Topology Through Inquiry

Michael Starbird and Francis Edward Su

November 21, 2018

Contents

Preface: four ways to use this book

Topology is an exciting **subject** to learn. Topological ideas surround us in daily life and in mathematical musings, but we can't fully enjoy their wonders until we learn topology. What makes a shoelace knotted? If point p and point q are part of a connected set, is there a path between them? Can we count to infinity and beyond? Can a surface have only one side or one edge? Topology is compelling because it can give us a new perspective and surprising answers to such questions.

Topology is an exciting subject to **learn**. Topology is not just an exciting subject to know—it is an exciting subject to *discover*.

One of the reasons that we, the authors, are writing this book comes from a personal source. For each of us, a topology class was the first setting in which we learned how to prove theorems on our own. Topology was the arena in which our personal relationship with mathematics was transformed. Proving theorems in topology was the experience that shifted our self-image from being purely consumers of mathematics to becoming *producers* of mathematics.

The presentation in this book invites readers, too, to feel the joy of discovering insights for themselves. They will not only learn some fascinating mathematics but will learn how to create insights and concepts by intent. We will preview a few of the delightful enticements of topology in Chapter One, but in this preface we describe our vision for how this book might be used by instructors and independent learners.

(1) A textbook for introductory topology: potential road maps

This book can be used as a textbook in an introductory topology class that is taught in an inquiry-based learning format. In this format, the instructor selects which exercises and theorems the students are to work on for the next sessions of the class. Then in class the students spend most of the time either in group work on unresolved theorems or in making and listening to presentations by students. The standing assignment for the students is to prove selected theorems and do selected exercises on their own, write up their own proofs and solutions, engage in group work during the classroom session, make presentations, and respond to presentations made by other students.

Sources about Inquiry Based Learning. There are many variations on how inquiry-oriented classes can be conducted. The accompanying *Instructors' Resource* describes some such methods and also includes recommendations for other sources of information and training about Inquiry Based Learning. The Instructor's Resource also includes several sample syllabi that describe in detail the daily running of the course.

Important note to instructors: It is not possible to do every theorem in any section during an introductory class, so only selected theorems and exercises should be assigned. The Instructor's Resource contains several possible threads through the material depending on instructor preferences.

Core General Topology: Here we list a collection of theorems and exercises that would provide students with the core ideas, examples, and theorems of point-set topology. Every chapter contains many additional concepts, examples, and theorems that an instructor may include. In each chapter, reading the introduction and conclusion section is recommended.

1. Cardinality—Chapter introduction; Section 2.1; Section 2.2 except Theorem 2.14 and Exercise 2.15; Theorem 2.16

2. Topological Space Fundamentals—Chapter introduction; Section 3.1; Section 3.2; Section 3.3.

3. Bases, Subspaces, Products—Section 4.1 through Exercise 4.5, then skip until Exercise 4.10 and skip rest of the section; Section 4.4; Section 4.5 through Exercise 4.35.

4. Separation Properties—Section 5.1except for Theorem 5.10-5.12 and 5.15; Section 5.2 can skip Exercise 5.18; Section 5.3 only 5.19, 5.20, 5.23.

5. Countable Features—Section 6.1 through Theorem 6.5; Section 6.2.

6. Compactness—Section 7.1 except Exercises 7.10-7.11; Section 7.2; Section 7.3 through 7.20.

7. Continuity and Homeomorphisms—Section 8.1; Section 8.2 skip 8.16-8.17; Section 8.3; Section 8.4 only Theorems 8.32, 8.34, 8.35, 8.36; Section 8.5 through 8.47.

8. Connectedness—Section 9.1 except 9.10 and 9.11.

9. Metric Spaces—Section 10.1 include 10.1, 10.3, 10.4, 10.6, 10.7, 10.8, 10.10, 10.11, 10.13, 10.18; in Section 10.2 do 10.19.

There are many possible variations for an introductory topology course. We sometimes include the Classification of 2-Manifolds as a part of the experience.

Classification of 2-Manifolds: Here is a path to give students an introduction to the classification of 2-manifolds after they have learned some general topology.

1. Introduction to Geometric and Algebraic Topology—Chapter 11 is an introduction to geometric topology. It is basically just a short reading assignment.

2. 2-Manifolds—Section 12.1; Section 12.2; Section 12.3; Section 12.4 or (12.6 and 12.7) (these are two different approaches); Section 12.5; 12.8; 12.9.

(2) Topology courses beyond an introductory course

This book contains far more material than could be covered in a single-semester course, and more than could be completely done in a year. So there are several alternatives for a second- or even third-semester course.

One possibility would be to treat point-set topology in the first semester and then do the more geometric and algebraic topology as a second-semester course.

Another possibility is for those who might have an interest in considering some of the more advanced topics in point-set topology. There is plenty of more advanced material in the point-set topology chapters so that an interesting second semester of point-set topology could be offered. Then the geometric and algebraic topology topics could be yet a third semester.

Geometric and Algebraic Topology: Here we list a collection of theorems and exercises that would provide students with the core ideas, examples, and theorems introducing geometric and algebraic topology. Every chapter contains many additional concepts, examples, and theorems that an instructor may include. In each chapter, reading the introduction and conclusion section would be good.

1. Start with the Classification of 2-Manifolds as described above.

2. Fundamental Group—Sections 13.1-13.4

3. Covering Spaces—Sections 14.1-14.4; Section 14.6

4. Manifolds and Complexes—Sections 15.1-15.4; (15.5 is a fun application)

5. $\mathbb{Z}_2$-homology—Sections 16.1-16.4

6. Applications of $\mathbb{Z}_2$-homology—Sections 17.1-17.4

7. Simplicial $\mathbb{Z}$-homology—Sections 18.1-18.4,18.7; (18.8 is quite cool)

8. Singular $\mathbb{Z}$-homology—19.1-19.5

(3) Independent Study Projects.

Another use for this book is as a source for many possible independent study projects. Many of the chapters include more advanced theorems than would be treated during a standard course, so many of those theorems or sections could be used as an independent study topic. Individual students or small groups working together could take a topic and work through the theorems and write a booklet describing their work. For example, the section on metrizability theorems or the section on the Cantor Set would be good topics for independent study. The student or group of students could be asked to come to grips with well-ordering, ordinal numbers, transfinite induction, and other concepts involved in the proofs of the basic metrization theorems. These concepts and techniques would form a satisfying, challenging collection of ideas that would be accessible to students during an independent study forum with help from the instructor. Many other such topics are available from this book including topics about issues around product spaces, continua, classification of 2-manifolds, and some algebraic topology among many others. Some of these possible independent study topics are described in the Instructor's Resource.

(4) Joyful Challenges for Independent Learners.

(This category includes those who may have skipped or given short shrift to point-set topology during their mathematical educations.) Yet another use of this book is simply for a person who wants to enjoy a rich collection of intriguing mathematical puzzles and challenges. Proving the theorems in this book can be an intrinsically rewarding and satisfying experience. So a person can simply take the whole book as a delightful collection of intellectual treats. Working on the proofs of the theorems can be a truly rewarding enterprise for those who enjoy thinking about challenges for their own sake. And these challenges have the added benefit that as you master them, you develop a robust understanding of a significant body of mathematics. Perhaps these theorems should be put on a theorem-a-day calendar or should appear in newspapers, where the harder theorems are suggested during the later days of the week.

A word about prerequisites.

Little preliminary mathematical knowledge is specifically required to undertake a study of this book; however, realistically speaking, a successful reader will probably need enough mathematical experience to be able to deal with mathematical abstraction. That mathematical maturity will be greatly enhanced while interacting with this book. At our schools, the topology class is generally offered as an upper-division undergraduate course that has the prerequisite of at least one proof-based course in abstract mathematics. Students would do well to have a basic grounding of elementary set theory such as understanding the concepts of sets, unions, intersections, and DeMorgan's Laws. For the sections on the fundamental group and homology groups, a basic introduction to group theory would be helpful background. The appendix we've provided summarizes the background from group theory that might be useful; for homology theory, a student would really only need the abelian material from the appendix.

Acknowledgments

We thank the Educational Advancement Foundation, the Initiative for Mathematics Learning by Inquiry, and Harry Lucas, Jr. for their generous support of the national project to foster effective mathematics instruction through inquiry. Their national support and their support of our work on this project have been profound. The Inquiry Based Learning Project has inspired us and has inspired many other faculty members and students across the country. The Educational Advancement Foundation, the Initiative for Mathematics Learning by Inquiry, the SIGMAA on Inquiry Based Learning, and other groups that promote effective mathematical instruction have a clear purpose of fostering methods of teaching that develop independent thinking and student creativity. We hope this project will contribute to making inquiry-based learning methods of instruction broadly available to many faculty members and students nationally.

We would like to thank everyone who has helped this project evolve and become more refined. Several individuals deserve special thanks for their considerable help with early drafts of this book. They include Cynthia Verjovsky and David Paige. Beyond considerable help with the manuscript, David also created many of the figures, for which we thank him heartily. Several faculty members have given us useful comments on drafts of this book. They include Joel Foisy, Caitlin Lienkaemper, Ben Lowenstein, Alex Cloud, Robert Bennett, Amzi Jeffs, Abram Sanderson, Elizabeth Kelley, Dagan Karp, Matt deLong, Christian Modjaiso, Jane Long, and Nicholas Scoville.

We would like especially to thank the many students who have used versions of these units in our classes. They helped us to make the presentation more effective and they energized us by rising to the challenges of personal responsibility, rigor, and curiosity. They inspired us to aim high in education—to realize that mathematics classes can genuinely help students to become better, more creative, more independent thinkers.

We would like to give personal thanks to our friends and family members whose support, encouragement, and insights suffuse this entire book and make all our projects possible and our lives fuller.

From Michael: I would especially like to acknowledge and thank Edward Burger for friendship, encouragement, and substantive contributions to this book. Although Ed is not a co-author of this book, many insights and perspectives that arose from work with Ed are prominently featured here. I would also like to thank my dear family, Roberta, Talley, and Bryn, whose love and support are priceless.

From Francis: I'm grateful to Mike Starbird for inspiring me as an undergraduate through an IBL topology class. It was that class that convinced me I could be a mathematician, and this book has its genesis in a course that Mike was teaching for many years. I also want to thank my wife Natalie for her constant encouragement.

Finally, we would like to express gratitude and appreciation to one another. An openness to suggestions and invariable good spirits have made this project a true joy.

And lastly, we would each like to add that any defects in this book are the other author's fault.

Chapter 1

Introduction: The Enchanting World of Topology

1.1 Enticements to Topology

The Möbius band and the Klein bottle may be the most famous objects of topology. These twisted surfaces are intriguing, thought-provoking, and beautiful—perfect descriptors of topology.

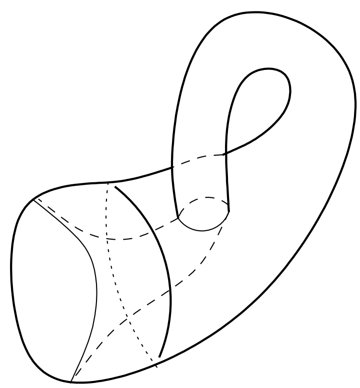


Figure 1.1: A Klein bottle.

But we've gotten ahead of ourselves. No introduction to topology should start without telling the most famous characterization of a topologist: A joke, among some admittedly nerdy people, is that a topologist is a person who can't tell the difference between a coffee cup and a doughnut. This apparent insult to topologists is really an observation that one branch of topology investigates when two absurdly elastic objects can be deformed continuously into one another—like elastic doughnuts and coffee cups can (try it). For this reason, topology enjoys the nickname "rubber sheet geometry." Actually, topology contains many facets besides rubber sheet geometry.

Perhaps the driving impulse behind topology is the urge to find the essence of familiar objects and concepts. The real line and continuous functions are part of familiar mathematics. When we isolate the fundamentals of those ideas, we create the subject of topology. Topology finds a way to isolate the most essential features that underlie ideas such as convergence, connectedness, continuity, and dimension.

One by-product of seeking essentials is that we discover incredible insights and entertaining examples that challenge our understanding of those fundamentals. For example, when we consider the basic idea of counting and extend that idea to infinity, we learn that infinity itself comes in more than one size. Or when thinking about connectedness, consider the graph of the function $f(x) = \sin(1/x)$ for $0 < x \leq 1$ together with the interval on the y-axis between -1 and 1. Should that set be considered connected or not connected?

The iconic Möbius band and the Klein bottle invite our curiosity to play. We can imagine being a 2-dimensional bug whose universe is a Möbius band while our cousin lives on the surface of a doughnut or a sphere. And immediately a cascade of questions delight us. What features are the same in our world and our cousin's? How can we tell the difference between our worlds our immediate neighborhoods look the same?

It may appear that such questions are trivial or inconsequential; however, the history of science has taught us repeatedly that explorations of abstract, mathematical concepts continually surprise us with their applicability to the real world, and topology is no exception.

The concepts of topology—including insights garnered from conceiving of a universe of unrealistically elastic material—have found applications from cosmology to biology. Descriptions of the potential shape and structure of our entire universe depend on topological ideas—including ideas arising from trying to distinguish a Möbius band from the surface of a sphere. At the other end of the size spectrum, topology provides insight into string theory, formulated to describe the most fundamental building blocks of the physical universe. And the structure and behavior of twisted DNA molecules can be studied using the topological concepts of knots and links.

Topology is a subject whose power arises from the impulse to abstract essential features from complex situations and then to let our curiosity roam while striving to truly understand what is essential about fundamental ideas.

The whole of topology arises from employing various learnable strategies of discovery and exploration. In the chapters to come, we will point out these methods of creativity and you will see how new mathematical ideas emerge.

1.2 Learning to Create Mathematics

> People are generally better persuaded by the reasons which they have themselves discovered than by those which have come in to the mind of others.
>
> —Blaise Pascal, *Pensees* #10

Every student spends the vast majority of the first years of mathematics instruction believing that mathematics is a subject that some brilliant dead people in the past created and that now we learn. It was during an inquiry-based-learning topology class that the authors' eyes were opened to the idea that mathematics is something that human beings create by virtue of thought that is within the reach of regular people—including ourselves—and not just specially endowed geniuses. So the theorems that appear in this book have played a special role in our development as mathematicians and, more generally, as people who seek to be effective thinkers in all areas of our lives.

Educational researchers now unanimously emphasize the importance of active struggle as a central experience for meaningful learning and for making the learning have a permanent effect. There are many inquiry-based learning methods of instruction; however, the fundamental ingredient is the students' engagement in doing mathematics for themselves.

So it is a special honor for us to be able to present the topological challenges in this book as opportunities for students or others to enjoy the process of discovering wonderful mathematical proofs and insights. Each theorem statement or exercise in this book is a puzzle to be thought through and added to an ever-growing toolbox of insights and techniques—insights and techniques both about topology and about how to think in general. The topological theorems in this book were among the first occasions when we personally proved theorems on our own. For us these challenges were a candy jar filled with delectable treats, and we hope that you too find many hours of pleasure from grappling with these intriguing ideas.

This book contains essentially no proofs of theorems. Instead, it presents theorem statements in an order designed to guide you to discover topology on your own.

Each of the remaining chapters is devoted to a topic in topology. After a brief introduction, the remaining sections of each chapter consist of examples, definitions, exercises, and theorem statements. Those sections are where the fun lies, because your challenge is to devise proofs of all the theorems and answers to all the exercises. In proving the theorems, you will be creating and exploring the wonderful mathematical techniques and ideas that are the essence of the subject.

The examples, exercises, and theorem statements are ordered in such a way that each theorem can be proved by the diligent reader. Of course, all the theorems in topology were originally proved by people who had never seen the ideas before. This book presents the statements in such a way that you will have the genuine experience of discovering the insights that make the theorems true.

Many of the theorems are difficult to prove or require an insight that may not occur to you right away. After you have gained additional experience, you will often look back on a theorem and see it as far more meaningful, clearer, and easier than it appeared at first. But there is no royal road to understanding. The struggle is where the learning comes—and where the enjoyment and satisfaction are.

No reader should expect to successfully prove all the theorems, and that is fine. Some theorems may require weeks of effort that is rewarded with an 'aha!' moment of joyful insight and resolution. Some theorems may seem impossible and you may reluctantly eventually feel the need to seek other sources to find a proof. However, having toiled on a theorem, you will find that the insights you encounter in a proof will be far more significant to you even if you do not succeed in discovering the proofs for yourselves.

Working with the concepts; grappling with the fine details of the definitions, examples, and theorems; and struggling to find the right way to look at an unfamiliar world is the experience that this book offers. The strategy of persevering through difficulties and enjoying the struggle as well as the triumphs makes the experience of this book meaningful. That process makes the mathematics come to life, but it also awakens many people to an extremely satisfying method for dealing with the whole range of unknown situations that arise constantly in all areas of life. Persistence, self-confidence, and skill at coming up with creative ideas are among the lessons that this experience helps to develop.

1.3 Introduction to Set Theoretic Topology

Our journey through topology will begin with the part of topology referred to as point-set topology or set theoretic topology. It arose from an impulse to find the essence of mathematics.

At the end of the nineteenth century, mathematicians embarked on a program whose aim was to axiomatize all of mathematics. The goal was to emulate the format of Euclidean geometry in the sense of explicitly stating a collection of definitions and unproved axioms and then proving all mathematical theorems from those definitions and axioms. The foundation on which this program rested was the concept of a set. Axioms for set theory were proposed and then the goal was to cast known mathematical theorems in set-theoretic terms. So the challenge for mathematicians was to take familiar objects, such as the real line, and familiar concepts, such as continuity and convergence, and recast them in terms of sets. From this effort arose the concept of a topological space, and the field of topology was born.

The first half of this book is an introduction to point-set or set theoretic topology. It begins with a chapter on Cardinality—the concept that extends to infinity the basic concept of counting. What should the analogy to simple counting be when you are trying to compare the sizes of sets that are infinite. The study of infinite sets could well be regarded as one of the triumphs of human thought. The concept of cardinality provides a foundational part of the basis of topology.

We next create the definition of a topological space by looking at familiar mathematical objects like the real line and plane with an eye toward finding out what is fundamental about certain subsets of those spaces. By abstracting features of sets that are important in the familiar definitions of continuity and convergence, we are led to the definition of a topological space. The definition of a topological space opens the door for the explorations that make up the rest of this book.

After defining a mathematical object that is as fundamental as a topological space, the next steps are to see the consequences of the definition. One thread of exploration involves creating examples of topological spaces that illustrate the range of possibilities for topological spaces. Another thread involves creating concepts that capture distinctions among features of spaces. These distinctions create categories of concepts.

One sequence of properties are called separation properties because they capture ideas concerning what types of subsets can be separated from one another by putting them in disjoint special sets described by the topology. Another collection of properties refer back to our concept of cardinality. This collection of properties explores the implications of having some features of a topological space being countable or uncountable—words describing cardinality. Another category of properties are the covering properties such as compactness, which turns out to capture many important features of spaces and maps.

As we create these properties of topological spaces and investigate their consequences, we find that the exploration becomes increasingly interesting and nuanced as we proceed. Part of the reason for the ever-increasing interest is that as we become aware of more features of topological spaces, we encounter ever-increasing numbers of potential interactions among those properties. So we find that covering properties together with countability properties have implications about separation properties, for example.

One of the consequences of seeking essentials is that the familiar objects and concepts that generated the ideas to begin with, later become increasingly fascinating as we come to appreciate their topological connections. After learning about the topological view of convergence and continuity, the definition of those ideas that we may have learned in calculus or analysis will become clearer and more meaningful. After learning about the topological distinctions of connectedness, we see that being connected has far more nuance and interest than it did before we explored the idea topologically. The idea of distance takes on a new meaning after we define the topological idea of a metric. And then we are led to see how the purely topological ideas that we developed in the earlier chapters are related to the property of having a metric, that is, a concept of distance in a space. Characterizing those topological spaces that are metrizable is appropriately one of the objectives in making the connection between set theoretic topology and the concept of metric that seems so fundamental to ideas such as Euclidean spaces.

The more we learn, the richer the world of topology becomes. We hope you enjoy the journey.

Part I

Point-Set Topology

Chapter 2

Cardinality: To Infinity and Beyond

We begin our exploration of set-theoretic topology by starting with perhaps the most basic mathematical idea—counting—and finding a way to generalize that notion to apply to infinite sets in a reasonable way. This exploration is fascinating and contains some of the most counterintuitive ideas in mathematics.

On the one hand, you will love to think through these ideas and relish their surprises. On the other hand, we do not want you to get so engrossed with these cardinality results that you delay your enjoyment of the topology to come. So, you may wish to read the definitions and theorems in those sections now, resolve to come back to them later, and begin proving theorems in the next chapter, where we introduce the idea of a *topology*.

2.1 Sets and Functions

The first accomplishment of the late 19th century program to axiomatize mathematics was to redefine the known fields such as calculus and analysis in terms of sets and functions between sets. Most of mathematics can now be viewed as the study of some collection of sets, usually with some structure added to them, along with a corresponding set of functions between those sets. Here we outline the basic concepts about sets and functions we will need to get started.

Definition. If A is a set and a is an element of A, we write $a \in A$.

Definition. Let A and B be sets. The set A is a **subset** of B if and only if every element of A is an element of B, and we write $A \subset B$. Note that we will always use the notation $\subset$ and not $\subseteq$. In particular, A is a subset of itself, i.e., $A \subset A$. We write $A = B$ if the elements of A and B are identical.

One common strategy to establish that $A = B$ is to show that both $A \subset B$ and $B \subset A$.

