, \ldots, P_m\}$ for some even integer $m \geq 4$ and for all j, $P_{j?1}$ and P_{j+1} and no other points are adjacent to P_j; moreover each path consisting of three consecutive points $P_{i?1}$, P_i, P_{i+1} turns at P_i by $45°$ or $90°$ or not at all if P_i is a pure point, and goes straight ahead if P_i is mixed.*

We investigate this theorem in this exercise.

(a) We need to first define the appropriate terms. Let X be a topological space. Two points x and y in X are *adjacent* if $x \neq y$ and the set $\{x, y\}$ is connected. Then let $N(x)$ to be the intersection of all neighborhoods of x.
Show that distinct elements x and y in a topological space X are adjacent if and only if $x \in N(y)$ or $y \in N(x)$.

(b) A point (x_1, x_2) in $\mathbb{Z}^2$ is called *pure* if x_1 and x_2 have the same parity. Otherwise, the point is *mixed*. Find $N(P)$ if P is a pure point or a mixed point.

(c) In Exercise 13 we show that the set $S_1 = \{(1,-1),(1,1),(-1,1),(-1,-1)\}$ is not a digital path and so not a digital Jordan curve. Which part of Theorem 19.13 does S_1 violate?

(d) In Exercise 13 we show that the set $S_2 = \{(0,0),(1,-1),(2,0),(1,1)\}$ is a digital Jordan curve. Show that in S_2, the property from Theorem 19.13 that x_{j-1} and x_{j+1} and no other points are adjacent to x_j is satisfied for each j.

(15) For each of the following, answer true if the statement is always true. If the statement is only sometimes true or never true, answer false and provide a concrete example to illustrate the the statement is false. You should provide **justification** for your responses. Throughout, let X_1 and X_2 be topological spaces and $X = X_1 \times X_2$ with the product topology.

(a) If $A_1 \subseteq X_1$ and $A_2 \subseteq X_2$, then $(X_1 \times X_2) \setminus (A_1 \times A_2) = (X_1 \setminus A_1) \times (X_2 \setminus A_2)$.

(b) If $A_1 \subseteq X_1$ and $A_2 \subseteq X_2$, then $\text{Bdry}(A_1 \times A_2) \subseteq \text{Bdry}A_1 \times \text{Bdry}A_2$.

(c) If $A_1 \subseteq X_1$ and $A_2 \subseteq X_2$, then $\text{Bdry}A_1 \times \text{Bdry}A_2 \subseteq \text{Bdry}(A_1 \times A_2)$.

(d) If O_1 is an open subset of X_1 and O_2 is an open subset of X_2, then $O_1 \times O_2$ is an open subset of $X_1 \times X_2$.

(e) If O_1 is a subset of X_1 and O_2 is a subset of X_2 and $O_1 \times O_2$ is an open subset of $X_1 \times X_2$, then O_1 is an open subset of X_1 and O_2 is an open subset of X_2.

(f) If C_1 is a closed subset of X_1 and C_2 is a closed subset of X_2, then $C_1 \times C_2$ is a closed subset of $X_1 \times X_2$.

(g) If C_1 is a subset of X_1 and C_2 is a subset of X_2 and $C_1 \times C_2$ is a closed subset of $X_1 \times X_2$, then C_1 is a closed subset of X_1 and C_2 is a closed subset of X_2.

Index