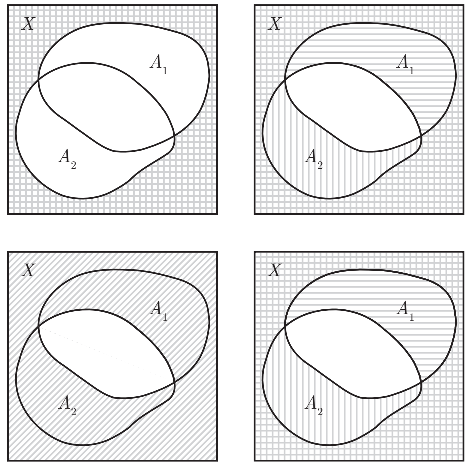


Figure 2.1: Sets involved in DeMorgan's Laws for two sets. The top two figures show that the complement of the union is the intersection of the complements. The bottom two figures show that the complement of the intersection is the union of the complements.

Definition. The **union** of A and B is the set $A \cup B$ of elements that are elements in either A or B. Thus $x \in A \cup B$ if and only if $x \in A$ or $x \in B$.

Definition. Let A and B be sets. The **intersection** of A and B is the set $A \cap B$ of elements that are elements in both A and B. Thus $x \in A \cap B$ if and only if $x \in A$ and $x \in B$.

Definition. Let $A \subset B$. The **complement** of A in B is the set of all elements that are in B but not in A, and is denoted $B - A$. Thus $x \in B - A$ if and only if $x \in B$ and $x \notin A$.

You may wish to try this exercise to see if the above definitions make sense.

Exercise 2.1. *For sets $A_1, A_2 \subset X$ show that*

$$X - (A_1 \cup A_2) = (X - A_1) \cap (X - A_2).$$

We can define the intersection and the union of collections of sets.

Definition. Let $A_1, A_2, ..., A_N$ be a collection of sets. Their **union** $A_1 \cup A_2 \cup \cdots \cup A_N$, also written $\bigcup_{i=1}^N A_i$, is the set

$$\{x \mid x \in A_i \text{ for at least one } i \in \{1, 2, ..., N\}\}.$$

Their **intersection** $A_1 \cap A_2 \cap \cdots \cap A_N$, also written $\bigcap_{i=1}^N A_i$, is the set

$$\{x \mid x \in A_i \text{ for every } i \in \{1, 2, ..., N\}\}.$$

The next theorem, known as DeMorgan's Laws, generalizes our first exercise and elegantly describes how unions, intersections, and complements are related.

Theorem 2.2. *(DeMorgan's Laws) Let X be a set, and let $\{A_k\}_{k=1}^N$ be a finite collection of sets such that $A_k \subset X$ for each $k = 1, 2, \ldots, N$. Then*

$$X - \left(\bigcup_{k=1}^{N} A_k \right) = \bigcap_{k=1}^{N} (X - A_k)$$

and

$$X - \left(\bigcap_{k=1}^{N} A_k \right) = \bigcup_{k=1}^{N} (X - A_k).$$

See Figure 2.1. The notion of union and intersection and DeMorgan's Laws can be established not just for finite collections of sets, but for infinite collection of sets as well. After learning the definition of the word *infinite*, you may wish to return and prove that DeMorgan's Laws hold for arbitrary infinite collections of sets.

Next, we introduce the notion of a function. There is a formal definition involving Cartesian products that we will encounter in a later chapter, but for now the following definition will suffice.

Definition. A **function** or **map** from a set X to a set Y, written $f : X \to Y$, assigns to each element of x in X an element $f(x)$ in Y. The set X is called the **domain** of f and the set Y is called the **codomain** of f. For a subset $A \subset X$, the **image** of A under f is the set

$$f(A) = \{f(a) \in Y \mid a \in A\}.$$

If $B \subset Y$, the **preimage** or **inverse image** of B under f is the set

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

Definition. A function $f : X \to Y$ is called an **injection** or an **injective function** or a **one-to-one function** if and only if $f(a) = f(b)$ implies $a = b$.

The function f is a **surjection** or a **surjective function** or an **onto function** if and only if for all $y \in Y$ there is an $x \in X$ such that $f(x) = y$.

A function that is both an injection and a surjection is a **bijection** or a **bijective function** or a **one-to-one correspondence**. (A one-to-one correspondence should not be confused with a one-to-one function which may lack surjectivity.)

Try these exercises to test your understanding of these concepts.

Exercise 2.3. *For a function $f : X \to Y$, and sets $A, B \subset Y$, show that $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$ and $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.*

Exercise 2.4. *If $f : X \to Y$ is injective and $y \in Y$, then $f^{-1}(y)$ contains at most one point.*

Exercise 2.5. *If $f : X \to Y$ is surjective and $y \in Y$, then $f^{-1}(y)$ contains at least one point.*

2.2 Cardinality and Countable Sets

A fundamental counting question is: "how many" elements are in a given set? Since our days as toddlers, we learned to signify the "size" of a non-empty set S by associating a natural number to it using a counting procedure. Counting is a ritual in which we point to successive elements of S and count out loud "$1, 2, 3, \ldots$" until all the elements of S have been exhausted; then we declare that the size of S is the very last number called.

However, there are sets for which this procedure will never terminate, because there is no last element. Such sets are called *infinite* and these present a problem for counting.

The way out is to realize that our counting procedure, when it works, is producing a one-to-one correspondence between the elements of S and the elements of $\{1, 2, 3, \ldots, n\}$ for some natural number n. This idea forms the basis of how we might think about the size of an infinite set S. We can focus on describing what it means for two sets to have the same size. It is reasonable to say that two sets S and T have the same size if there is a way to pair up the elements of S with the elements of T by a one-to-one correspondence.

Since the word "size" is a bit casual, we define a new word, "cardinality", which captures a natural criterion for asserting that two sets have the same size.

Definition. Two sets, A and B, have the same **cardinality** if and only if there exists a bijection $f : A \to B$. The cardinality of a set A is denoted $|A|$.

Notice that we didn't really define the cardinality of A, but nevertheless the notation $|A|$ is helpful, because now we can write such statements as $|A| = |B|$. Formally, $|A|$ could be thought of as the equivalence class of all sets with the same cardinality as A, but that definition is confusingly abstract, though correct. For example, we could define the number 5 to be the equivalence class of all sets that can be put into one-to-one correspondence with the set $\{1, 2, 3, 4, 5\}$, but most toddlers would be puzzled by that introduction to counting.

Fortunately, we do define the cardinality of a finite set as the number of elements in that set.

Definition. A set X is **finite** if and only if it is empty or there is a one-to-one correspondence $f : X \to \{1, 2, \ldots, n\}$ where n is an element of $\mathbb{N}$, in which case we say X **has cardinality** n and write $|X| = n$. A set that is not finite is **infinite**.

These first theorems will get you accustomed to the definition of cardinality by asking you to demonstrate some one-to-one correspondences between sets. It is not necessary to write down a formula for the correspondence. If you present a pattern that indicates an unambiguous way to correspond the elements of one set with the elements of another set, then you have shown the existence of a one-to-one correspondence.

Some of the most familiar infinite sets are the natural numbers, the integers, the rational numbers, and the real numbers.

Definition. Throughout this book we will use the following notation:

- $\mathbb{N}$ denotes the set of **natural numbers**: $\{1, 2, 3, \ldots\}$, i.e., the positive integers.
- $\mathbb{Z}$ denotes the set of all **integers**: $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
- $\mathbb{Q}$ denotes the set of **rational numbers**: $\{\frac{m}{n} \in \mathbb{R} \mid m, n \in \mathbb{Z}, n \neq 0\}$.
- $\mathbb{R}$ denotes the set of **real numbers**, that is, the set of all decimal numbers.

The first theorem asks you to prove that the even positive integers have the same cardinality as $\mathbb{N}$, the natural numbers.

Theorem 2.6. *Let* $2\mathbb{N}$ *denote the even positive integers* $\{2, 4, 6, \ldots\}$*. Then* $2\mathbb{N}$ *has the same cardinality as* $\mathbb{N}$*, that is,* $|2\mathbb{N}| = |\mathbb{N}|$.

Next you will prove that the set of all integers—positive, negative, and zero—has the same cardinality as the natural numbers.

Theorem 2.7. *The set* $\mathbb{Z}$ *has the same cardinality as* $\mathbb{N}$*, that is,* $|\mathbb{Z}| = |\mathbb{N}|$.

A basic fact about $\mathbb{N}$ is that with the usual $<$ ordering, every non-empty subset of $\mathbb{N}$ has a least element. This fact is useful in proving the following theorem.

Theorem 2.8. *Every subset of* $\mathbb{N}$ *is either finite or has the same cardinality as* $\mathbb{N}$.

Definition. A set is **countable** if and only if it is finite or has the same cardinality as $\mathbb{N}$. A set is **uncountable** if and only if it is not countable.

So, a countable set is either finite or has the same cardinality as $\mathbb{N}$. Notice that the elements of a countable set can be viewed as a sequence indexed by natural numbers. The next theorem shows that the set of natural numbers is in some sense the smallest infinite set.

Theorem 2.9. *Every infinite set has a countably infinite subset.*

Any injection from a finite set to itself is automatically a surjection. However, such is not the case with infinite sets.

Theorem 2.10. *A set is infinite if and only if there is an injection from the set into a proper subset of itself.*

The union of finitely many countable sets or even countably many countable sets is still a countable set.

Theorem 2.11. *The union of two countable sets is countable.*

Theorem 2.12. *The union of countably many countable sets is countable.*

There are infinitely many rational numbers between every two integers; nevertheless, the set of rational numbers has the same cardinality as the set of natural numbers.

Theorem 2.13. *The set $\mathbb{Q}$ is countable.*

There are often more ways than one to prove a theorem. So see if you can find several different proofs of the next result. After you have found a proof or two, you might look at the comments after the theorem that are suggestions leading to three different approaches.

Theorem 2.14. *The set of all finite subsets of a countable set is countable.*

> **Effective Thinking Principle.** *Seek Different Views of the Same Result.* Understanding a theorem often resides in the proof rather than the statement. So alternative proofs provide additional insight—look for them.

Here are some possible approaches to creating several different proofs of the preceding theorem:

1. Stratify the set of all subsets by size.
2. Suppose your countable set is the set of primes.
3. Suppose your countable set is the natural numbers. Concatenating the elements of each finite set does not quite work. Can you make it work by somehow changing the commas into digits?

The following exercise would be an example of a practical application of these ideas if the ocean were an infinitely large plane and we had an infinite amount of time to work.

> **Effective Thinking Principle.** *Do Special Cases.* By doing special cases, particularly simple cases, you teach yourself how to do harder problems.

Exercise 2.15. *Suppose a submarine is moving in the plane along a straight line at a constant speed such that at each hour, the submarine is at a lattice point, that is, a point whose two coordinates are both integers. Suppose at each hour you can explode one depth charge at a lattice point that will hit the submarine if it is there. You do not know the submarine's direction, speed, or its current position. Prove that you can explode one depth charge each hour in such a way that you will be guaranteed to eventually hit the submarine.*

2.3 Uncountable Sets and Power Sets

So far, all the sets we have encountered have been countable: the integers, the rationals, the countable union of countable sets, and the set of finite subsets of a countable set. We might think that every infinite set has the same cardinality as any other infinite set. After all, once we have gotten to infinity, that might well be as far as we could hope to go. So now we begin the process of looking for sets that may have a larger cardinality than the set of natural numbers. The set of real numbers, that is, the set of all decimal numbers, is a natural set to investigate.

If there were a one-to-one correspondence between the natural numbers and the real numbers, then it would be possible to write the numbers $1, 2, 3, 4, 5, \ldots$ in a column and next to each of those natural numbers we could write a decimal number in such a way that every decimal number would eventually appear on the list. So here is a challenge for you to think about. Suppose we handed you a long piece of paper that had $1, 2, 3, 4, 5, \ldots$ in one column with a decimal number next to each of those numbers. Can you devise a process by which you could write down a decimal number that is provably not anywhere on this infinitely long list? In other words, can you describe a procedure to write down a single decimal number that is not equal to the decimal number next to 1, that is not equal to the decimal number next to 2, that is not equal to the decimal number next to 3, and so on forever? If so, you are well on your way to proving one of the most famous results in mathematics: Cantor's Theorem.

Theorem 2.16 (Cantor's Theorem)**.** *The cardinality of the set of natural numbers is not the same as the cardinality of the set of real numbers. That is, the set of real numbers is uncountable.*

You have now proved conclusively that infinite sets come in more than one size!! What an amazing and counterintuitive insight.

Cantor's Theorem has opened the door to the idea that infinite sets come in different sizes, so we can start exploring ways to create infinite sets with different cardinalities. Power sets play a central role in this exploration.

Definition. For any set A, the set of all subsets of A is called the **power set** of A and is denoted 2^A.

Recall that the empty set, denoted $\varnothing$, is a subset of any set and hence is always an element of a power set.

Exercise 2.17. *Suppose $A = \{a, b, c\}$. Explicitly write out 2^A, the power set of A.*

The following theorem justifies the 2^A notation for the power set of A.

Theorem 2.18. *If a set A is finite, then the power set of A has cardinality $2^{|A|}$, that is, $|2^A| = 2^{|A|}$.*

The following easy theorem shows that every set has the same cardinality as a subset of its power set. Later we will see that the power set of any set has a cardinality that is strictly greater than the cardinality of the set itself. At this point we have not yet even defined the idea of cardinalities being greater or less than one another, but we will do so soon.

Theorem 2.19. *For any set A, there is an injection from A into 2^A.*

One way to think about the power set of a set is to think about a set of functions into the two-point set $\{0, 1\}$. An example of a function from the set $A = \{a, b, c\}$ to the set $\{0, 1\}$ is the function f defined by: $f(a) = 1$, $f(b) = 0$, and $f(c) = 1$.

Theorem 2.20. *For a set A, let P be the set of all functions from A to the two point set $\{0, 1\}$. Then $|P| = |2^A|$.*

In the case of the natural numbers $\mathbb{N}$, we can make each subset of $\mathbb{N}$ correspond in a natural way with a sequence of 0's and 1's.

Theorem 2.21. *There is a one-to-one correspondence between $2^{\mathbb{N}}$ and the set of all infinite sequences of 0's and 1's.*

Thinking of the power set in these ways may be helpful in proving the following theorem, whose proof is similar to that of Cantor's Theorem above.

Theorem 2.22 (Cantor's Power Set Theorem)**.** *There is no surjection from a set A onto 2^A. Thus for any set A, the cardinality of A is not the same as the cardinality of its power set. In other words, $|A| \neq |2^A|$.*

2.4 The Schroeder-Bernstein Theorem

To show that two sets A and B have the same cardinality, the fundamental challenge is to produce a bijection from one set to the other. In many cases, it may be easier to produce two injections: one from set A to B and one from set B to A. These two injections give us the sense that A is smaller than B and B is smaller than A, so morally speaking, they should be the same size. The Schroeder-Bernstein theorem states that this intuition is justified—we will be able to use these injections to construct a bijection between A and B.

The next exercises gives you practice in using two injections to produce a single bijection. You will be developing the insights to allow you to prove the Schroeder-Bernstein theorem.

Exercise 2.23. *Consider $A = [0,1]$ and $B = [0,1)$ and injections $f(x) = x/3$ from A to B and $g(x) = x$ from B to A. Construct a bijection h from A to B such that on some points of A, $h(x) = f(x)$, and for the other points of A, $h(x) = g^{-1}(x)$.*

You may find it helpful to draw A and B as parallel vertical lines so that you can track where various sets go under f and g^{-1}. Can you identify on what subsets of A you are required to use f and on what subsets you are required to use g^{-1}?

Exercise 2.24. *Consider $A = [0,1]$ and $B = [0,1)$ and injections $f(x) = x/3$ from A to B and $g(x) = x/2$ from B to A. Construct a bijection h from A to B such that on some points of A, $h(x) = f(x)$, and for the other points of A, $h(x) = g^{-1}(x)$.*

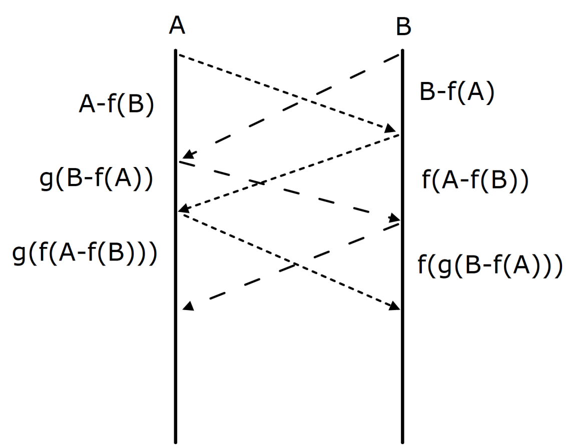


Figure 2.2: Examples that may help you visualize a proof of the Schroeder-Bernstein theorem.

As you work to prove the Schroeder-Bernstein Theorem below, you may find it helpful to draw the sets A and B in the hypothesis as parallel vertical lines and picture the functions f and g to be like the functions f and g in the preceding exercise so that you can keep track of and label where various sets go under f and g^{-1}. Try to identify and label the subsets of A on which you are required to use f and identify and label the subsets of A on which you are required to use g^{-1}.

Theorem 2.25 (Schroeder-Bernstein)**.** *If A and B are sets such that there exist injections f from A into B and g from B into A, then $|A| = |B|$.*

The Schroeder-Bernstein Theorem helps us to order cardinalities.

Definition. The cardinality of A is **less than or equal** to the cardinality of B, written $|A| \leq |B|$, if and only if there is an injective function from A into B. And the cardinality of A is **less than** the cardinality of B, written $|A| < |B|$ if and only if $|A| \leq |B|$, but B is not less than or equal to A, that is, there is no injective function from B into A.

The Schroeder-Bernstein Theorem justifies the notation "$\leq$", since it implies that if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$, as we would expect.

We could reformulate the Schroeder-Bernstein Theorem using onto functions.

Theorem 2.26 (Schroeder-Bernstein)**.** *If A and B are sets such that there exist a surjective function $f : A \to B$ and a surjective function $g : B \to A$, then $|A| = |B|$.*

The principal remaining issue about ordering cardinalities is the possibility that some pair of sets may not be comparable. That is, could there be two sets A and B such that no injection from A to B exists and no injection from B to A exists? It turns out that the Well-Ordering Principle, which we will discuss in the next section, implies that every pair of sets is comparable, and therefore that cardinalities are ordered.

For now, let's return to some other questions about cardinality of sets that we can settle using the Schroeder-Bernstein Theorem. This next theorem tells us that intervals have the same cardinality as the whole real line.

Theorem 2.27. $|\mathbb{R}| = |(0, 1)| = |[0, 1]|$.

Going up in dimension does not raise the cardinality; the unit square and the unit interval have the same cardinality.

Theorem 2.28. *Let $[0, 1] \times [0, 1]$ denote the Cartesian product of two closed unit intervals. Then*

$$|[0, 1] \times [0, 1]| = |[0, 1]|.$$

Cantor's Power Set Theorem (2.22) tells us that there are sets whose cardinality is larger than the cardinality of $\mathbb{R}$. A specific set whose cardinality is greater than that of the reals is the set of all functions from the reals to the reals. The following theorem states that the set of all real-valued functions has the same cardinality as the power set of $\mathbb{R}$.

Theorem 2.29. *The set of all functions $f : \mathbb{R} \to \mathbb{R}$ has the same cardinality as $2^{\mathbb{R}}$.*

By the way, if we restrict our attention to the smaller class of *continuous* functions from $\mathbb{R}$ to $\mathbb{R}$, then this set has the same cardinality as $\mathbb{R}$ (Theorem 8.8). But it would be better to work on that later after you have proved some theorems about continuous functions.

Cantor's Theorem (2.16) showed us that the cardinality of the reals is greater than the cardinality of the natural numbers. The next theorem states that the cardinality of the reals is the same as the cardinality of the power set of the natural numbers. Therefore, we see that Cantor's Theorem about the uncountability of $\mathbb{R}$ is actually a special case of Cantor's Power Set Theorem (2.22).

Theorem 2.30. $|\mathbb{R}| = |2^{\mathbb{N}}|$.

Cantor's Power Set Theorem implies that there are infinitely many different sizes of infinity.

Theorem 2.31. *There are infinitely many different infinite cardinalities.*

We end this section with some truly bizarre issues about cardinality. We know that the cardinality of $\mathbb{R}$ is uncountable. But this fact leaves open the question of whether the cardinality of $\mathbb{R}$ is the next bigger cardinality above the cardinality of $\mathbb{N}$. Perhaps there is some uncountable set whose cardinality is strictly less than the cardinality of $\mathbb{R}$. Cantor believed that there was no such intermediate set and tried in vain to prove it; this assertion came be known as the Continuum Hypothesis:

Continuum Hypothesis. *There is no uncountable set whose cardinality is greater than the cardinality of* $\mathbb{N}$ *yet less than the cardinality of* $\mathbb{R}$.

But in 1940, Kurt Gödel proved that the Continuum Hypothesis is consistent with the standard axioms of set theory (the Zermelo-Fraenkel axioms)—that is, no contradiction would arise if the Continuum Hypothesis were added as a new axiom. And in the 1960's, Paul Cohen proved that the negation of the Continuum Hypothesis is also consistent with the Zermelo-Fraenkel axioms. Together, this means the Continuum Hypothesis is independent of the Zermelo-Fraenkel Axioms of Set Theory—it *provably* cannot be proved or disproved with the usual axioms of set theory.

As you see, the study of infinite cardinalities presents us with some truly surreal surprises.

2.5 The Axiom of Choice

Dealing with infinite sets presents us with some truly weird phenomena. One of the weirdest is called the Banach-Tarski Paradox. It asserts that it is possible to take a solid unit ball, divide it into a finite number of subsets, and then rigidly move those subsets to create two solid unit balls. Of course, such a thing sounds ridiculous, and we do not recommend that you try it at home.

The subsets are not at all like pieces of pie. They are certainly not measurable sets. This apparent impossibility becomes even stranger when we learn that it is equivalent to a rather bland sounding statement, namely, if we have a set of non-empty sets, then we can create a set that contains an element from each of those sets. That assertion is called the Axiom of Choice. Who would have guessed that such an innocuous statement could imply that you can take a single solid ball, break it into a few pieces, and reassemble them to create two solid balls of the same size.

As we learn more about the Axiom of Choice and the Banach-Tarski Paradox, we come to feel that the Banach-Tarski Paradox is not really as surprising as it first appears. Like the fact of different sizes of infinity, once we understand the arguments, a new normalcy occurs. We will not explore the Banach-Tarski Paradox here, because that would lead us too far astray. Instead, this section will introduce you to three important and equivalent statements in set theory: Zorn's Lemma, the Axiom of Choice, and the Well-Ordering Principle. They are accepted as fundamental axioms and are used freely in most standard mathematics. We will use them in this book.

If you have only worked with finite sets, these statements will not seem that interesting to you. Their power comes from applying the statements to infinite sets. In this section we give the relevant definitions and then state Zorn's Lemma, the Axiom of Choice, and the Well-Ordering Principle. You may just wish to familiarize yourself with these statements, and not prove the theorems in this section right now. Feel free to use these statements as needed in the rest of the book.

Definition. A set X is **partially ordered** by the relation $\leq$ if and only if, for any elements x, y, and z in X,

1. $x \leq x$,
2. if $x \leq y$ and $y \leq z$, then $x \leq z$,
3. if $x \leq y$ and $y \leq x$, then $x = y$.

Note that two elements in a partially ordered set may not be comparable. A partially ordered set is sometimes called a **poset**. We write $x < y$ to signify that $x \leq y$ but $x \neq y$.

Example. For any set X, any collection of subsets of X is partially ordered under the relation of set inclusion $\subset$.

Definition. Let X be a poset with relation $\leq$. An element a in X is a **least element** if and only if for any $x \in X$ with $x \leq a$, it must be the case that $x = a$. An element m in X is a **maximal element** if and only if for any x in X with $m \leq x$, it must be that $m = x$.

Exercise 2.32. *Given a set X, consider the poset P of all subsets of X partially ordered by inclusion. Show that X is the unique maximal element of P, and show that the empty set is the unique least element of P.*

Exercise 2.33. *Construct an example of a poset with several maximal elements and several least elements.*

Definition. A poset is **totally ordered** if and only if it is partially ordered and every two elements are comparable (that is, for all x and y, either $x \leq y$ or $y \leq x$).

Example. The real line $\mathbb{R}$ with the $\leq$ relation is totally ordered.

Definition. A set is **well-ordered** if and only if it is totally ordered and every non-empty subset has a least element.

Example. The set of natural numbers $\mathbb{N}$ is well-ordered.

Exercise 2.34. *Show that $\mathbb{R}$ with the $\leq$ relation is totally ordered but not well-ordered.*

The next three statements—Zorn's Lemma, the Axiom of Choice, and the Well-Ordering Principle—are equivalent to each other. Each is regarded as an *axiom*: a statement that seems so obvious we simply accept it as true. These statements are generally accepted and used freely in most branches of mathematics. They are known to be independent of the usual (Zermelo-Fraenkel) axioms of set theory, so it would be possible to investigate strange collections of axioms for set theory that do not include them, but we will leave such exploration for interested set theorists.

Definition. Let P be a poset with relation $\leq$ and let A be a subset of P. An element b in P is an **upper bound** of A if and only if for every $a \in A$, $a \leq b$.

Zorn's Lemma. *Let X be a partially ordered set in which each totally ordered subset has an upper bound in X. Then X has a maximal element.*

The Axiom of Choice states that we can create a new set from a collection of sets in a certain way. We will denote an arbitrary collection of sets like this: $\{A_\alpha\}_{\alpha \in \lambda}$. This notation is a fancy way of expressing the idea that a collection can be indexed. For instance, the set $\{A_1, A_2, A_3\}$ has three elements, each of which is a set, so we could also write it as $\{A_i\}_{i \in \{1,2,3\}}$. Now, an arbitrary collection might be a very large infinite set, possibly uncountable. So we think of λ as a potentially very large index set and α as an index. In deference to mathematical tradition, we avoid using i as the index if λ could be uncountable.

Now we can state the Axiom of Choice:

Axiom of Choice. *Let $\{A_\alpha\}_{\alpha \in \lambda}$ be a set of non-empty sets. Then there is a function $f : \lambda \to \bigcup_{\alpha \in \lambda} A_\alpha$ such that for each α in λ, $f(\alpha)$ is an element of A_α.*

This function "chooses" one element from each set. Said another way, the Axiom of Choice allows one to construct a set that contains one element from each of the sets in a given set of non-empty sets. This Axiom of Choice seems like an obvious way to construct a new set, but, as we pointed out it does lead to some surprising consequences, such as the *Banach-Tarski Paradox*.

Well-Ordering Principle. *Every set can be well-ordered. That is, every set can be put in one-to-one correspondence with a well-ordered set.*

To see how surprising this principle is, try constructing a well-ordering of an uncountable set like $\mathbb{R}$. You will have a hard time. And yet this principle says that we can assume there is a well-ordering, even if we can't find one! You'll be forgiven if this statement doesn't seem quite as obvious as the Axiom of Choice and Zorn's Lemma. But since it is equivalent to the others, it must be equally compelling. The proof of the equivalence of these three statements is challenging; as a start you might want to try to prove one or more of the six directed implications.

Theorem 2.35. *Zorn's Lemma, the Axiom of Choice, and the Well-Ordering Principle are equivalent.*

In any case, let's just feel free to use any one of these statements whenever they are useful.

2.6 Ordinal numbers

Ordinal numbers are a fascinating topic, because they answer the question "is it possible to count an uncountable set one element at a time?" The ordinal numbers give us a rich collection of interesting examples that we can use later. However, this whole section can be skipped on first reading.

> **Effective Thinking Principle.** *Understand Simple Things Deeply.* Understanding basic ideas with great depth often leads to great insights.

One of the first things we learn to do as toddlers is to count. But what is counting? Let us try to answer this basic question and see where it leads us.

Suppose a toddler has carefully lined up five toys. There are really two features of counting them. One is to say, "one, two, three, four, five." The other related way is to say, "first, second, third, fourth, fifth." The first method is associated with deciding how many toys there are—their cardinality. The second method records the order in which the toys fall. That concept of ordered counting is what ordinal numbers capture.

In common English, an ordinal number refers to the numerical position of an object that is in an ordered list: first, second, third, and so on. The Arabic numerals denote the ordinal numbers with which we are most familiar: $0, 1, 2, 3, \ldots$ (incidentally, with ordinal numbers it is customary to start counting at 0). Continuing this list through all the finite natural numbers gives us an infinite set of ordinal numbers with the same order type as the natural numbers. The concept of ordinal numbers extends this list by creating well-ordered sets of "numbers" that start with the finite numbers, but then just keep going.

We can get a clue about how to accomplish this extension by thinking about what must be true about a set of ordinal numbers based on properties of well-ordered sets. For instance, since we are creating a set of ordinal numbers that will be well-ordered, each ordinal number α must have an immediate next ordinal. Here's why. Suppose there is any ordinal larger than α in our set. Then consider the set of all ordinals larger than α. The property of well-ordering tells us that that set must have a smallest element, which must be the immediate successor of α. It is denoted $\alpha + 1$. Notice that this reasoning applies to sets of ordinals as well as to individual ordinals. For example, consider the set of all the finite ordinals. If there are any ordinals larger than all the finite ordinals, we can think of all the ordinals larger than all those finite ordinals. Then that collection of ordinals must have a least element–which is traditionally denoted ω_0 and read as "omega nought".

We have been a bit vague so far, but this kind of reasoning leads us to a clever method for constructing a list of ordinal numbers that goes on indefinitely, namely, we think of each ordinal number as literally *being* the set of its predecessors.

We start with the empty set, which we identify with the ordinal 0, and work our way up. This strategy gives us a whole new way to look at the finite ordinals, and it has the advantage that this method can be extended beyond the finite ordinals. So here are the first few ordinals:

- $0 = \varnothing$
- $1 = \{\varnothing\}$
- $2 = \{\varnothing, \{\varnothing\}\}$
- $3 = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$
- $\cdots$

Notice that the number of elements in each of the sets above is recorded by its ordinary name. For example, the ordinal number $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$ is a set with 3 elements. We can continue this process of constructing the next ordinal by declaring it to be the set of its predecessors to create an ordinal number for every natural number.

But there are more ordinal numbers "beyond" these finite ordinals, and we can use the exact same method to create them! For we can now take the set of all the finite ordinals $\{0, 1, 2, 3, ...\}$ and declare that set to be the next ordinal, thus creating the smallest ordinal greater than any of the finite ordinals. As we mentioned earlier, we call that first infinite ordinal ω_0. It is the first infinite ordinal number, i.e., the first ordinal number that is an infinite set. Notice that ω_0 is also the union of all the finite ordinals. Indeed, it is also the union of any infinite collection of finite ordinals.

Whenever we have created an ordinal α (which is a set), we have a mechanism for creating its **successor**, denoted $\alpha + 1$ or $S(\alpha)$ (S for successor), namely, the set containing all the elements of α together with α itself. In other words,

$$\alpha + 1 = S(\alpha) = \alpha \cup \{\alpha\}.$$

So the set α is a subset of $\alpha + 1$ and α is also an element of $\alpha + 1$.

Notice that we can create an ordinal that follows any set of ordinals similarly to how we constructed ω_0 as the union of all the finite ordinals. Given a set of ordinals $\{\alpha_\beta\}_{\beta\in\lambda}$ with no largest ordinal, the union $\bigcup_{\beta\in\lambda} \alpha_\beta$ is the smallest ordinal that is larger than all of the α_β's.

So let's construct some more ordinals. The successor of ω_0 is $\omega_0 + 1$; its successor is $\omega_0 + 2$; its successor is $\omega_0 + 3$. Note that all of these ordered sets have the same cardinality, even though they are different order types! This is the difference between ordinal numbers and the so-called *cardinal numbers* that are used to describe the cardinality of a set. There are the finite cardinalities which all correspond to finite ordinals, and the smallest infinite cardinality is often denoted $\aleph_0$, called "aleph nought."

If we continue constructing ordinal successors to $\omega_0 + 3$, we obtain ordinals $\omega_0 + n$ for each $n \in \mathbb{N}$. The set of all such ordinals is the next ordinal, and it is called $2\omega_0$. We can continue this process to create $2\omega_0 + 1, 2\omega_0 + 2, \ldots; \ldots k\omega_0, k\omega_0 + 1, k\omega_0 + 2, \ldots$, and so on. The ordinal numbers named in the previous sentence are sets, each one being the set of its predecessors. Notice that the process outlined in the previous sentence is a countable process creating a countable number of countable sets.

Of course, there is no necessity to ever stop this process of creating more ordinals. Taking the union of all the ordinals that we created above is yet another countable ordinal. But there is no reason to ever stop. We can imagine continuing this process indefinitely. The next conceptual leap is to imagine that we have continued this process so long that we have created all the countable ordinals. We then can take the union of all those countable ordinals to create the first uncountable ordinal number, called ω_1 (whose cardinality is denoted $\aleph_1$, the first cardinality larger than $\aleph_0$). The ordinal ω_1 is simply the set of all the countable ordinals. Every ordinal preceding it (hence in it) is countable. Then the successor of ω_1 is $\omega_1 + 1$; its successor is $\omega_1 + 2$, and these have the same cardinality as ω_1. Continuing for a while, eventually we will reach an ordinal whose cardinality is larger—we call that $\aleph_2$.

Proceeding in this fashion, we can continue creating ordinals indefinitely—creating ordinals of ever increasing cardinality. An interesting fact about all these cardinalities is that they are well-ordered too! So if we continue denoting new cardinalities by cardinal numbers $\aleph_\alpha$, the index α is an ordinal number! Cardinal numbers are a fascinating topic in their own right, though we won't say more about them here.

Conceptually, we can think about all the ordinals; however, the collection of all ordinals is so vast that it is not even a set. The reason is that if it were a set, then the set of all its subsets would have a higher cardinality than it, contradicting the idea that there are ordinals whose cardinality is equal to the cardinality of any given set.

Suppose we have created ordinals in this way. Then one ordinal α is less than another ordinal β if and only if α is an element of β and, also, α is a subset of β. For example, notice above that the set that is defined as 2 is both an element of 3 and is a subset of 3. So the ordinal numbers are ordered by containment; that is, given two ordinals (which, remember, are sets), one of those sets is contained in the other set. Also, ordinals are ordered by membership; that is, given two ordinals, one of those ordinals is a set that is an element of the other ordinal (which is a set of sets).

So we are ready now to give a formal definition of ordinal number. This definition is not illuminating, because it is so abstract, but it captures exactly the features of the sets that we have called ordinals and it pins down what features we want to say are essential to saying that a set is an ordinal number.

> **Effective Thinking Principle.** *Definitions Should Capture Essentials.* After we have isolated essential features of a concept, pin them down with a definition that contains those essentials and as little else as possible.

Definition. An **ordinal number** is a set α such that

1. every element of α is also a subset of α,
2. the elements of α are strictly ordered by membership, that is, an ordinal $\beta \in \alpha$ is less than an ordinal $\gamma \in \alpha$ if and only if β is a member of γ.

> **Effective Thinking Principle.** *Explore Consequences of Definitions.* After making a definition, identify immediate implications of the definition in order to understand its meaning.

Theorem 2.36. *1. If α is an ordinal number, then any element of α is also an ordinal.*

2. $S(\alpha) := \alpha \cup \{\alpha\}$, *the* successor *of* α, *is also an ordinal.*

3. *The union of any set of ordinals is an ordinal.*

4. *The ordinal numbers are naturally ordered by inclusion.*

5. *The intersection of any set of ordinals is an ordinal contained in that set of ordinals and is the least element in the set. Hence, any set of ordinal numbers has a least element. Hence, ordinals are well-ordered.*

To get used to dealing with ordinal numbers, let's make some observations about the first uncountable ordinal number, ω_1, and then prove some consequences.

1. Every ordinal less than ω_1 is a countable set.

2. Every non-empty set of ordinals less than ω_1 has a least element, since ordinals are well-ordered.

3. There are uncountably many countable ordinals less than ω_1.

Those properties are all you will need to prove the following theorems.

Theorem 2.37. *Let $\{\alpha_i\}_{i\in\omega_0}$ be a countable set of countable ordinal numbers; that is, each $\alpha_i < \omega_1$. Then there is an ordinal β such that $\alpha_i < \beta$ for each i and $\beta < \omega_1$.*

Let's give a name to ordinals that do not have immediate predecessors—ordinals such as ω_0 and ω_1 .

Definition. A **limit ordinal** is an ordinal that is not a successor of another ordinal.

Theorem 2.38. *For any countable set of countable ordinals $\{\alpha_i\}_{i\in\omega_0}$, there is a countable limit ordinal γ such that for every ordinal $\beta < \gamma$, there exists an α_i such that $\beta < \alpha_i < \gamma$.*

Theorem 2.39. *Let A and B be unbounded sets of ordinals in ω_1, that is, for every ordinal $\delta \in \omega_1$, there is an ordinal $\alpha \in A$ such that $\delta < \alpha$ and an ordinal $\beta \in B$ such that $\delta < \beta$. Then there exists a limit ordinal γ in ω_1 such that γ is a limit of ordinals from A and is also a limit of ordinals from B.*

An amazing fact (that can be shown by something called transfinite induction, equivalent to the Axiom of Choice) is that every well-ordered set is order-isomorphic to exactly one ordinal number.

2.7 To Infinity and Beyond

For thousands of years, people thought the study of infinity was a topic for poets or mystics. Isolating the essence of what it means for two collections to have the same number of things, namely, the idea of a bijection, opened a whole world of intrigue for us to explore and enjoy. That basic idea about cardinality allowed us to reason about infinity and to prove wonderful theorems that make the study of infinity nuanced and full of richness and variety. We started with several theorems about countably infinite sets—the smallest size of infinity. We then were able to see that infinity itself comes in different sizes. That amazing discovery opened whole new worlds to explore and questions to consider.

You may or may not have worked through many theorems in this chapter on Cardinality. We suggested you might choose to just get a flavor of it without delving too deeply into it so as not to delay your coming introduction to topology. In any case, it is now time for us to leave the intriguing issues of cardinalities and axioms about sets, and introduce the idea of a topological space, an idea that is built on the foundation of set theory.

Chapter 3

Topological Spaces: Fundamentals

The formal study of topology arose in the late 19th and early 20th century out of a desire to free the ideas of analysis from the concept of distance, and to ground these ideas in the study of sets. By this time, many ideas of topology had been around a while; Euler had studied the Königsberg bridge problem in 1736, the word *topologie* had been in use by Listing by 1847, Riemann began studying the surfaces that bear his name in the 1850's, and the Möbius band was introduced in the 1860's. In the 1870's the idea of closed sets and open sets emerged in work by Cantor, and the idea of a neighborhood in work by Weierstrass. Poincaré formalized the idea of connectivity and introduced the idea of homology in the 1890's. By the early 1900's, the idea of a metric space was introduced by Fréchet.

One of the basic goals of mathematics in the early 1900's was to attempt to put all of mathematics on an axiomatic foundation, following the format of Euclidean geometry. This impulse led mathematicians to seek essential features that reside at the root of spaces like the real numbers and other Euclidean spaces and at the core of ideas such as continuity. Sets seemed about as basic as we can get, so the goal was to try to think about how ideas such as continuity could be described using sets rather than the concept of distance.

3.1 Rubber Sheet Geometry and Special Sets

Effective Thinking Principle. *Imagine Alternative Worlds.* One way to create concepts is to imagine our world but with some feature altered, and then follow the consequences.

When we think about replacing the idea of distances by something based on sets, we might want to explore the consequences of living in a world in which objects can be expanded or contracted. Expansion and contraction is just one way to manipulate objects, so we might go further and explore the consequences of imagining a world in which objects can be stretched and bent at will. One informal way to think of the subject of topology is as the study of properties of objects that do not change when we continuously stretch them or distort them. If an object suddenly became elastic, what properties would still remain? For instance, an object that was connected would still remain connected if stretching were allowed, so connectedness is a "topological" property. How can we build a mathematical theory that is equipped to look at deformable objects?

One basic challenge is to generalize the idea of what it means for points in a space to be 'close' to each other. At first, it may seem that the idea of closeness is a geometric feature that could not be meaningful in a deformable object, since two points that seem close together could be made far apart by stretching the object. Distances are not preserved by deformations.

On the other hand, if you imagine an elastic map of Texas, it is qualitatively apparent that the border of Texas would remain close to Texas even if we greatly distorted the map. Upon some reflection we can clarify what we mean by this observation, namely, any "region" around a border point contains points of Texas. That feature would remain true no matter how much we stretched the map. The idea of a border point, which we shall see presages the idea of a "limit point," does not need to use a specific distance. Instead, it suggests that we specify what set constitutes a "region" around a border point.

Pinning down these intuitive ideas leads to the formal definition of a topological space as well as generalizations of the notions of convergence and continuity that we first encountered in calculus.

Effective Thinking Principle. *Push Analogies.* When using an analogy to create an idea, see whether the analogy can be refined by looking at more specific features.

The concept of a point on the border of Texas suggests that we consider the notion of convergence of a sequence. In calculus, we learn that a sequence of numbers converges to number x if, informally speaking, for any target interval you name around x, the terms of the sequence eventually get close to x (within the target interval) and stay there. Here again, notice how the idea of closeness is measured not by a single region but by a collection of regions (in this case, intervals).

In calculus, the idea of a continuous function is also determined by a collection of regions. Intuitively speaking, a continuous function f maps points that are close to x in the domain to points that are close to $f(x)$ in the co-domain. A more formal definition says that for any target interval you name around $f(x)$ there is an interval around x that completely maps inside the target interval around $f(x)$.

Some subsets of the reals are not pertinent to our discussions about convergence and continuity. For example, the set of rational numbers is not a set that comes up when we define convergence or continuity. So some subsets of the reals arise when we develop the ideas of convergence and continuity while some subsets are not mentioned even implicitly in those definitions.

Let's try a thought experiment. Can we describe convergence and continuity without using distances? In general terms, a continuous function is one that maps points that are close to each other in the domain to points that are close to each other in the range. The formal definition of continuity of a function f at a point x says that for any pre-specified distance around $f(x)$ in the co-domain, there is a distance in the domain such that all points whose distance from x is less than that distance get mapped to within the pre-specified distance from $f(x)$ in the range. Let's rephrase that definition without using the idea of distance. We could say, a function f is continuous at x if given any special set around $f(x)$ in the range, there is a special set in the domain around x such that all the points in the domain's special set get mapped into the range's special set.

We could phrase the concept of convergence in that same general manner. We could say that an infinite list of points converges to a limit point should mean that for any special set around the limit point, all but a finite number of points in our list of points lie in that special set.

Of course, if all we meant by 'special set' was ϵ-neighborhood, we would not have accomplished our goal of generalizing the ideas of convergence or continuity. We would simply have rephrased them. So now we break free of the concept of distance and think far more abstractly. The concept of convergence basically is saying that in some set X we have designated a special collection of subsets that we will use to define the meaning of convergence. A sequence of points converges to a limit point if for any one of our special sets around the limit point, all but finitely many of our set of points lie inside that special set.

Thinking about continuity in the abstract, we have two sets X and Y, the domain X and the range Y. There are some specially designated subsets of X and specially designated subsets of Y. Then we want to say that a function $f : X \to Y$ is continuous if for any point $x \in X$ and any specially designated subset V of Y containing $f(x)$, there is a specially designated subset U of X containing x such that $f(U) \subset V$.

These specially designated subsets are not just ϵ-neighborhoods, but they are generalizations of that idea. So this analysis has turned our attention toward the idea of specifying certain subsets of our space as being distinguished as sets in which we have special interest with respect to considerations of limit points, convergence, and continuity.

Definition of a Topology

So far we have described a format for the phraseology of concepts like convergence and continuity; however, we have not defined the conditions that would make a reasonable collection of distinguished subsets. Again let's return to our generative example of the real numbers with our usual sense of distance and see whether there are some conditions on the distinguished subsets that seem especially pertinent. We want our collection of distinguished subsets to reflect the essential features that make convergence and continuity work.

In describing convergence and continuity, we find ourselves selecting a challenge set around a point and then concluding that some points are in that challenge set. In the case of convergence, all except a finite number of points are in the challenge set. So if we have two challenge sets containing the same point, then all except a finite number of points must be in the intersection of the two challenge sets. This observation suggests that the intersection of two distinguished sets should also be a distinguished set. We can reach the same intuition when we think about continuity. So let's settle on the insight that the intersection of two distinguished sets should itself be in our collection of distinguished sets.

Let's think about a function $f : X \to Y$. Recall that our intuition told us that f is continuous if for any point $x \in X$ and any specially designated subset V of Y containing $f(x)$, there is a specially designated subset U of X containing x such that $f(U) \subset V$. However, we might want to think more globally and realize that many points may all go to the same point y in Y. So if f is continuous and y is a point in Y and V is a specially designated set containing y, then for each such point x for which $f(x)$ goes to y, there is a designated set U_x in X with x in U_x such that $f(U_x) \subset V$. That means the union of all such designated subsets U_x must also have the property that their union's image is in V. This property suggests that arbitrary unions of distinguished sets should be declared to be distinguished.

> **Effective Thinking Principle.** *Settle on an Idea and Explore the Consequences.* After identifying essential elements, make a definition and explore the implications.

The two properties that finite intersections of distinguished subsets and arbitrary unions of distinguished subsets are distinguished form the essential ingredients in formulating the definition of a topology. The definition is completed by making the whole space X be distinguished, which is equivalent to saying that each point of X is in some one of the distinguished subsets. Finally, for technical reasons it is convenient to include the empty set as distinguished. We are now ready for the formal definition of a topology and a topological space.

3.2 Open Sets and the Definition of a Topological Space

Definition. Suppose X is a set. Then $\mathcal{T}$ is a **topology** on X if and only if $\mathcal{T}$ is a collection of subsets of X such that

1. $\varnothing \in \mathcal{T}$,
2. $X \in \mathcal{T}$,
3. if $U \in \mathcal{T}$ and $V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$, and
4. if $\{U_\alpha\}_{\alpha\in\lambda}$ is any collection of sets of $\mathcal{T}$, then $\bigcup_{\alpha\in\lambda} U_\alpha \in \mathcal{T}$.

Recall that the notation $\bigcup_{\alpha\in\lambda} U_\alpha$ means the arbitrary union over a possibly uncountable index set λ.

Definition. A **topological space** is an ordered pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a topology on X. We use the word **space** to mean topological space unless otherwise noted.

Note that a set X may admit many different topologies. Then $(X, \mathcal{T})$ and $(X, \mathcal{T}')$ are *different* topological spaces if $\mathcal{T} \neq \mathcal{T}'$, even though the underlying set X is the same. When it is clear what topology a space $(X, \mathcal{T})$ has, we will denote the space by X.

The definition of a topological space arose from an abstraction of our familiar concepts of limits and continuity in Euclidean spaces. The properties of the sets in a topology were properties that were satisfied by the usual open sets of $\mathbb{R}$, so we will use the term **open** to refer to sets in a topology.

Definition. A set $U \subset X$ is called an **open set** in $(X, \mathcal{T})$ if and only if $U \in \mathcal{T}$.

So we can think of a topology on X as specifying what subsets of X will be considered open. (By the way, it is customary to use the letters U and V to denote open sets.) Keep in mind that open sets are *elements* of the topology $\mathcal{T}$ and *subsets* of the space X. Elements of X, on the other hand, are the *points* of the space X.

Effective Thinking Principle. *Explore Consequences of Definitions.* After making a definition, explore consequences and reformulations.

The definition of a topology includes the condition that the arbitrary union of open sets is open. The definition also specifies that the intersection of two open sets is open, but that condition actually implies that any *finite* intersection of open sets is open.

Theorem 3.1. *Let $\{U_i\}_{i=1}^n$ be a* finite *collection of open sets in a topological space $(X, \mathcal{T})$. Then $\bigcap_{i=1}^n U_i$ is open.*

Exercise 3.2. *Why does your proof not prove the false statement that the infinite intersection of open sets is necessarily open?*

To check whether a subset U of X is an open set, we only need to confirm that each point of U is in an open set that is contained in U.

Theorem 3.3. *A set U is open in a topological space $(X, \mathcal{T})$ if and only if for every point $x \in U$, there exists an open set U_x such that $x \in U_x \subset U$.*

We sometimes call an open set containing x a **neighborhood of** x. Thus a set U is open if and only if every point has a neighborhood that lies within U.

Effective Thinking Principle. *Generalizations Should Generalize.* If a definition is intended to generalize a specific example, check to make certain it does so.

Let's check that our definition of a topological space captures the relevant features of the prototype that spawned it, namely, the real number line and the familiar concept of "open sets" from calculus.

Example. The **standard topology** $\mathcal{T}_{\text{std}}$ on $\mathbb{R}$ is defined as follows: a subset U of $\mathbb{R}$ belongs to $\mathcal{T}_{\text{std}}$ if and only if for each point p of U there is some $\varepsilon_p > 0$ such that the interval $(p - \varepsilon_p, p + \varepsilon_p)$ is contained in U. We may write $\mathbb{R}_{\text{std}}$ for $(\mathbb{R}, \mathcal{T}_{\text{std}})$, although anytime we see $\mathbb{R}$ without any topology mentioned, we should assume it has the standard topology.

We can generalize the standard topology on $\mathbb{R}$ to **Euclidean space** $\mathbb{R}^n$, the set of all n-tuples of real numbers. In $\mathbb{R}^n$, recall the **Euclidean distance** between points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ is given by

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

We use the Euclidean distance to produce **open balls** in $\mathbb{R}^n$, which will play the role that open intervals did in $\mathbb{R}$. Recall in $\mathbb{R}^n$, the **open ball** of radius $\varepsilon > 0$ around point $p \in \mathbb{R}^n$ is the set

$$B(p, \varepsilon) = \{x \mid d(p, x) < \varepsilon\}.$$

Example. The **standard topology** $\mathcal{T}_{\text{std}}$ on $\mathbb{R}^n$ is defined as follows: a subset U of $\mathbb{R}^n$ belongs to $\mathcal{T}_{\text{std}}$ if and only if for each point p of U there is an $\varepsilon_p > 0$ such that $B(p, \varepsilon_p) \subset U$.

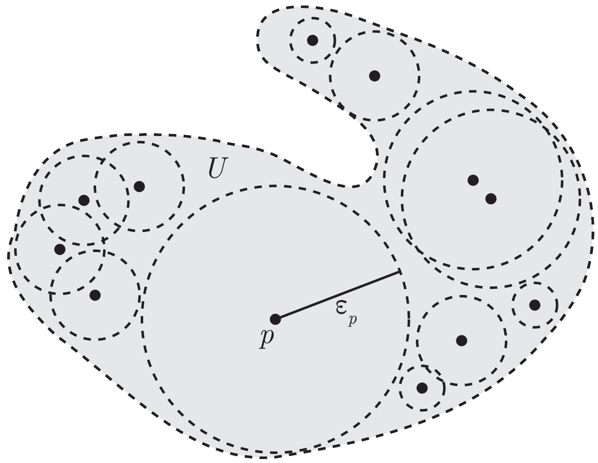


Figure 3.1: An open set U in $\mathbb{R}^n$.

Exercise 3.4. *Verify that $\mathcal{T}_{\text{std}}$ is a topology on $\mathbb{R}^n$; in other words, it satisfies the four conditions of the definition of a topology.*

Let us consider some other examples of topological spaces.

Example. Given a set X, let $\mathcal{T} = 2^X$, the set of all subsets of X. This topology $\mathcal{T}$ is called the **discrete topology** on X. We call the space $(X, 2^X)$ a **discrete topological space**. In the discrete topology, every subset of X is an open set.

Note the spelling: *discrete* topology, not *discreet* topology! The discrete topology on $\mathbb{R}$ is different from the standard topology on $\mathbb{R}$ because some sets are open in one topology but not in the other—for example, a single point is open in the discrete topology but not open in the standard topology on $\mathbb{R}$.

Example. Given a set X, the collection $\mathcal{T} = \{\varnothing, X\}$ is called the **indiscrete topology** on X. We call the space $(X, \{\varnothing, X\})$ an **indiscrete topological space**. In this space, there are only two open sets: the empty set and the entire set X.

Thus, the discrete topology has the largest possible collection of open sets that any topology can have, while the indiscrete topology has the smallest possible collection of open sets that any topology can have. A fun way to think about a topology is to imagine it as a pair of glasses: when you put them on, the open sets are the sets that you can "see." So the discrete topology glasses allow you to see everything very sharply; in fact, you can see individual points. But with the indiscrete topology, there's only one thing to see.

Example. Given a set X, the **finite complement** (or **co-finite**) **topology** on X is described as follows: a subset U of X is open if and only if $X - U$ is finite or $U = \varnothing$.

Recall that a countable set is one that is either finite or countably infinite.

Example. Given a set X, the **countable complement (or co-countable) topology** on X is described as follows: a subset U of X is open if and only if $X - U$ is countable or $U = \varnothing$.

Exercise 3.5. *Verify that the discrete, indiscrete, finite complement, and countable complement topologies are indeed topologies for any set* X*.*

Exercise 3.6. *Describe some of the open sets you get if* $\mathbb{R}$ *is endowed with the topologies described above (standard, discrete, indiscrete, finite complement, and countable complement). Specifically, identify sets that demonstrate the differences among these topologies, that is, find sets that are open in some topologies but not in others. For each of the topologies, determine if the interval* $(0, 1)$ *is an open set in that topology.*

Although a finite intersection of open sets is an open set, an infinite intersection of open sets need not be open.

Exercise 3.7. *Give an example of a topological space and a collection of open sets in that topological space that show that the* infinite *intersection of open sets need not be open.*

3.3 Limit Points and Closed Sets

One of the most basic concepts in topology is the concept of a limit point of a set. This idea captures the topological abstraction of the idea of a limit point that occurs in calculus.

Definition. Let $(X, \mathcal{T})$ be a topological space, A be a subset of X, and p be a point in X. Then p is a **limit point** of A if and only if for each open set U containing p, $(U - \{p\}) \cap A \neq \varnothing$. Notice that p may or may not belong to A.

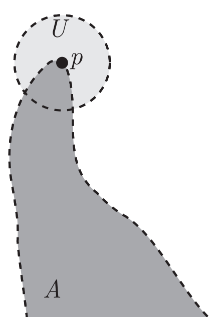


Figure 3.2: The point p is a limit point of A if every open set U containing p contains other points of A. Drawing a picture like this one can greatly help you understand the definition.

In other words, p is a limit point of A if *every* open set containing p intersects A at *some point other than* p itself. Thus, the concept of a limit point gives us a way of capturing the idea of a point being “arbitrarily close” to a set *without* using the concept of distance. Instead, we use the idea of open sets in a topology.

So if we change the topology, then what is considered “arbitrarily close” may change, as the next exercise shows.

Exercise 3.8. *Let* $X = \mathbb{R}$ *and* $A = (1, 2)$. *Verify that* 0 *is a limit point of* A *in the indiscrete topology and the finite complement topology, but not in the standard topology nor the discrete topology on* $\mathbb{R}$.

An important step in understanding the definition of being a limit point is to understand what it means to *not* be a limit point. The following theorem is useful in theorems to come.

Theorem 3.9. *Suppose* $p \notin A$ *in a topological space* $(X, \mathcal{T})$. *Then* p *is* not *a limit point of* A *if and only if there exists a neighborhood* U *of* p *such that* $U \cap A = \varnothing$.

Definition. Let $(X, \mathcal{T})$ be a topological space, A be a subset of X, and p be a point in X. If $p \in A$ but p is not a limit point of A, then p is an **isolated point** of A.

Exercise 3.10. *If* p *is an isolated point of a set* A *in a topological space* X, *then there exists an open set* U *such that* $U \cap A = \{p\}$.

> **Effective Thinking Principle.** *Create Examples.* When learning any definition in mathematics, one helpful step is to construct several examples that illustrate the meaning of the definition.

The following exercise encourages you to undertake that process with the goal of better understanding the idea of a limit point.

Exercise 3.11. *Give examples of sets* A *in various topological spaces* $(X, \mathcal{T})$ *with*

1. *a limit point of* A *that is an element of* A;
2. *a limit point of* A *that is not an element of* A;
3. *an isolated point of* A;
4. *a point not in* A *that is not a limit point of* A.

The definition of limit point forms the central idea in the definitions of closure and closed set.

Definition. Let $(X, \mathcal{T})$ be a topological space, and $A \subset X$. Then the **closure** of A in X, denoted $\overline{A}$ or $\mathrm{Cl}(A)$ or $\mathrm{Cl}_X(A)$, is the set A together with all its limit points in X.

Definition. Let $(X, \mathcal{T})$ be a topological space and $A \subset X$. The subset A is **closed** if and only if $\overline{A} = A$, in other words, if A contains all its limit points.

Exercise 3.12.

1. *Which sets are closed in a set X with the discrete topology?*

2. *Which sets are closed in a set X with the indiscrete topology?*

3. *Which sets are closed in a set X with the finite complement topology?*

4. *Which sets are closed in a set X with the countable complement topology?*

At first the following theorem may appear to be stating the obvious; however, it requires you to carefully understand the definitions of closure and closed set. It's a good idea to draw a picture.

Effective Thinking Principle. *Draw a Picture.* Part of the value of drawing a picture is to see features in your picture that you did not intentionally put there. For example, if you draw a triangle without thinking about it, the sum of the angles will equal 180 degrees. A picture can suggest insights that will help with your argument. Get in the habit of drawing pictures!

Theorem 3.13. *For any topological space $(X, \mathcal{T})$ and $A \subset X$, $\overline{A}$ is closed. That is, for any set A in a topological space, $\overline{\overline{A}} = \overline{A}$.*

A basic relationship between open sets and closed sets in a topological space is that they are complements of each other.

Theorem 3.14. *Let $(X, \mathcal{T})$ be a topological space. Then the set A is closed if and only if $X - A$ is open.*

Removing a closed set from an open set leaves an open set.

Theorem 3.15. *Let $(X, \mathcal{T})$ be a topological space, and let U be an open set and A be a closed subset of X. Then the set $U - A$ is open and the set $A - U$ is closed.*

The properties of closed sets in a topological space mirror the properties of open sets. From that perspective, the four defining properties of a topological space are captured in the following theorem about closed sets.

Theorem 3.16. *Let $(X, \mathcal{T})$ be a topological space. Then:*

i) $\varnothing$ *is closed.*

ii) X *is closed.*

iii) The union of finitely many closed sets is closed.

iv) Let $\{A_\alpha\}_{\alpha\in\lambda}$ *be a collection of closed subsets in* $(X,\mathcal{T})$*. Then* $\bigcap_{\alpha\in\lambda} A_\alpha$ *is closed.*

The theorem above shows that an alternative way to build up a theory of topology would be to specify all subsets that are "closed," and demand that they satisfy the above four properties.

The following exercises will help you understand closed sets and their relationships to open sets.

Exercise 3.17. *Give an example to show that the union of infinitely many closed sets in a topological space may be a set that is not closed.*

Exercise 3.18. *Give examples of topological spaces and sets in them that:*

1. *are closed, but not open;*
2. *are open, but not closed;*
3. *are both open and closed;*
4. *are neither open nor closed.*

These counterexamples show that the words 'closed' and 'open' are not antonyms!

Exercise 3.19. *State whether each of the following sets are open, closed, both, or neither.*

1. *In* $\mathbb{Z}$ *with the finite complement topology:* $\{0,1,2\}$, $\{$*prime numbers*$\}$, $\{n \mid |n| \geq 10\}$.
2. *In* $\mathbb{R}$ *with the standard topology:* $(0,1), (0,1], [0,1], \{0,1\}, \{1/n \mid n\in\mathbb{N}\}$.
3. *In* $\mathbb{R}^2$ *with the standard topology:* $\{(x,y) \mid x^2+y^2=1\}$, $\{(x,y) \mid x^2+y^2>1\}$, $\{(x,y) \mid x^2+y^2\geq 1\}$.

One way to think about the closure of a set is as the intersection of all the closed sets that contain it.

Theorem 3.20. *For any set* A *in a topological space* X*, the closure of* A *equals the intersection of all closed sets containing* A*, that is,*

$$\overline{A} = \bigcap_{B\supset A, B\in\mathcal{C}} B$$

where $\mathcal{C}$ *is the collection of all closed sets in* X*.*

Informally, we can say $\overline{A}$ is the "smallest" closed set that contains A. To gain some intuition about the process of taking the closure of a set, it is valuable to consider the closures of various sets in various topological spaces. The next exercise asks you to do so.

Exercise 3.21. *Pick several different subsets of $\mathbb{R}$, and find their closures in:*

1. *the discrete topology;*

2. *the indiscrete topology;*

3. *the finite complement topology;*

4. *the standard topology.*

The next theorem includes the fact that the closure of the union of two sets is the union of the closures of those two sets. Try proving the following in two different ways: (i) using the definition of a limit point, and (ii) using Theorem 3.20.

Theorem 3.22. *Let A and B be subsets of a topological space X. Then*

1. *$A \subset B$ implies $\overline{A} \subset \overline{B}$.*

2. *$\overline{A \cup B} = \overline{A} \cup \overline{B}$.*

Effective Thinking Principle. *Extending Theorems.* A habit that will help you to understand mathematics is always to look for the extent to which a theorem can be extended or not.

The following exercise asks you undertake this exploration with respect to the previous theorem.

Exercise 3.23. *Let $\{A_\alpha\}_{\alpha \in \lambda}$ be a collection of subsets of a topological space X. Then is the following statement true?*

$$\overline{\bigcup_{\alpha \in \lambda} A_\alpha} = \bigcup_{\alpha \in \lambda} \overline{A}_\alpha.$$

Sometimes the closure of a set may not be completely obvious. Let's explore the closures of two interesting subsets of $\mathbb{R}^2$ with the standard topology.

Exercise 3.24. *In $\mathbb{R}^2$ with the standard topology, describe the limit points and closure of each of the following two sets:*

1. $S = \left\{\left(x, \sin\left(\frac{1}{x}\right)\right) \mid x \in (0,1)\right\}$. *The closure of the set S is called the* topologist's sine curve.

2. $C = \{(x, 0) \mid x \in [0,1]\} \cup \bigcup_{n=1}^{\infty} \left\{\left(\frac{1}{n}, y\right) \mid y \in [0,1]\right\}$. *The closure of the set C is called the* topologist's comb.

The following exercise is difficult, but the example that you will produce is fascinating. If you cannot do this exercise now, do not worry. It will reappear again later in the book.

Exercise 3.25. *In the standard topology on $\mathbb{R}$, there exists a non-empty subset C of the closed unit interval $[0, 1]$ that is closed, contains no non-empty open interval, and where no point of C is an isolated point.*

3.4 Interior and Boundary

Just as we explored the concept of the "smallest" closed set containing A, we can consider what is the "largest" open set *contained in* A.

Definition. The **interior** of a set A in a topological space X, denoted A° or $\operatorname{Int}(A)$, is defined as:

$$\operatorname{Int}(A) = \bigcup_{U \subset A,\, U \in \mathcal{T}} U.$$

Points of $\operatorname{Int}(A)$ are called **interior points** of A.

Theorem 3.26. *Let A be a subset of a topological space X. Then p is an interior point of A if and only if there exists an open set U with $p \in U \subset A$.*

Exercise 3.27. *Show that a set U is open in a topological space X if and only if every point of U is an interior point of U.*

The following definition of the word "boundary" lets us take any set A in a topological space and prove that the whole space is equal to the disjoint union of the interior of A, the boundary of A, and the interior of the complement of A.

Definition. The **boundary** of A, denoted $\operatorname{Bd}(A)$ or ∂A, is defined to be $\overline{A} \cap \overline{X - A}$.

Theorem 3.28. *Let A be a subset of a topological space X. Then $Int(A)$, $Bd(A)$ and $Int(X - A)$ are disjoint sets whose union is X.*

Exercise 3.29. *Pick several different subsets of $\mathbb{R}$, and for each one, find its interior and boundary using:*

1. *the discrete topology;*
2. *the indiscrete topology;*
3. *the finite complement topology;*
4. *the standard topology.*

3.5 Convergence of Sequences

When you learned about convergence of sequences in calculus, you were dealing with the standard topology on the real line. Our challenge now is to make a definition that extends that concept to the setting of a general topological space.

Definition. A **sequence** in a topological space X is a function from $\mathbb{N}$ to X. The image of i under this function is a point of X denoted x_i and we traditionally write the sequence by listing its images: $x_1, x_2, x_3, ...$ or in shorter form: $(x_i)_{i \in \mathbb{N}}$.

Definition. A point $p \in X$ is a **limit of the sequence** $(x_i)_{i \in \mathbb{N}}$, or, equivalently, $(x_i)_{i \in \mathbb{N}}$ **converges** to p (written $x_i \to p$), if and only if for every open set U containing p, there is a $N \in \mathbb{N}$ such that for all $i > N$, the point x_i is in U.

Theorem 3.30. *Let A be a subset of the topological space X and let p be a point in X. If $\{x_i\}_{i \in \mathbb{N}} \subset A$ and $x_i \to p$, then p is in the closure of A.*

As we shall see later, in some topological spaces, the converse of the previous result is not true. But it is true for $\mathbb{R}^n$.

Theorem 3.31. *In standard topology on $\mathbb{R}^n$, if p is a limit point of a set A, then there is a sequence of points in A that converge to p.*

In the standard topology on the real line, convergent sequences converge to a unique limit. That uniqueness property does not hold for all topological spaces.

Exercise 3.32. *Find an example of a topological space and a convergent sequence in that space, where the limit of the sequence is not unique.*

Exercise 3.33.

1. *Consider sequences in $\mathbb{R}$ with the finite complement topology. Which sequences converge? To what value(s) do they converge?*

2. *Consider sequences in $\mathbb{R}$ with the countable complement topology. Which sequences converge? To what value(s) do they converge?*

After we have created some additional interesting topological spaces and after we have defined continuity in topological spaces in the chapters ahead, we will explore convergence further, but for now let's be satisfied with understanding the definition of convergence and the basic properties of convergence that we have seen above.

3.6 Topological Essentials

In this chapter, we have constructed the foundation on which the whole study of topology is built. From our familiar examples of the real numbers and higher dimensional Euclidean spaces, we extracted essential set-theoretic features that undergird ideas such as convergence and continuity. That exploration led to the definition of a topological space, which is the topic of study for the rest of this book and much, much more.

One of the amazing realities of mathematics and our understanding of the world is that when we identify truly central essentials and create ideas that capture them, the exploration of those concepts becomes an incredibly rich and limitless adventure. We now stand at the base of a huge, invisible mountain of knowledge, currently unaware of what lies before us. But every step ahead will expose beautiful vistas of the world of topology.

Chapter 4

Bases, Subspaces, Products: Creating New Spaces

In the last chapter we defined a topology as a collection of sets—the open sets—which abstracted the intuitive idea of "closeness." Specifying which sets are in a topology can be a difficult process. One way to simplify the specification of a topology is to describe building blocks from which all the open sets arise. This strategy is analogous to describing a collection of basis vectors that generate a vector space, as we do in linear algebra. The impulse to identify generators of a topology leads to the concept of a *basis* for a topology. In this chapter we will develop the concept of a basis and use it to build new topological spaces.

Other strategies for defining new topological spaces involve starting with one or more known spaces and creating new spaces from the old ones. Considering a subset of a topological space gives us a way to create a *subspace* from an old space. And taking a Cartesian *product* of topological spaces is another way to create a new space. Subspaces and product spaces have topologies that naturally flow from the original spaces.

By the end of this chapter, you will have created many fascinating, new topological spaces to explore and enjoy.

4.1 Bases

One of the most powerful strategies for coming up with new ideas in one field is to consider concepts in other fields and see whether analogous reasoning might apply. In linear algebra, a basis is a collection of vectors that generate the whole vector space. In number theory, the set of primes generate all natural numbers above 1. Let's see where the impulse to find a generating set might lead us in topology.

Effective Thinking Principle. *Borrow Strategies from Other Fields.* Powerful concepts from other fields can inspire new insights either directly or by analogy.

As you recall, we defined a subset U of $\mathbb{R}$ to be an open set in the standard topology if we can find an open interval contained in U around every point in U. Thus, every open set in the standard topology on $\mathbb{R}$ is the union of these simpler open sets (the open intervals).

That example motivates the definition of a basis for a topology. Because arbitrary unions of open sets are open, a topological space can have extremely complicated open sets. It is often convenient to describe a (simpler) subcollection of open sets (like intervals in the example of $\mathbb{R}$ with the standard topology) that *generate* all open sets in a given topology by taking unions. This strategy leads to the definition of a *basis* for a topology.

Definition. Let $\mathcal{T}$ be a topology on a set X and let $\mathcal{B} \subset \mathcal{T}$. Then $\mathcal{B}$ is a **basis** for the topology $\mathcal{T}$ if and only if every open set in $\mathcal{T}$ is the union of elements of $\mathcal{B}$. If $B \in \mathcal{B}$, we say B is a **basis element** or **basic open set**. Note that B is an *element* of the basis $\mathcal{B}$, but a *subset* of the space X.

Note: by definition, an empty union is the empty set, so any basis $\mathcal{B}$ will generate the empty set as a union of none of the elements of $\mathcal{B}$. (We recommend you spend an empty amount of time thinking about the empty set.)

Given a topology on some space X, how can we test whether a collection of subsets forms a basis for that topology? The next theorem gives an answer.

Theorem 4.1. *Let $(X, \mathcal{T})$ be a topological space and $\mathcal{B}$ be a collection of subsets of X. Then $\mathcal{B}$ is a basis for $\mathcal{T}$ if and only if:*

1. *$\mathcal{B} \subset \mathcal{T}$, and*

2. *for each set U in $\mathcal{T}$ and point p in U there is a set V in $\mathcal{B}$ such that $p \in V \subset U$.*

The plural of basis is *bases*. A particular topology can have many different bases. This next exercise describes a couple of different bases for the standard topology on $\mathbb{R}$.

Exercise 4.2.

1. *Let $\mathcal{B}_1 = \{(a, b) \subset \mathbb{R} \mid a, b \in \mathbb{Q}\}$. Show that $\mathcal{B}_1$ is a basis for the* standard topology *on $\mathbb{R}$.*

2. *Let $\mathcal{B}_2 = \{(a, b) \cup (c, d) \subset \mathbb{R} \mid a < b < c < d$ are distinct irrational numbers$\}$. Show that $\mathcal{B}_2$ is also a basis for the* standard topology *on $\mathbb{R}$.*

Suppose you are given a set X and a collection $\mathcal{B}$ of subsets of X. Under what circumstances is there a topology for which $\mathcal{B}$ is a basis? Note how this question is different from the question that motivated Theorem 4.1. That theorem describes when a collection of sets forms a basis for a *given* topology. The next theorem describes when a collection of sets forms a basis for *some topology* on X.

Theorem 4.3. *Suppose X is a set and $\mathcal{B}$ is a collection of subsets of X. Then $\mathcal{B}$ is a basis for some topology on X if and only if:*

1. *each point of X is in some element of $\mathcal{B}$, and*
2. *if U and V are sets in $\mathcal{B}$ and p is a point in $U \cap V$, there is a set W in $\mathcal{B}$ such that $p \in W \subset (U \cap V)$.*

Theorem 4.3 allows us to describe topological spaces by first specifying a set X and then a collection $\mathcal{B}$ of subsets of X satisfying the two conditions listed in the theorem. Then the topology $\mathcal{T}$ with basis $\mathcal{B}$ is the collection of all possible unions of basis elements.

Example. We can define an alternative topology on $\mathbb{R}$, called the **lower limit topology**, generated by a basis consisting of all sets of the form $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$. Denote this space by $\mathbb{R}_{\mathrm{LL}}$. It is sometimes called the **Sorgenfrey line** or $\mathbb{R}^1_{bad}$.

Exercise 4.4. *Show that the basis proposed above for the lower limit topology is in fact a basis.*

As we shall see, $\mathbb{R}_{\mathrm{LL}}$ is a topological space with many interesting (and bad) properties. Its topology has more sets than the standard topology on $\mathbb{R}$.

Theorem 4.5. *Every open set in $\mathbb{R}_{\mathrm{std}}$ is an open set in $\mathbb{R}_{\mathrm{LL}}$, but not vice versa.*

Definition. Suppose $\mathcal{T}$ and $\mathcal{T}'$ are two topologies on the same underlying set X. If $\mathcal{T} \subset \mathcal{T}'$, then we say $\mathcal{T}'$ is *finer* than $\mathcal{T}$. Alternatively, we say $\mathcal{T}$ is *coarser* than $\mathcal{T}'$. We say *strictly coarser* or *strictly finer* if $\mathcal{T} \neq \mathcal{T}'$.

If you find it hard to remember this definition, think of a comb with more teeth as *finer* than one with fewer teeth! Theorem 4.5 shows that the topology of $\mathbb{R}_{\mathrm{LL}}$ is strictly finer than the standard topology of $\mathbb{R}$. Not every pair of topologies on the same set is comparable using the finer and coarser relationship.

Exercise 4.6. *Give an example of two topologies on $\mathbb{R}$ such that neither is finer than the other, that is, the two topologies are not comparable.*

The strategy of describing a topology by describing a basis for it allows us to create many topological spaces.

Example. We construct a topological space called the **Double Headed Snake**. Let $\mathbb{R}_{+00}$ be the set consisting of $\mathbb{R}_+$ (the positive real numbers) together with two points which we'll call $0'$ and $0''$. Put a topology on it generated by a basis consisting of all intervals in $\mathbb{R}_+$ of the form (a, b) or else of the form $(0, b) \cup \{0'\}$ or $(0, b) \cup \{0''\}$ for $b \in \mathbb{R}_+$.

Exercise 4.7. *Check that the collection of sets that we specify as a basis in the Double Headed Snake actually forms the basis for a topology.*

Exercise 4.8. *In the Double Headed Snake, show that every point is a closed set; however, it is impossible to find disjoint open sets U and V such that $0' \in U$ and $0'' \in V$.*

Example. Let $\mathbb{R}_{\text{har}}$ be the set $\mathbb{R}$ with a topology whose basis is all sets of the form (a, b) or $(a, b) - H$ where $H = \{1/n\}_{n\in\mathbb{N}}$. You should check that these sets forms the basis for a topology.

Exercise 4.9.

1. *In the topological space $\mathbb{R}_{\text{har}}$, what is the closure of the set $H = \{1/n\}_{n\in\mathbb{N}}$?*
2. *In the topological space $\mathbb{R}_{\text{har}}$, what is the closure of the set $H^- = \{-1/n\}_{n\in\mathbb{N}}$?*
3. *Is it possible to find disjoint open sets U and V in $\mathbb{R}_{\text{har}}$ such that $0 \in U$ and $H \subset V$?*

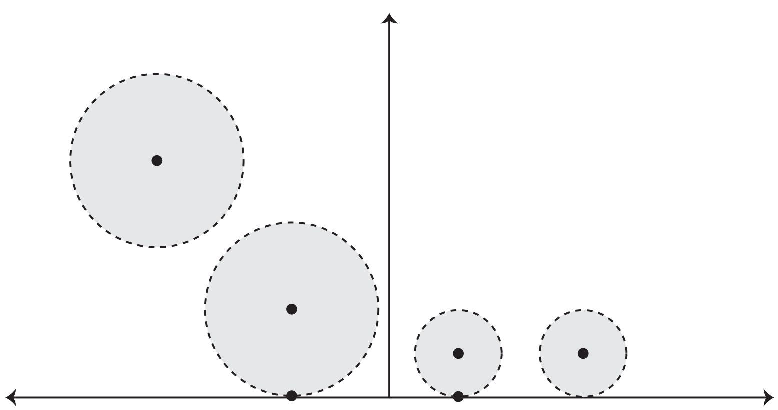


Figure 4.1: Basic open sets of the Sticky Bubble Topology on the upper half plane.

Example. Let $\mathbb{H}_{\text{bub}}$ be the upper half plane $\{(x, y) : x, y \in \mathbb{R}, y \geq 0\}$ with a topology whose basis consists of:

1. all balls $B((x, y), r)$ where $0 < r \leq y$, and
2. all sets $B((x, y), r) \cup \{(x, 0)\}$ where $r = y > 0$.

We will call this space the **Upper Half Plane with the Sticky Bubble Topology**.

The next exercise foreshadows some of the so-called *separation properties* of topologies that we will study later.

Exercise 4.10. *1. In $\mathbb{H}_{\mathrm{bub}}$, what is the closure of the set of rational points on the x-axis?*

2. In $\mathbb{H}_{\mathrm{bub}}$, which subsets of the x-axis are closed sets?

3. In $\mathbb{H}_{\mathrm{bub}}$, let A be a countable set on the x-axis and let z be a point on the x-axis not in A. Then there exist disjoint open sets U and V such that $A \subset U$ and $z \in V$. (Do you need the countability hypothesis on A?)

4. In $\mathbb{H}_{\mathrm{bub}}$, let A and B be countable sets on the x-axis such that A and B are disjoint. Then there exist disjoint open sets U and V such that $A \subset U$ and $B \subset V$.

5. In $\mathbb{H}_{\mathrm{bub}}$, let A be the rational numbers and let B be the irrational numbers. Do there exist disjoint open sets U and V such that $A \subset U$ and $B \subset V$?

There is an interesting topology that we can place on the integers $\mathbb{Z}$.

Example. Let $\mathbb{Z}_{\mathrm{arith}}$ be the set $\mathbb{Z}$ with a topology whose basis elements are arithmetic progressions, i.e., sets of the form $\{az + b : z \in \mathbb{Z}\}$ for $a, b \in \mathbb{Z}$, $a \neq 0$.

Exercise 4.11. *Check that the arithmetic progressions form a basis for a topology on $\mathbb{Z}$.*

In this topology, integers that are in arithmetic progressions are considered "close". So the topology encodes the arithmetic structure. In fact, this topology can be used to prove a standard theorem from number theory!

Theorem 4.12. *There are infinitely many primes.*

As a hint, start by examining the set $p\mathbb{Z}$ where p is prime, and show that it is a closed set in $\mathbb{Z}_{\mathrm{arith}}$. Recall also that a prime number in $\mathbb{Z}$ is a natural number greater than 1 whose only positive factors are 1 and itself, and that every natural number greater than 1 can be written as a product of prime factors.

4.2 Subbases

We saw in Section 4.1 how a basis determines a topology, namely, each open set is a union of basis elements. We can specify topologies in an even more condensed form by means of a *subbasis*, which generates a topology when you allow both arbitrary unions *and* finite intersections.

Definition. Let $(X, \mathcal{T})$ be a topological space and let $\mathcal{S}$ be a collection of subsets of X. Then $\mathcal{S}$ is a **subbasis** for $\mathcal{T}$ if and only if the collection $\mathcal{B}$ of all finite intersections of sets in $\mathcal{S}$ is a basis for $\mathcal{T}$. An element of $\mathcal{S}$ is called a *subbasis element* or a **subbasic open set**.

Exercise 4.13. *A basis for a topology is also a subbasis for that topology.*

The following theorem describes conditions under which we can conclude that a collection of subsets is a subbasis for a given topology:

Theorem 4.14. *Let $(X, \mathcal{T})$ be a topological space and let $\mathcal{S}$ be a collection of subsets of X. Then $\mathcal{S}$ is a subbasis for $\mathcal{T}$ if and only if*

1. *$\mathcal{S} \subset \mathcal{T}$, and*

2. *for each set U in $\mathcal{T}$ and point p in U there is a finite collection $\{V_i\}_{i=1}^n$ of elements of $\mathcal{S}$ such that*

$$p \in \bigcap_{i=1}^{n} V_i \subset U .$$

The standard topology on $\mathbb{R}$ has a subbasis consisting of rays.

Exercise 4.15. *Let $\mathcal{S}$ be the following collection of subsets of $\mathbb{R}$: $\{x \mid x < a$ for some $a \in \mathbb{R}\}$ and $\{x \mid a < x$ for some $a \in \mathbb{R}\}$. Then $\mathcal{S}$ is a subbasis for $\mathbb{R}$ with the standard topology.*

As with bases, we also want to answer the question of when a given collection $\mathcal{S}$ of subsets of a set X is a subbasis for *some* topology on X.

Theorem 4.16. *Suppose X is a set and $\mathcal{S}$ is a collection of subsets of X. Then $\mathcal{S}$ is a subbasis for some topology on X if and only if every point of X is in some element of $\mathcal{S}$.*

The preceding theorem can thus be used to describe a topology by presenting a subbasis that generates it.

Exercise 4.17. *Let $\mathcal{S}$ be the following collection of subsets of $\mathbb{R}$: $\{x \mid x < a$ for some $a \in \mathbb{R}\}$ and $\{x \mid a \leq x$ for some $a \in \mathbb{R}\}$. For what topology on $\mathbb{R}$ is $\mathcal{S}$ a subbasis?*

4.3 Order Topology

Definition. Let X be a set totally ordered by $<$. Let $\mathcal{B}$ be the collection of all subsets of X that are any of the following forms:

$$\{x \in X \mid x < a\} \quad \text{or} \quad \{x \in X \mid a < x\} \quad \text{or} \quad \{x \in X \mid a < x < b\}$$

for $a, b \in X$. Then $\mathcal{B}$ is a basis for a topology $\mathcal{T}$, called the **order topology** on X.

Exercise 4.18. *Let X be a set totally ordered by $<$. Let $\mathcal{S}$ be the collection of sets of the following forms*

$$\{x \in X \mid x < a\} \quad \textit{or} \quad \{x \in X \mid a < x\}$$

for $a \in X$. Then $\mathcal{S}$ forms a subbasis for the order topology on X.

Exercise 4.19. *Verify that the order topology on $\mathbb{R}$ with the usual $<$ order is the standard topology on $\mathbb{R}$.*

Definition. Given sets A and B, their **product** (or **Cartesian product**) $A \times B$ is the set of all ordered pairs (a, b) such that $a \in A$ and $b \in B$. If A and B are totally ordered by $<_A$ and $<_B$, respectively, then the **dictionary order** or **lexicographic order** $<$ on $A \times B$ is specified by defining $(a_1, b_1) < (a_2, b_2)$ if $a_1 <_A a_2$, or if $a_1 = a_2$ and $b_1 <_B b_2$.

Example. The square $[0, 1] \times [0, 1]$ with the lexicographic order and its associated order topology is called the **lexicographically ordered square**.

To understand the topology of the lexicographically ordered square, we can gain intuition by visualizing open sets.

Exercise 4.20. *Draw pictures of various open sets in the lexicographically ordered square.*

Various subsets of the lexicographically ordered square have surprising closures. The next exercise is instructive; be sure to do it, and exclaim, "Oh" or "Wow" when you discover the surprises.

Exercise 4.21. *In the lexicographically ordered square find the closures of the following subsets:*

$$\begin{aligned}
A &= \left\{ \left(\frac{1}{n}, 0 \right) \mid n \in \mathbb{N} \right\}. \\
B &= \left\{ \left(1 - \frac{1}{n}, \frac{1}{2} \right) \mid n \in \mathbb{N} \right\}. \\
C &= \{(x, 0) \mid 0 < x < 1\}. \\
D &= \left\{ \left(x, \frac{1}{2} \right) \mid 0 < x < 1 \right\}. \\
E &= \left\{ \left(\frac{1}{2}, y \right) \mid 0 < y < 1 \right\}.
\end{aligned}$$

Recall that an ordered set is well-ordered if every non-empty subset has a least element. Let's investigate the lexicographic ordering of the countable product of copies of $\mathbb{N}$.

Exercise 4.22. *Assume that $\mathbb{N}$ has the usual order. Let $\mathbb{N}^\omega$ denote the Cartesian product of a countable number of copies of the space $\mathbb{N}$. It can be endowed with the dictionary order in a natural way. Show that $\mathbb{N}^\omega$ with the dictionary order topology is uncountable, is not well-ordered, and any set that does not have a least element does have a limit point.*

The remaining theorems in this section depend on knowing about the ordinal numbers (see Section 2.6). The ordinal numbers can be regarded as topological spaces by giving them the order topology.

Example. For each ordinal α, the collection of predecessors of α with the order topology forms a space called α.

Every infinite set of ordinals less than the first uncountable ordinal has a limit point.

Theorem 4.23. *Consider the topological space ω_1 consisting of all ordinals less than ω_1, the first uncountable ordinal, with the order topology. Let A be an infinite set of ordinals in ω_1. Then there is an ordinal $\beta < \omega_1$ that is a limit point of A.*

The next theorem about the space ω_1 shows a surprising feature of unbounded closed sets.

Theorem 4.24. *Let A and B be unbounded closed sets in the topological space ω_1. Then $A \cap B \neq \varnothing$.*

4.4 Subspaces

If $(X, \mathcal{T})$ is a topological space and Y is a subset of X, then there is a natural topology that the topology $\mathcal{T}$ induces on Y, formed by intersecting open sets of X with the subset Y.

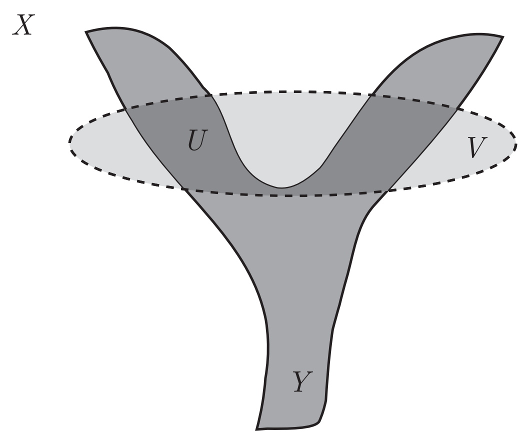


Figure 4.2: U is open in the subspace topology on Y.

Definition. Let $(X, \mathcal{T})$ be a topological space. For $Y \subset X$, the collection

$$\mathcal{T}_Y = \{U \mid U = V \cap Y \text{ for some } V \in \mathcal{T}\}$$

is a topology on Y called the **subspace topology**. It is also called the **relative topology** on Y **inherited** from X. The space $(Y, \mathcal{T}_Y)$ is called a (topological) **subspace** of X. If $U \in \mathcal{T}_Y$ we say U is **open in** Y.

Theorem 4.25. *Let $(X, \mathcal{T})$ be a topological space and $Y \subset X$. Then the collection of sets $\mathcal{T}_Y$ is in fact a topology on Y.*

Exercise 4.26. *Consider $Y = [0, 1)$ as a subspace of $\mathbb{R}_{\text{std}}$. In Y, is the set $[1/2, 1)$ open, closed, neither, or both?*

Exercise 4.27. *Consider a subspace Y of the topological space X. Is every subset $U \subset Y$ that is open in Y also open in X?*

Closed sets in a space are related to closed sets in a subspace in essentially in the same way open sets in a space are related to open sets in a subspace.

Theorem 4.28. *Let $(Y, \mathcal{T}_Y)$ be a subspace of $(X, \mathcal{T})$. A subset $C \subset Y$ is closed in $(Y, \mathcal{T}_Y)$ if and only if there is a set $D \subset X$, closed in $(X, \mathcal{T})$, such that $C = D \cap Y$.*

Corollary 4.29. *Let $(Y, \mathcal{T}_Y)$ be a subspace of $(X, \mathcal{T})$. A subset $C \subset Y$ is closed in $(Y, \mathcal{T}_Y)$ if and only if $\mathrm{Cl}_X(C) \cap Y = C$.*

What is the connection between a basis for a space and a basis for the subspace?

Theorem 4.30. *Let $(X, \mathcal{T})$ be a topological space, and (Y, T_Y) be a subspace. If $\mathcal{B}$ is a basis for $\mathcal{T}$, then $\mathcal{B}_Y = \{B \cap Y | B \in \mathcal{B}\}$ is a basis for $\mathcal{T}_Y$.*

Exercise 4.31. *Consider the following subspaces of the lexicographically ordered square:*

1. $D = \left\{\left(x, \frac{1}{2}\right) \mid 0 < x < 1\right\}$.
2. $E = \left\{\left(\frac{1}{2}, y\right) \mid 0 < y < 1\right\}$.
3. $F = \{(x, 1) \mid 0 < x < 1\}$.

As sets they are all lines. Describe their relative topologies, especially noting any connections to topologies you have seen already.

4.5 Product Spaces

There's a natural way to project a Cartesian product to each of its coordinates.

Definition. Let X and Y be two sets. The **projection functions** $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are defined by $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$.

If sets X and Y have topologies, there is a natural topology on $X \times Y$.

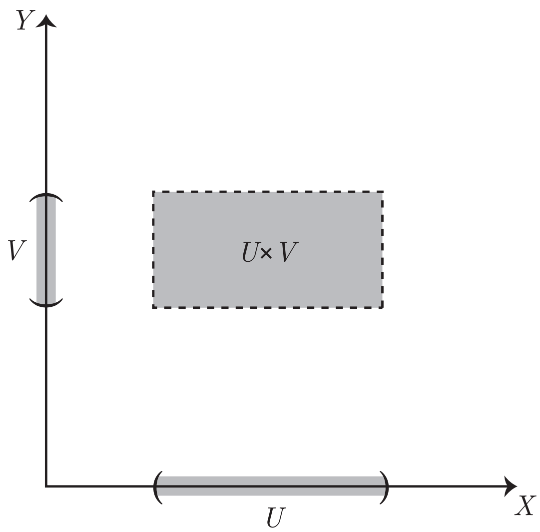


Figure 4.3: A basic open set $U \times V$ in $X \times Y$, where U is open in X and V open in Y.

Definition. Suppose X and Y are topological spaces. The **product topology** on the product $X \times Y$ is the topology whose basis is all sets of the form $U \times V$ where U is an open set in X and V is an open set in Y.

Exercise 4.32. *Verify that the collection of basic open sets above satisfies the conditions of Theorem 4.3, thus confirming that this collection is the basis for a topology.*

Thus open sets of the product topology look like unions $\cup_{\alpha\in\lambda}(U_\alpha \times V_\alpha)$ where λ is a (possibly uncountable) indexing set.

Exercise 4.33. *Draw examples of basic and arbitrary open sets in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ using the standard topology on $\mathbb{R}$. Find (i) an open set in $\mathbb{R} \times \mathbb{R}$ that is not the product of open sets, and (ii) a closed set in $\mathbb{R} \times \mathbb{R}$ that is not the product of closed sets.*

Exercise 4.34. *Is the product of closed sets closed?*

Exercise 4.35. *Show that the product topology on $X \times Y$ is the same as the topology generated by the subbasis of inverse images of open sets under the projection functions, that is the subbasis is $\{\pi_X^{-1}(U) \mid U$ open in $X\} \cup \{\pi_Y^{-1}(V) \mid V$ open in $Y\}$.*

Exercise 4.36. *Using the standard topology on $\mathbb{R}$, is the product topology on $\mathbb{R}\times\mathbb{R}$ the same as the standard topology on $\mathbb{R}^2$?*

Think about how you might generalize the above constructions to define a topology on the Cartesian product of several sets, such as $X \times Y \times Z$. As the above exercises suggest, there are two strategies: one would be to take the product of open sets in each coordinate, and the other is to define a topology by a subbasis of inverse images of open sets under the projection functions. Either approach produces the same topology for a finite product.

But the two approaches give different topologies for products of *infinitely* many spaces. And the most natural topology for an infinite product is perhaps not what one might first expect—surprisingly, the subbasis approach produces a more natural topology.

First we should define what we mean by an infinite product of spaces. To do this, first think about a finite product, such as $X_1 \times X_2 \times X_3$. A shorthand notation for this product is $\Pi_{i\in\{1,2,3\}}X_i$. This Π notation is similar to the Σ notation for an indexed sum (and, in fact, the sound of the Greek letter Pi reminds us of a Product just as Sigma reminds us of a Sum.)

A generic element of $\Pi_{i\in\{1,2,3\}}X_i$ is (x_1, x_2, x_3) where $x_1 \in X_1$, $x_2 \in X_2$, and $x_3 \in X_3$. However, another way to view (x_1, x_2, x_3) is as a *function* f that takes in the coordinate number i and spits out the coordinate x_i, in other words, we can think of a generic element of $\Pi_{i\in\{1,2,3\}}X_i$ as a function

$$f : \{1, 2, 3\} \to \cup_{i=1}^{3} X_i$$

such that $f(i) \in X_i$. This function's domain is the index set $\{1, 2, 3\}$ and its codomain is the union of the factors X_i.

The value of thinking of a product as a function is that it is now apparent how to define an infinite product; namely, we can just replace the domain $\{1, 2, 3\}$ by an infinite index set.

Definition. Let $\{X_\alpha\}_{\alpha\in\lambda}$ be a collection of topological spaces. The **product** $\prod_{\alpha\in\lambda} X_\alpha$, or **Cartesian product**, is the set of functions

$$\{f : \lambda \to \bigcup_{\alpha\in\lambda} X_\alpha \mid \text{for all } \alpha \in \lambda, f(\alpha) \in X_\alpha\}.$$

Here $f(\alpha)$ is called the **α-th coordinate** of f. The spaces X_α are sometimes called **factors** of the infinite product. Thus a point in the product may be thought of as a function that associates to each α an element $f(\alpha)$ of the factor X_α.

Lest you think the infinite product construction is just an abstract oddity, we mention that such products are at the heart of some profound ideas in physics. For instance, the set $\ell^2 = \{(z_1, z_2, z_3, ...) \mid z_i \in \mathbb{C},\ \sum_{n=1}^{\infty} |z_n|^2 < \infty\}$ is a subset of the countably infinite product of copies of $\mathbb{C}$. With extra algebraic structure, ℓ^2 is an example of a *Hilbert space*—a complete inner product space. Hilbert spaces arise in quantum mechanics.

How do we put a topology on a (possibly infinite) product?

Definition. For each β in λ, define the **projection function** $\pi_\beta : \prod_{\alpha\in\lambda} X_\alpha \to X_\beta$ by $\pi_\beta(f) = f(\beta)$. We define the **product topology** on $\prod_{\alpha\in\lambda} X_\alpha$ to be the one generated by the subbasis of sets of the form $\pi_\beta^{-1}(U_\beta)$ where U_β is open in X_β.

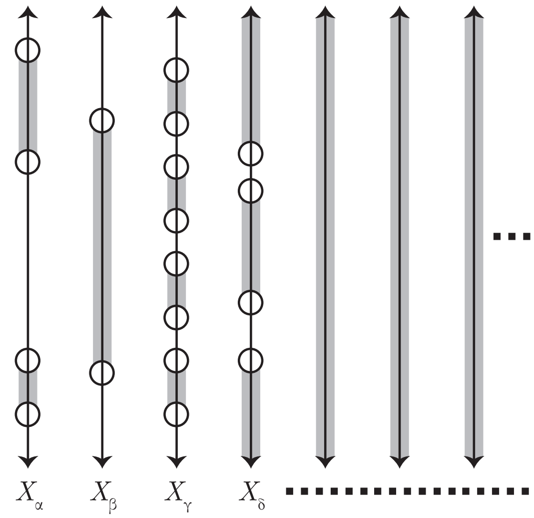


Figure 4.4: A basic open set in the product topology restricts points in only finitely many factors.

Exercise 4.37. *A* basis *for the* product topology *on* $\Pi_{\alpha\in\lambda}X_\alpha$ *is the collection of all sets of the form* $\Pi_{\alpha\in\lambda}U_\alpha$ *where* U_α *is open in* X_α *for each* α *and* $U_\alpha = X_\alpha$ *for all but finitely many* α.

Why do you think the product topology is sometimes also called the *finite gate topology*?

Example. If each factor X_α is the same space X, then the product space $\Pi_{\alpha\in\lambda}X_\alpha$ is sometimes denoted X^λ, where λ is the index set of the product or an ordinal number representing it. Thus $\mathbb{R}^3$ is the three-fold product of copies of $\mathbb{R}$. The countable product of copies of $\mathbb{R}$ is sometimes denoted $\mathbb{R}^\mathbb{N}$ or $\mathbb{R}^\omega$, where ω represents ω_0, the first, countable ordinal.

Example. The space $\{0,1\}^A = \Pi_{a\in A}\{0,1\}$ is a product of discrete two-point spaces, one for each element a in A. An element of this space is a function $f : A \to \{0,1\}$. Note that for $a \in A$, the projection function π_a takes the function f to $f(a)$, and note that $\{0,1\}$ has the discrete topology. Consider these subsets of $\{0,1\}^A$:

$$U(a,0) = \pi_a^{-1}(0) = \{f \in \{0,1\}^A \mid f(a) = 0\}.$$

$$U(a,1) = \pi_a^{-1}(1) = \{f \in \{0,1\}^A \mid f(a) = 1\}.$$

Then $\mathcal{S} = \{U(a,\delta) \mid a \in A, \delta \in \{0,1\}\}$ is a subbasis for the product topology on $\{0,1\}^A$.

This space $\{0,1\}^A$ is also sometimes written 2^A, which you may recall is the notation we reserved for the power set of A, because there is a natural bijection between $\{0,1\}^A$ and the power set of A. In particular, each function $f : A \to \{0,1\}$ can be bijectively associated with the subset of all points of A where f is 1. This bijection naturally endows the power set 2^A with a topology that comes from the product topology on $\{0,1\}^A$. Then for each a in A, the subbasic set $U(a,1)$ corresponds to the set of all subsets of A that contain a and $U(a,0)$ corresponds to the set of all subsets of A that do not contain a. So in thinking about 2^A, it may be helpful to pass back and forth between viewing elements of 2^A as subsets of A or as functions $f : A \to \{0,1\}$.

Exercise 4.38. *Let $\mathcal{T}$ be the topology on 2^X with basis generated by the subbasis $\mathcal{S}$.*

1. *Every basic open set in 2^X is both open and closed.*
2. *Show that if a collection of subbasic open sets of 2^X has the property that every point of 2^X lies in at least one of those subbasic open sets, then there are two subbasic open sets in that collection such that every point of 2^X lies in one of those two subbasic sets.*
3. *Show that if a collection of basic open sets of 2^X has the property that every point of 2^X lies in at least one of those basic open sets, then there are a finite number of basic open sets in that collection such that every point of 2^X lies in one of those basic sets.*

Exercise 4.39. *In the product space $2^{\mathbb{R}}$, what is the closure of the set Z consisting of all elements of $2^{\mathbb{R}}$ that are 0 on every rational coordinate, but may be 0 or 1 on any irrational coordinate? Equivalently, thinking of $2^{\mathbb{R}}$ as subsets of $\mathbb{R}$, what is the closure of the set Z consisting of all subsets of $\mathbb{R}$ that do not contain any rational?*

Recall from Theorem 3.30 that if a sequence $(x_i)_{i\in\mathbb{N}}$ converges to $x \in X$ and each $x_i \in A$, then $x \in \overline{A}$.

> **Effective Thinking Principle.** *Consider the Converse.* We can deepen our understanding of a theorem by considering whether or not the converse is true.

Exercise 4.40. *Find a subset A of $2^{\mathbb{R}}$ and a limit point x of A such that no sequence in A converges to x. For an even greater challenge, determine whether you can find such an example if A is countable.*

You may have been wondering why the product topology, with a basis of sets with finitely many restricted coordinates as in Exercise 4.37, is more natural than a topology with a basis of sets in which we allow all coordinates to be restricted, i.e., where basic open sets are *boxes*: products of open sets in each coordinate? This alternative construction yields another possible topology on an infinite product, called the *box topology*.

Definition. A *basis* for the **box topology** on $\Pi_{\alpha\in\lambda}X_\alpha$ is the collection of all sets of the form $\Pi_{\alpha\in\lambda}U_\alpha$ where U_α is open in X_α for each α.

Every open set in the product topology is open in the box topology, but not vice versa. Thus the box topology is finer than the product topology.

Theorems and exercises throughout the coming chapters will demonstrate that the product topology is a nicer and more natural topology than the box topology.

Exercise 4.41. *Let $\mathbb{R}^\omega$ be the countable product of copies of $\mathbb{R}$. So every point in $\mathbb{R}^\omega$ is a sequence $(x_1, x_2, x_3, ...)$. Let $A \subset \mathbb{R}^\omega$ be the set consisting of all points with only positive coordinates. Show that in the product topology, $\mathbf{0} = (0, 0, 0, ...)$ is a limit point of the set A, and there is a sequence of points in A converging to $\mathbf{0}$. Then show that in the box topology, $\mathbf{0} = (0, 0, 0, ...)$ is a limit point of the set A, but there is no sequence of points in A converging to $\mathbf{0}$.*

Exercise 4.42. *Show that the set $2^{\mathbb{N}}$ in the box topology is a discrete space, whereas the set $2^{\mathbb{N}}$ in the product topology has no isolated points.*

4.6 A Bounty of New Spaces

Before starting this chapter, our collection of topological spaces was rather limited. The concepts of bases, subbases, products, and subspaces gave us an explosion of new topological spaces to explore and enjoy. Giving the familiar real line the funky lower limit topology produced a new space nicknamed $\mathbb{R}^1_{bad}$. The lexicographically ordered square has unexpected limit points and closures. The Sticky Bubble Topology is full of interesting closed sets that can or cannot be put in disjoint open sets. The topology on $\mathbb{Z}$ using arithmetic progressions gave us a proof of the infinitude of primes that Euclid never dreamed of.

This ever-expanding collection of topological spaces suggests the topological concepts we will explore in the chapters ahead. The 2-Headed Snake had a pair of points that cannot be put in disjoint open sets. The Sticky Bubble Topology presented challenges about putting pairs of disjoint sets into disjoint open sets. The product topology gave us examples suggesting properties of spaces that differentiate one topological space from another. Concepts arise from examples and questions. This chapter has enriched our stock of examples and opened up to us a world of questions that we will begin exploring in the next chapter.

Chapter 5

Separation Properties: Separating This from That

Recall that one of the main motivations for topology is to understand the essential properties of a space that make ideas from calculus work, such as convergence and continuity. Specifically, we want to understand those ideas without referring to the concept of distance. We will discover that some pertinent features of the topology revolve around the question of whether open sets exist in the topology that separate various subsets of the space from one another. Let's consider the idea of convergence from calculus, for example. A convergent sequence of real numbers cannot converge to two different values x and y. Why? Because we can find disjoint open intervals around x and y. The tail of the convergent sequence could not simultaneously be in both of those intervals, so the sequence cannot converge to both x and y. The key step in this little proof is the existence of a pair of disjoint open intervals around x and y.

The *separation axioms* explore the existence of pairs of open sets that separate objects in the space. In our example about convergence, we separated pairs of points, but separating various other kinds of sets, such as closed sets, is also of interest. This exploration gives rise to a hierarchy of so-called separation properties.

Some of the names of these separation properties consist of a capital T with a subscript. Those unimaginative names have historical significance. In 1914, Felix Hausdorff wrote one of the foundational books about topology called *Grundzüge der Mengenlehre* (*Elements of Set Theory*). In that book, Hausdorff defined what a topological space is and then defined and explored properties of topological spaces. Since he wrote in German, he naturally used the German word for separation, namely, *Trennung*. He defined the separation properties in a numbered list, so the separation properties are often referenced by the numbers appearing in that list: T_1, T_2, and so on. These names seem rather unimaginative; however, the alternative names associated with these separation properties are, if anything, even less informative. 'Regular' and 'normal' are two other words used to describe spaces with defined separation properties. These bland names do not help us to remember their meaning, but, as generations have before you, you too will soon happily converse about regular spaces and normal spaces without having any concern about the anemic choice of vocabulary.

5.1 Hausdorff, Regular, and Normal Spaces

Now back to the mathematics. Here are some separation properties.

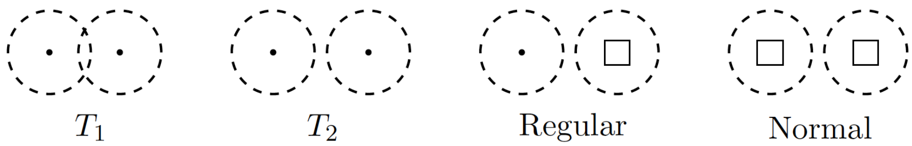


Figure 5.1: Four separation properties.

Definition. Let $(X, \mathcal{T})$ be a topological space.

1. X is a T_1**-space** if and only if for every pair x, y of distinct points there are open sets U, V such that U contains x but not y, and V contains y but not x.

2. X is **Hausdorff**, or a T_2**-space**, if and only if for every pair x, y of distinct points there are *disjoint* open sets U, V such that $x \in U$ and $y \in V$.

3. X is **regular** if and only if for every point $x \in X$ and closed set $A \subset X$ not containing x, there are disjoint open sets U, V such that $x \in U$ and $A \subset V$. A T_3**-space** is any space that is both T_1 and regular.

4. X is **normal** if and only if for every pair of disjoint closed sets A, B in X, there are *disjoint* open sets U, V such that $A \subset U$ and $B \subset V$. A T_4**-space** is any space that is both T_1 and normal.

The most important property of a T_1-space is that points are closed.

Theorem 5.1. *A space $(X, \mathcal{T})$ is T_1 if and only if every point in X is a closed set.*

For the topological spaces that you know, it is fun to determine which separation axioms they satisfy. We will soon ask you to construct a chart listing examples along the top and separation properties down the side and in each box answer the question of whether the example of the column has the property of the row. Here are a few of those exercises to warm up with.

Exercise 5.2. *Let X be a space with the finite complement topology. Show that X is T_1.*

Exercise 5.3. *Show that $\mathbb{R}_{\text{std}}$ is Hausdorff.*

Exercise 5.4. *Show that $\mathbb{H}_{\text{bub}}$ is regular.*

Exercise 5.5. *Show that $\mathbb{R}_{\text{LL}}$ is normal.*

Let's explore some of these separation properties in our familiar example: $\mathbb{R}^2$ with the standard topology. Part (3) of the following exercise points out that a plausible approach to proving the normality of $\mathbb{R}^2$ does not work; however, as you will show in part (4), $\mathbb{R}^2$ is, in fact, normal.

Exercise 5.6.

1. *Consider $\mathbb{R}^2$ with the standard topology. Let $p \in \mathbb{R}^2$ be a point not in a closed set A. Show that $\inf\{d(a,p) \mid a \in A\} > 0$. (Recall that $\inf E$ is the greatest lower bound of a set of real numbers E.)*
2. *Show that $\mathbb{R}^2$ with the standard topology is regular.*
3. *Find two disjoint closed subsets A and B of $\mathbb{R}^2$ with the standard topology such that*
$$\inf\{d(a,b) \mid a \in A \text{ and } b \in B\} = 0.$$
4. *Show that $\mathbb{R}^2$ with the standard topology is normal.*

The separation axioms are related to one another in a hierarchical fashion suggested by the T_i nomenclature.

Theorem 5.7.

1. *A T_2-space (Hausdorff) is a T_1-space.*
2. *A T_3-space (regular and T_1) is a Hausdorff space, that is, a T_2-space.*
3. *A T_4-space (normal and T_1) is regular and T_1, that is, a T_3-space.*

A word of caution as you read the literature: topologists sometimes differ on the definitions of *regular* and *normal* spaces. Because of the above theorem, some books include the T_1 condition as part of the definitions of regular and normal spaces, thereby making a normal space automatically regular, and a regular space automatically Hausdorff. However, we will use the more common definitions of regular and normal that do not include the T_1 condition.

The following theorems give alternative characterizations of regularity and normality that can sometimes be useful.

Theorem 5.8. *A topological space X is regular if and only if for each point p in X and open set U containing p there exists an open set V such that $p \in V$ and $\overline{V} \subset U$.*

Theorem 5.9. *A topological space X is normal if and only if for each closed set A in X and open set U containing A there exists an open set V such that $A \subset V$, and $\overline{V} \subset U$.*

The next theorem tells us that in normal spaces, closed sets can actually be separated by open sets that are not only disjoint, but whose closures are also disjoint.

Theorem 5.10. *A topological space X is normal if and only if for each pair of disjoint closed sets A and B, there are disjoint open sets U and V such that $A \subset U$, $B \subset V$, and $\overline{U} \cap \overline{V} = \varnothing$.*

The next characterization of normality is affectionately known as the "Incredible Shrinking Theorem".

Theorem 5.11 (The Incredible Shrinking Theorem)**.** *A topological space X is normal if and only if for each pair of open sets U, V such that $U \cup V = X$, there exist open sets U', V' such that $\overline{U'} \subset U$ and $\overline{V'} \subset V$, and $U' \cup V' = X$.*

The previous theorem can actually be extended to a yet more Incredible Shrinking Theorem. Using the Well-Ordering Principle and a technique called transfinite induction, one can prove the following: if X is normal, then for every collection $\{U_\alpha\}_{\alpha\in\lambda}$ of open sets such that $\cup_{\alpha\in\lambda} U_\alpha = X$ and each point of X is in only finitely many of the U_α's, there exist open sets $\{U'_\alpha\}_{\alpha\in\lambda}$ such that for each $\alpha \in \lambda$, we have $\overline{U'_\alpha} \subset U_\alpha$ and $\cup_{\alpha\in\lambda} U'_\alpha = X$.

Effective Thinking Principle. *Find Examples to Distinguish Concepts.* We have now been introduced to a collection of new definitions. A good way to get accustomed to new definitions is to find examples that manifest the differences in the various properties.

Exercise 5.12.

1. *Describe an example of a topological space that is T_1 but not T_2.*
2. *Describe an example of a topological space that is T_2 but not T_3.*
3. *Describe an example of a topological space that is T_3 but not T_4.*

Exercise 5.13. *Construct a table, listing our previous examples of topological spaces as column titles, and listing the separation properties as row titles. In each box, answer the question of whether the example of the column has the property of the row. Here are the spaces to use as column titles:*

1. $\mathbb{R}_{std}$
2. $\mathbb{R}^n_{std}$

3. *indiscrete topology*
4. *discrete topology*
5. *finite complement topology*
6. *countable complement topology*
7. *lower limit topology,* $\mathbb{R}_{\mathrm{LL}}$
8. *double headed snake,* $\mathbb{R}_{+00}$
9. $\mathbb{R}_{har}$
10. *Sticky Bubble Topology,* $\mathbb{H}_{\mathrm{bub}}$
11. *arithmetic progression topology,* $\mathbb{Z}_{arith}$
12. *lexicographically ordered square*
13. 2^X

Here are the properties to use as row titles:

1. T_1
2. *Hausdorff*
3. *regular*
4. *normal*

One of the more challenging problems in this table is the following entry.

Exercise 5.14. *Show that* $\mathbb{H}_{\mathrm{bub}}$ *is not normal.*

In the next sections, we will introduce additional properties and additional examples to add to your chart.

In some cases whole categories of spaces can be dealt with at once. You can add "order topologies" as a column in your chart and use the following theorem to fill in the separation properties of all ordered spaces with the order topology.

Theorem 5.15. *Order topologies are* T_1*, Hausdorff, regular, and normal.*

5.2 Separation Properties and Products

Taking products is a method for creating new topological spaces from existing ones. In your chart of spaces and properties add a column for products. For each property, ask whether if you start with two spaces, each having that property, will the product necessarily have that property? For instance

Theorem 5.16. *Let X and Y be Hausdorff. Then $X \times Y$ is Hausdorff.*

Theorem 5.17. *Let X and Y be regular. Then $X \times Y$ is regular.*

For the above problem, Theorem 5.8 may make your life easier.

It turns out the the product of two normal spaces need not be normal. Recall that $\mathbb{R}_{\mathrm{LL}}$ is normal.

Exercise 5.18. *Show that $\mathbb{R}_{\mathrm{LL}} \times \mathbb{R}_{\mathrm{LL}}$ is not normal. It may help to consider the "negative diagonal" line L.*

5.3 A Question of Heredity

A natural question to ask is: given a topological space satisfying certain properties, what properties do its subspaces "inherit"?

Definition. Let P be a topological property (such as T_1, Hausdorff, etc.). A topological space X is **hereditarily** P if and only if for each subspace Y of X, the space Y has property P when Y is given the relative topology from X.

Theorem 5.19. *Every Hausdorff space is hereditarily Hausdorff.*

Theorem 5.20. *Every regular space is hereditarily regular.*

However, *not* every normal space is hereditarily normal(!)— there actually exist normal spaces that have non-normal subspaces.

Exercise 5.21.

1. *Prove that the space $2^{\mathbb{R}}$ is normal.*
2. *Prove that if you remove a single point from $2^{\mathbb{R}}$, the resulting subspace is not normal.*

This exercise asks you to prove two things, both of which are difficult. To prove that $2^{\mathbb{R}}$ is normal, you may find Exercise 4.38 useful. Or you could wait until we get to the compactness chapter, after which this result will not be as difficult. Part 2 above is difficult, and you are on your own.

There is a famous example of a space that is normal but not hereditarily normal that goes by the piratical title of the *Tychonoff Plank*.

Example. The **Tychonoff Plank** is the product of two ordinal spaces: $(\omega_0 + 1) \times (\omega_1 + 1)$.

Exercise 5.22. *(Walking the Tychonoff Plank, or Mutiny on the Boundary)*

1. *Show that the Tychonoff Plank is normal.*

2. *Show that the Tychonoff Plank minus the single point (ω_0, ω_1) is not normal.*

Although not every subspace of a normal space is normal, certain subspaces do inherit normality.

Theorem 5.23. *Let A be a closed subset of a normal space X. Then A is normal when given the relative topology.*

Establishing whether a subset of a normal space is normal can be challenging. The next exercise asks you to investigate a particular subset of 2^X. However, this exercise is extremely difficult, and we recommend that you skip it.

Exercise 5.24.

1. *Prove that for any set X, 2^X is normal. (This part is not really different from showing that $2^{\mathbb{R}}$ is normal, which you did in a previous exercise.)*

2. *Recall that there is a one-to-one correspondence between the points of 2^X and subsets of X, as follows: recall that each point of 2^X is a function $f : X \to \{0, 1\}$, so $f^{-1}(1)$ is a subset of X. Let $C \subset 2^X$ consist of those points that take on the value 1 on only a countable set of coordinates, that is, C is the set of functions $f : X \to \{0, 1\}$, for which $f^{-1}(1)$ is countable. Prove that C with the subspace topology is normal.*

In trying to prove that a subspace is normal, we naturally must consider two disjoint sets that are closed and disjoint in the subspace topology. The next exercise contains an observation about such sets.

Exercise 5.25. *Let Y be a subspace of a topological space X, and let A and B be two disjoint closed subsets of Y in the subspace topology. Show that both $\overline{A} \cap B = \varnothing$ and $A \cap \overline{B} = \varnothing$, where the closures are taken in X.*

Such pairs of sets that do not contain limit points of the other set deserve a name.

Definition. Two sets A and B in a space X are **separated** if and only if both $\overline{A} \cap B = \varnothing$ and $A \cap \overline{B} = \varnothing$.

Definition. A space X is **completely normal** if and only if for any two separated sets A and B there exist disjoint open sets U and V such that $A \subset U$ and $B \subset V$. A T_5**-space** is any space that is both T_1 and completely normal.

Theorem 5.26. *The space X is a completely normal space if and only if X is hereditarily normal.*

Ordered spaces, that is, spaces whose topology is generated from an order on the points, are all normal. But as you'll see in the next exercise, the relative topology on subsets of ordered spaces is not necessary generated by the ordering of the points in the subset. So the hereditary normality of ordered spaces is not immediate.

Exercise 5.27.

1. *Recall that $\mathbb{R}$ is an order topology. Find a subset of $\mathbb{R}$ where the subspace topology is not the order topology on the subset.*
2. *Find a line in the lexicographically ordered square whose relative topology is the discrete topology on this line, but this is not the order topology on the subset.*
3. *Notice that $\mathbb{R}_{LL}$ is not an order topology. Find a line in the lexicographically ordered square whose relative topology is the lower limit topology.*

The above exercise shows us that the relative topology on a subset of an ordered space is not necessarily an order topology. Nevertheless, all subspaces of order topologies are normal.

Theorem 5.28. *Order topologies are hereditarily normal.*

5.4 The Normality Lemma

In trying to prove that a topological space is normal, we are faced with the challenge of producing two disjoint open sets that contain a pair of disjoint closed sets. The following theorem describes conditions under which it is possible to construct disjoint open sets around a pair of sets that may or may not be closed.

Theorem 5.29 (The Normality Lemma)**.** *Let A and B be subsets of a topological space X and let $\{U_i\}_{i\in\mathbb{N}}$ and $\{V_i\}_{i\in\mathbb{N}}$ be two collections of open sets such that*

1. *$A \subset \bigcup_{i\in\mathbb{N}} U_i$,*
2. *$B \subset \bigcup_{i\in\mathbb{N}} V_i$,*
3. *for each i in $\mathbb{N}$, $\overline{U}_i \cap B = \varnothing$ and $\overline{V}_i \cap A = \varnothing$.*

Then there exist open sets U and V such that $A \subset U$, $B \subset V$, and $U \cap V = \varnothing$.

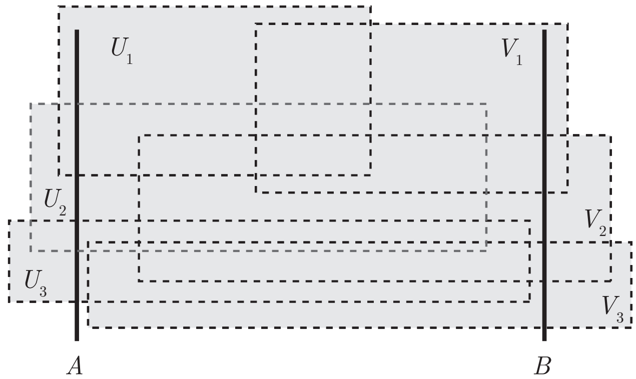


Figure 5.2: The Normality Lemma.

Not all subspaces of normal spaces are normal, but we showed in Theorem 5.23 that closed subspaces of a normal space are normal. The next theorem shows that subspaces of normal spaces that are countable unions of closed sets are also normal.

Theorem 5.30. *If X is normal and $C = \cup_{i\in\mathbb{N}} K_i$ is the union of closed sets K_i in X, then the subspace C is normal.*

Another application of the Normality Lemma shows that countable regular spaces are normal.

Theorem 5.31. *Suppose a space X is regular and countable. Then X is normal.*

Theorem 5.32. *Suppose a space X is regular and has a countable basis. Then X is normal.*

Definition. A space X is **perfectly normal** if and only if X is normal and each closed set A in X is the intersection of countably many open sets.

Theorem 5.33. *Let X be a perfectly normal space. Then X is completely normal.*

5.5 Separating This from That

Topological spaces differ in which pairs of sets can be separated by disjoint open sets. In this section, we saw a whole hierarchy of separation properties. Among T_1 spaces the hierarchy went from T_1 to Hausdorff to regular to normal to completely normal to perfectly normal. When we got to normality in this hierarchy, issues of inheritance of properties to subsets and products became more interesting. Perhaps it requires a special kind of person to enjoy the myriad challenges of determining which examples of spaces exhibit which properties, but we hope you are one of those special people.

The separation properties are characteristics of topological spaces that can help us to distinguish one space from another. As is the case with most any meaningful subject, when we examine topological spaces more closely, we become attentive to increasingly detailed and increasingly interesting features of them. In this chapter we examined spaces from the perspective of how various kinds of subsets can be separated from one another by disjoint open sets. In the next chapters, we will look at features of spaces that focus on other kinds of characteristics. This strategy of exploring topological spaces is a metaphor for one of the most powerful methods of understanding our world more deeply, namely, choosing a perspective and undertaking explorations using that lens; and then choosing another perspective and exploring the same world again using that alternative lens. We have a whole optometrist's storehouse of lenses ahead. Enjoy.

Chapter 6

Countable Features of Spaces: Size Restrictions

In this chapter, we will explore several properties of topological spaces that have to do with countability—countability of a dense set, countability of a basis, countability of a local version of a basis. Recall from Chapter 2 that a *countable* set is one that is finite or has the same cardinality as $\mathbb{N}$.

In some sense, these countability properties impose a restriction on the ”size” of the space by restricting the size (the cardinality) of some of its topological features. Countability affords us a systematic way to study the space, namely, countability allows us to go through an enumerated list, one item after another. We will see that the countability properties often have implications about other topological properties of the space.

For instance, one of the final theorems of the previous chapter asserted that a regular space that is countable must be normal. One way to view that theorem is that the countability property strengthens the separation property of the space: a regular space with only a countable number of points now has further structure—it must be normal. The proof of that theorem relied on creating a step-by-step process that dealt with the points one at a time.

As another example, in this chapter you will prove that in any space with a countable basis, an uncountable set has further structure—it must have a limit point. The proof uses the enumerability of the basis to perform a construction that produces a limit point.

6.1 Separable Spaces, An Unfortunate Name

We introduce the idea of a *dense* subset of a topological space.

Definition. Let A be a subset of a topological space X. Then A is **dense** in X if and only if $\overline{A} = X$.

So you can think of a dense subset A of a space X as a set that permeates the space—every point in X can be approached arbitrarily closely by points of A.

Exercise 6.1. *Show that A is dense in X if and only if every non-empty open set of X contains a point of A.*

Definition. A topological space X is **separable** if and only if X has a countable dense subset.

The choice of the word *separable* for the property described above is an unfortunate one (due to Fréchet), since it is not related to the separation properties we described in the previous section. Nor is it related to the concept of *separated* sets which will be discussed when we learn about connected sets.

You can think of a separable space as one that is not too large, because there is a countable set which every open set intersects, so there is a countable set of points that is "close" to every point in the space.

Exercise 6.2. *Show that $\mathbb{R}_{std}$ is separable. With which of the topologies on $\mathbb{R}$ that you have studied is $\mathbb{R}$ not separable?*

Exercise 6.3. *Add 'separable' as a new property in your chart, and complete your chart by deciding which of the spaces we've studied are separable.*

In exploring a new property, it is natural to ask whether it is inherited by subspaces.

Exercise 6.4. *Find a separable space that contains a subspace that is not separable in the subspace topology.*

Fortunately the property of being separable is better behaved when we take products.

Theorem 6.5. *If X and Y are separable spaces, then $X \times Y$ is separable.*

Theorem 6.6. *The space $2^{\mathbb{R}}$ is separable.*

This theorem is a special case of the following more general fact, which says if the size of the indexing set is relatively small, then the product will have a countable dense set in it.

Exercise 6.7. *Let $\{X_\beta\}_{\beta\in\mu}$ be a collection of separable spaces where $|\mu| \leq 2^{\omega_0}$, then $\prod_{\beta\in\mu} X_\beta$ is separable.*

In a sense, a separable space is not too large since every point in it is a limit point of some countable set. The next, challenging theorem shows us that indeed the cardinality of any separable Hausdorff space is limited.

Exercise 6.8. *If X is a separable, Hausdorff space, then $|X| \leq \left|2^{2^{\mathbb{N}}}\right|$.*

To prove it, you may wish to think of a point in $2^{2^{\mathbb{N}}}$ as a set of subsets of the countable dense set.

6.2 2^{nd} Countable Spaces

Another way to measure the size of a space is to measure the size of a basis for its topology.

Definition. A space X is 2^{nd} **countable** if and only if X has a countable basis.

So in a 2^{nd} countable space every open set can be built up from a countable collection of basic open sets. You may be wondering two things: (1) why do we define 2^{nd} countable before defining 1^{st} countable? and (2) why do we name it 2^{nd} countable instead of something more descriptive, such as "basis-countable"?

Both these questions have the same answer: they were introduced as axioms about topological spaces by Hausdorff in *Mengenlehre*.

Theorem 6.9. *Let X be a 2^{nd} countable space, then X is separable.*

The following exercise asks you to establish the separability or 2^{nd} countability of a few specific spaces. Notice that in order to show that a space is *not* 2^{nd} countable, you must show that *any possible basis* cannot be countable, not just a given basis. Equivalently, in order to show that a given topological space is not 2^{nd} countable, you need to show that any countable collection of open sets in the topology cannot generate the whole topology.

Exercise 6.10.

1. *The space $\mathbb{R}_{std}$ is 2^{nd} countable (and hence separable).*
2. *The space $\mathbb{R}_{LL}$ is separable but not 2^{nd} countable.*
3. *The space $\mathbb{H}_{\text{bub}}$ is separable but not 2^{nd} countable.*

These countability properties suggest that in some way the space is not too large. If you try to cram a "large" set in a space that isn't too large, what do you expect to happen? The set should have a limit point!

Theorem 6.11. *Every uncountable set in a 2^{nd} countable space has a limit point.*

The property of 2^{nd} countability behaves well with respect to heredity and products.

Exercise 6.12. *A 2^{nd} countable space is hereditarily 2^{nd} countable.*

Exercise 6.13. *If X and Y are 2nd countable spaces, then $X \times Y$ is 2nd countable.*

6.3 1^{st} Countable Spaces

> **Effective Thinking Principle.** *Consider Ideas at Different Scales.* One of the strategies of mathematics is, after conceiving of a property at some level, to consider whether that same idea could be valuable at a different scale.

The property of 2^{nd} countability refers to a basis for the whole topology; however, we could consider a local version of that idea. Thinking in that way leads to the concept of a local basis at a point. After we have pinned down the idea of a local basis at a point, then we can give a version of 2^{nd} countability at a local level.

Definition. Let p be a point in a space X. A collection of open sets $\{U_\alpha\}_{\alpha\in\lambda}$ in X is a **neighborhood basis for** p if and only if: (i) each U_α contains p, and (ii) every open set containing p contains some U_α.

Definition. A topological space X is 1^{st} **countable** if and only if every point of X has a countable neighborhood basis.

The next theorem shows the relationship between 2^{nd} countability and 1^{st} countability.

Theorem 6.14. *Let X be a 2^{nd} countable space. Then X is 1^{st} countable.*

> **Effective Thinking Principle.** *Look for Implications of Definitions.* Sometimes a definition immediately implies an apparently stronger version or versions of the definition. Seek them out and formulate them.

One of the simplifying insights about countable neighborhood bases is that they can be chosen to be nested, that is, they can be indexed by the natural numbers in such a way that each subsequent set is a subset of the previous one.

Theorem 6.15. *If X is a topological space, $p \in X$, and p has a countable neighborhood basis, then p has a nested countable neighborhood basis.*

The next exercise asks you to investigate the 1^{st} countability properties of some specific spaces.

Exercise 6.16. 1. *The space $\mathbb{R}_{LL}$ is 1^{st} countable.*

2. *The space $\mathbb{H}_{\text{bub}}$ is 1^{st} countable.*

3. *The space $2^{\mathbb{R}}$ is not 1^{st} countable.*

Exercise 6.17. *You may as well extend your table of spaces and properties by adding new rows for the properties 1^{st} countable and 2^{nd} countable and determining those properties for each of your spaces.*

One important feature of a 1^{st} countable space is that in such a space, every limit point is "reachable" by some convergent sequence. Recall from Exercises 4.40 and 4.41 that there are spaces in which limit points are not accessible by convergent sequences.

Theorem 6.18. *Suppose x is a limit point of the set A in a 1^{st} countable space X. Then there is a sequence of points $\{a_i\}_{i\in\mathbb{N}}$ in A that converges to x.*

The property of 1^{st} countability also behaves well with respect to heredity and products.

Exercise 6.19. *A 1^{st} countable space is hereditarily 1^{st} countable.*

Exercise 6.20. *If X and Y are 1st countable spaces, then $X \times Y$ is 1st countable.*

6.4 The Souslin Property

For fun, we include some theorems about the Souslin property, but they are not central for understanding the main ideas that follow in subsequent chapters.

Definition. A space X has the **Souslin property** if and only if X does *not* contain an uncountable collection of disjoint open sets.

Exercise 6.21. *Show that the real line with the standard topology is Souslin.*

The previous exercise is really an example of the following theorem.

Theorem 6.22. *A separable space has the Souslin property.*

You might think that having the Souslin property is basically the same as being separable; however, there are counterexamples. The spaces of the form 2^X might be places to look for such examples.

Theorem 6.23. *For any set X, the topological space 2^X has the Souslin property.*

Exercise 6.24. *Find a Souslin space that is not separable.*

Theorem 6.23 is a special case of a more general result, namely:

Theorem. *Let $\{X_\beta\}_{\beta\in\mu}$ be a collection of separable spaces, then $\prod_{\beta\in\mu} X_\beta$ is Souslin.*

The Souslin property is part of an investigation about basic properties of the real line that turn out to push us to the boundary of set theory and beyond. Products of Souslin spaces take us to perhaps even weirder realms. It turns out that the question of whether the product of Souslin spaces is Souslin is independent of the standard axioms of set theory.

6.5 Count on it

This chapter invited you to explore the implications of restricting the size (the cardinality) of some features of spaces. You considered spaces that have a countable dense set (separable spaces); spaces that have a countable basis (2^{nd} countable spaces); spaces where each point has a local countable basis (1^{st} countable spaces); and spaces that do not contain an uncountable collection of disjoint open sets (Souslin spaces).

The exploration of these countability properties of spaces demonstrated great strategies for creating and learning mathematical ideas. When you saw a new idea, you connected it with previous ideas. For example, once a property was identified, you investigated whether or not subspaces or products of those spaces would still have that property. After a property was identified, such as 2^{nd} countability, you considered whether a local version would be meaningful (in this case, the concept of 1^{st}) and explored its consequences.

The special feature of countability is that a countable set can be listed in order like the natural numbers. That ordering gives you the option to deal with the situation one step at a time. In this chapter, you thought about countable features of spaces. In the next chapter, you will consider a version of size restriction of spaces associated with covering spaces with open sets.

Chapter 7

Compactness: The Next Best Thing to Being Finite

The next properties we will study are the "covering" properties, so called because they involve collections of open sets (called, appropriately, "open covers") that cover the space or a subset of the space. Like the countability properties, these covering properties in some sense give us a measure of the size of the space in relation to its topology.

Open covers of topological spaces or subsets of topological spaces arise naturally. Frequently, we are presented with a scenario in which every point of the space or a subset of the space has an open set containing it with some property or other. Many proofs that involve open coverings become accessible if infinite open covers can be replaced by sub-collections of open sets that still cover but are finite.

For example, in a Hausdorff space, if you are given a point p and a set A not containing p, then for every point $q \in A$, the Hausdorff property assures you that there exist disjoint open sets U_q and V_q with $p \in U_q$ and $q \in V_q$. The collection of open sets $\{V_q\}_{q \in A}$ is an open cover of A. In general, it is impossible to put p and A into disjoint open sets; however, if there were a finite number of those open sets V_q that completely covered A, then you would be able to put p and A into disjoint open sets. (Do you see why?) Such a finite collection of open sets V_q would make A's behavior like a set with only a finite number of points, and you would be able to separate p from A. In other words, if you somehow knew that the open cover $\{V_q\}_{q \in A}$ had a finite sub-collection that also contained every point of A, then you could succeed in separating p from A.

As in the example above, replacing infinite open covers by finite sub-collections of open sets that still cover frequently allows you to do something that you would not otherwise be able to do. One strategy of mathematics is to create a concept that captures desirable features that you would like to have available in trying to prove a theorem or set of theorems. For example, the concepts of differentiability or continuity focus our attention on functions with desirable properties that allow us to deduce results that are not true for arbitrary functions. As you will see in this chapter, many important topological spaces have this desirable property that every open cover has a sub-cover that is finite.

In this chapter you will explore open covers of topological spaces. You will define concepts related to these open covers and will discover the rich world of covering properties and the consequences that emerge.

7.1 Compact Sets

Definition. Let A be a subset of X and let $\mathcal{C} = \{C_\alpha\}_{\alpha\in\lambda}$ be a collection of subsets of X. Then $\mathcal{C}$ is a **cover** of A if and only if $A \subset \bigcup_{\alpha\in\lambda} C_\alpha$. The collection $\mathcal{C}$ is an **open cover** of A if and only if $\mathcal{C}$ is a cover of A and each C_α is open. A **subcover** $\mathcal{C}'$ of a cover $\mathcal{C}$ of A is a subcollection of $\mathcal{C}$ whose elements form a cover of A.

For instance, the open sets $\{(-n, n)\}_{n\in\mathbb{N}}$ form an open cover of $\mathbb{R}$. A subcover of this cover is $\{(-n, n)\}_{n\geq 5}$, because these sets still cover all of $\mathbb{R}$.

Definition. A space X is **compact** if and only if every open cover of X has a finite subcover.

So for instance, in the example above, does the open cover $\{(-n, n)\}_{n\in\mathbb{N}}$ of $\mathbb{R}$ have a *finite* subcover? If you think about it, you may see that this cover has no finite subcover, because any finite collection of these sets has a largest set that contains all the others, and points outside this set are not covered by this finite collection. So $\mathbb{R}$ cannot be compact, because not every open cover has a finite subcover.

Be careful: many students new to the concept think the definition of compactness says a space is compact if and only if it has a finite cover. No! If having a finite open cover were the definition, the concept would be useless, because every space has a finite open cover—namely, just cover the space X by one set: X itself, which is open.

To show a space is compact, you must prove that for any open cover that you are given, that is, for every possible open cover, you can prove that that open cover has a finite subcover.

Theorem 7.1. *Let X be a finite topological space. Then X is compact.*

A compact space may not be finite, but in some respects, it is the next best thing. Compact spaces enjoy many properties that finite sets do. For instance, as we shall see, not all subsets of real numbers have a maximum, but finite subsets and compact subspaces do.

Theorem 7.2. *Let C be a compact subset of $\mathbb{R}_{std}$. Then C has a maximum point, that is, there is a point $m \in C$ such that for every $x \in C$, $x \leq m$.*

And the "smallness" of a compact space X is reflected in the fact that every infinite subset E has a limit point.

Theorem 7.3. *If X is a compact space, then every infinite subset of X has a limit point.*

Corollary 7.4. *If X is compact and E is a subset of X with no limit point, then E is finite.*

Compactness is framed in terms of open sets and unions; the next two theorems give equivalent formulations in terms of closed sets and intersections. Let us first define the *finite intersection property* of a collection of sets:

Definition. A collection of sets has the **finite intersection property** if and only if every finite subcollection has a non-empty intersection.

This definition can be used in an alternative characterization of compactness:

Theorem 7.5. *A space X is compact if and only if every collection of closed sets with the finite intersection property has a non-empty intersection.*

The next theorem shows that compactness is equivalent to the following property: for every (possibly infinite) collection of closed sets whose intersection lies in an open set, the intersection of some finite number of those closed sets lies in that open set. Notice that one direction is simply a corollary of the previous theorem.

Theorem 7.6. *A space X is compact if and only if for any open set U in X and any collection of closed sets $\{K_\alpha\}_{\alpha\in\lambda}$ such that $\bigcap_{\alpha\in\lambda} K_\alpha \subset U$, there exist a finite number of the K_α's whose intersection lies in U.*

We have been speaking about compactness of a topological space X, but we can just as easily speak of the compactness of a subspace A without confusion, because by the definition of the subspace topology, a cover of A by open sets $\{U_\alpha\}$ in X restricts to a cover of A by relative open sets $\{U_\alpha \cap A\}$ in A. So "every open cover has a finite subcover" has the same meaning whether we regard A as a subspace of X or as its own topological space.

Exercise 7.7. *If A and B are compact subsets of X, then $A \cup B$ is compact. Suggest and prove a generalization.*

As the next theorems show, there is a tight connection between compact sets and closed sets, namely, closed subsets of compact spaces are compact and compact subsets of Hausdorff spaces must be closed.

Theorem 7.8. *Let A be a closed subspace of a compact space. Then A is compact.*

You may find an especially pleasing proof of the next theorem.

Theorem 7.9. *Let A be a compact subspace of a Hausdorff space X. Then A is closed.*

Effective Thinking Principle. *Explore Limits of Theorems; Add Hypotheses.* When you discover a theorem, explore its limitations and possible extensions by systematically weakening the hypotheses and checking to see whether the theorem is still true and strengthening the conclusion to see whether you can deduce more than you originally thought.

The following exercises may be illuminating.

Exercise 7.10. *Construct an example of a compact subset of a topological space that is not closed.*

Exercise 7.11. *Must the intersection of two compact sets be compact? Add hypotheses, if necessary. Extend any theorems you discover, if possible.*

Covering properties and separation properties of spaces are related. Proofs of the following theorems use the interplay of covering and separation properties in a delightful way. For instance the next theorem will show that a compact Hausdorff space is normal. As an intermediate step you may wish to first show that a compact, Hausdorff space is regular.

Theorem 7.12. *Every compact, Hausdorff space is normal.*

We end this introductory section on compactness by making an observation about compactness in spaces whose topology is generated by a basis. In the definition of compactness, arbitrary open covers appear. Suppose the topology is generated by a basis. Would it be sufficient to consider only open covers by basic open sets to determine whether the space is compact?

Theorem 7.13. *Let $\mathcal{B}$ be a basis for a space X. Then X is compact if and only if every cover of X by basic open sets in $\mathcal{B}$ has a finite subcover.*

7.2 The Heine-Borel Theorem

Which subsets of the real line $\mathbb{R}$ are compact? Recall that when we refer to $\mathbb{R}$ and do not mention a topology, then we are talking about $\mathbb{R}_{\text{std}}$.

Let's look at the first non-trivial example: a closed interval. That a closed interval is compact is a basic insight in the topology of the real line. To prove it, you will have to use some axiom about the real numbers such as the axiom that every bounded set has a least upper bound.

Theorem 7.14. *For any $a \leq b$ in $\mathbb{R}$, the subspace $[a, b]$ is compact.*

The next theorem completely characterizes the sets in $\mathbb{R}_{\text{std}}$ that are compact. This theorem, known as the Heine-Borel Theorem, is one of the fundamental theorems about the topology of the line.

Recall a set A in $\mathbb{R}^1$ is **bounded** if and only if there is a number M such that $A \subset [-M, M]$.

Heine-Borel Theorem 7.15. *Let A be a subset of $\mathbb{R}_{std}$. Then A is compact if and only if A is closed and bounded.*

> **Effective Thinking Principle.** *Look at Examples to Understand Theorems.* One of the best ways to make a theorem more meaningful is to see its implications in specific cases.

Exercise 7.16. *Consider the rationals $\mathbb{Q}$ with the subspace topology inherited from $\mathbb{R}$. Find a set A in $\mathbb{Q}$ that is closed and bounded but not compact.*

Knowing that compact subsets of real numbers are closed allows you to give a possibly different proof of the fact that compact subsets of $\mathbb{R}$ have a maximum.

Theorem 7.17. *Every compact subset C of $\mathbb{R}$ contains a maximum in the set C, i.e., there is an $m \in C$ such that for any $x \in C$, $x \leq m$.*

7.3 Compactness and Products

> **Effective Thinking Principle.** *Consider Related Spaces.* Once you have proved a theorem for one space, see whether it is true for related spaces.

Let's explore whether compactness is preserved when you take products of spaces. You will get off to a promising start by proving that the product of two compact spaces is compact. First prove a theorem known as the "tube lemma":

Theorem 7.18 (The tube lemma)**.** *Let $X \times Y$ be a product space with Y compact. If U is an open set of $X \times Y$ containing the set $x_0 \times Y$, then there is some open set W in X containing x_0 such that U contains $W \times Y$ (called a "tube" around $x_0 \times Y$).*

The tube lemma is a good start toward proving that compactness is preserved when taking products.

Theorem 7.19. *Let X and Y be compact spaces. Then $X \times Y$ is compact.*

Repeated application of this theorem shows that any finite product of compact spaces is compact.

Knowing that products of compact spaces are compact suggests that some theorems about the line might be true in higher dimensions. The Heine-Borel Theorem characterizes compact sets in the real line. It is natural to ask what the analogous theorem would be for higher dimensional Euclidean spaces. Viewing $\mathbb{R}^n_{std}$ as a product of copies of $\mathbb{R}_{std}$ is one way to generalize the Heine-Borel Theorem to $\mathbb{R}^n_{std}$.

Heine-Borel Theorem 7.20. *Let A be a subset of $\mathbb{R}^n$ with the standard topology. Then A is compact if and only if A is closed and bounded.*

Theorem 7.13 stated that it was sufficient to consider only basic open covers to determine compactness; this fact follows rather straightforwardly from the definitions. In contrast, the Alexander Subbasis Theorem below is not straightforward to prove. It states that considering subbasic open covers (that is, open covers each element of which is a set in the subbasis) suffices to determine compactness. Recall that a subbasis for a topology is a collection of open sets in the topology with the property that the collection of all finite intersections of sets in the subbasis forms a basis for the topology.

The proof of the Alexander Subbasis Theorem is difficult. You might consider using Zorn's Lemma or the Well-Ordering Principle (see Chapter 2) as you try to prove it. In fact, you will have to use some axiom like that since the Alexander Subbasis Theorem is equivalent to Zorn's Lemma or the Axiom of Choice or the Well-Ordering Principle, which are all equivalent to one another.

Alexander Subbasis Theorem 7.21. *Let $\mathcal{S}$ be a subbasis for a space X. Then X is compact if and only if every subbasic open cover has a finite subcover.*

Exercise 7.22. *Use the Alexander Subbasis Theorem to prove that the space 2^X is compact for every X.*

In fact, you can now show that any product of compact spaces is compact, even infinite products. That assertion is the Tychonoff Theorem, named after the topologist Andrey Nikolayevich Tychonoff or Tikhonov who lived from 1906 until 1993 and proved this theorem in 1930.

Tychonoff's Theorem 7.23. *Any product of compact spaces is compact.*

Exercise 7.24. *Consider the set $[0, 1]^\omega$ and show that the Tychonoff Theorem is not true if the box topology is used instead of the product topology.*

7.4 Countably Compact, Lindelöf Spaces

We now consider some covering properties related to compactness. The strategy for exploring these new covering properties is to first define them and then systematically go through our results from compactness and see how those results manifest themselves in the context of the new covering properties.

Effective Thinking Principle. *Consider Analogies of Previous Results.* After developing variations of previous concepts, look at previous results and see what analogous results hold.

Definition. A space X is **countably compact** if and only if every countable open cover of X has a finite subcover.

Definition. A space X is **Lindelöf** if and only if every open cover of X has a countable subcover.

It is evident that a compact space is countably compact; and a compact space is also Lindelöf. In the reverse direction, we have the following result.

Theorem 7.25. *Every countably compact and Lindelöf space is compact.*

In some sense, the preceding theorem shows that compactness can be broken into two steps—taking an arbitrary cover and making it countable and then taking a countable cover and making it finite. This result encourages us to look again at our previous results where compactness was in the hypothesis with an eye toward investigating whether both halves of this two step process were needed in drawing the conclusion.

Effective Thinking Principle. *Weaken Hypotheses if Possible.* To understand theorems better and to improve them if possible, identify exactly what aspects of the hypotheses were actually used in the proof.

One of the early theorems you proved about compactness concerned limit points. You proved earlier that in a compact space, every infinite set has a limit point. Actually, only the countable compactness property was needed to draw that conclusion. In fact, the issue of convergence basically characterizes countably compact spaces.

Theorem 7.26. *Let X be a T_1 space. Then X is countably compact if and only if every infinite subset of X has a limit point.*

In a Lindelöf space, it is possible to have an infinite set with no limit point; however, recall that you proved earlier that every uncountable set in a 2^{nd} countable space must have a limit point. The following theorem shows that the same must be true in a Lindelöf space.

Theorem 7.27. *If X is a Lindelöf space, then every uncountable subset of X has a limit point.*

Exercise 7.28. *Formulate and prove theorems about Lindelöf and countably compact spaces analogous to the theorems you proved relating compactness with collections of closed sets with the finite intersection property.*

Closed subsets of compact spaces are compact. The similar statement is true for Lindelöf and countably compact spaces.

Theorem 7.29. *If A is a closed subspace of a countably compact (respectively, Lindelöf) space, then A is countably compact (respectively, Lindelöf).*

Earlier you proved a connection between compactness and normality. Recall that in proving that a compact, Hausdorff space is normal, you first proved that a compact, Hausdorff space is regular. Then you proved that a regular, compact Hausdorff space is normal. The following theorem observes that after regularity is established, the Lindelöf condition of the space is all that is needed to infer the conclusion. You may find it useful to use the Normality Lemma in your proof.

Theorem 7.30. *Every regular, Lindelöf space is normal.*

In 7.13 you proved that only basic open covers need be considered when determining compactness of spaces. That encourages us to consider the analogous questions for our new covering properties.

Theorem 7.31. *Let $\mathcal{B}$ be a basis for a space X. Then X is Lindelöf if and only if every cover of X by basic open sets in $\mathcal{B}$ has a countable subcover.*

Corollary 7.32. *Every 2^{nd} countable space is Lindelöf.*

So in a 2^{nd} countable space (which is Lindelöf), regularity implies normality. We have also seen that in 2^{nd} countable spaces, uncountable subsets have limit points, and limit points are reachable by convergent sequences. So we see that 2^{nd} countable spaces have special properties, which we will explore further when we study *metric spaces* in Chapter 10.

You have seen that a natural connection exists between bases and the Lindelöf property. The following exercise is the next logical question, because it asks whether considering basic open sets is sufficient to determine countable compactness. Although counterexamples exist, the authors cannot think of one, but perhaps you can.

Exercise 7.33. *Can you think of a topological space in which every countable open cover by basic open sets has a finite subcover and yet not every countable open cover has a finite subcover?*

You saw that the product of compact spaces is compact. So it is natural to investigate whether other covering properties are preserved when we take products. The answers are no.

Exercise 7.34. *Show that $\mathbb{R}_{LL}$ is Lindelöf, but $\mathbb{R}_{LL} \times \mathbb{R}_{LL}$ is not Lindelöf.*

There are countably compact spaces whose product is not countably compact; however, they are not easy to produce.

> **Effective Thinking Principle.** *Look at Examples.* After learning new concepts, investigate examples to see how those properties manifest themselves.

The ordinal numbers (Section 2.6) give us some good examples of compact and countably compact spaces.

Theorem 7.35. *The space ω_1 of countable ordinals is countably compact but not compact.*

Theorem 7.36. *The space $\omega_1 + 1$, which includes all countable ordinals together with the ordinal ω_1, is compact.*

Exercise 7.37. *Extend your table of spaces and properties by adding new rows for the properties compact, Lindelöf, and countably compact and determining those properties for each of your spaces.*

7.5 Paracompactness

One strategy for developing new mathematics is to take a concept that applies to a whole space and exploring local versions of it. You saw this strategy at work when you first investigated spaces with a countable basis and then explored the local version, namely, 1^{st} countable spaces. Here we will take the idea of compactness, that is, every open cover having a finite subcover, and asking what kind of a local version of this property might be possible. This definition may appear technical-sounding at first, but it is a natural result of creating a local version of compactness. The concept of paracompactness, which we define below, is central to the idea of metrizability, which we consider in a future chapter; however, this section may be skipped for now without loss of continuity.

Definition. A collection $\mathcal{B} = \{B_\alpha\}_{\alpha\in\lambda}$ of subsets of a space X is **locally finite** if and only if for each point p in X there is an open set U containing p such that U intersects only finitely many elements of $\mathcal{B}$.

Example. Let $\mathcal{B} = \{[n, n+1] \subset \mathbb{R} \mid n \text{ is an integer}\}$. Then $\mathcal{B}$ is a locally finite collection in $\mathbb{R}_{\text{std}}$.

Generally, the closure of an infinite union of subsets of a topological space may be larger than the union of the closures of the individual sets (recall Exercise 3.23). However, if the sets are locally finite then the union of the closures is the closure of the union even for infinite collections.

Theorem 7.38. *Let $\mathcal{B} = \{B_\alpha\}_{\alpha\in\lambda}$ be a locally finite collection of subsets of a space X. Then*

$$\overline{(\bigcup_{\alpha\in\lambda} B_\alpha)} = \bigcup_{\alpha\in\lambda} \overline{B}_\alpha.$$

The next definition may seem difficult to follow at first; however, the idea is that we have a cover of a space and we simply break each of the elements of that cover into smaller pieces to get a cover by sets each of which is a subset of one of our original elements of the cover.

Definition. Let $\mathcal{B} = \{B_\alpha\}_{\alpha\in\lambda}$ be a cover of X. Then $\mathcal{C} = \{C_\beta\}_{\beta\in\mu}$ is a **refinement** of $\mathcal{B}$ if and only if (i) $\mathcal{C}$ is a cover of X and (ii) for each $\beta \in \mu$ there is an $\alpha \in \lambda$ such that $C_\beta \subset B_\alpha$. The collection $\mathcal{C}$ is an **open refinement** if and only if each C_β is an open set.

The ability to create locally finite refinements of open covers turns out to allow us to recognize that spaces have useful properties that are not obviously apparent. So we give a name to spaces with such refinements of covers. It is traditional to include the Hausdorff property as part of the the definition of paracompact spaces.

Definition. A space X is **paracompact** if and only if every open cover of X has a locally finite open refinement and X is Hausdorff.

Clearly, every compact, Hausdorff space is paracompact. Paracompact spaces enjoy some properties that compact spaces do.

Theorem 7.39. *Let A be a closed subspace of a paracompact space. Then A is paracompact.*

Previously we saw that compact, Hausdorff spaces are normal. Here we see that paracompact spaces are normal. (Remember that part of the definition of paracompact is that the space is Hausdorff.)

Theorem 7.40. *Every paracompact space is normal.*

This next theorem encourages you to find a way to change countable covers into locally finite covers. As you work on proving this theorem, you will inevitably come to grips with the fact that the definition of paracompactness automatically involves three open covers of the space—the open cover you are originally given, the open cover that is the refinement that you create, and the open cover implicit in the definition of being locally finite, that is, for each point in the space there is an open set containing it that intersects only finitely many of the open sets in the refinement, so that collection of open sets used to verify local finiteness is a third open cover. So as you construct a locally finite open refinement of your original open cover, you might want to think overtly about constructing the open cover of sets that are used in verifying local finiteness. The following theorem is challenging, but not impossible.

Theorem 7.41. *Every regular, T_1, Lindelöf space is paracompact.*

We will return to a further discussion of paracompactness in Chapter 10 on metric spaces.

7.6 Covering Up Reveals Strategies for Producing Mathematics

This chapter introduced a whole array of concepts that start with open covers of topological spaces. The first and most basic concept was the idea of a space being compact—that is, every open cover has a finite subcover. One reason for making that definition was the realization that frequently arguments were available if we knew our cover was finite. In some instances, having a finite cover was almost as useful as having a finite space. The arguments in those two cases were similar.

One great feature of this exploration of compactness is that it illustrated strategies of developing mathematics very well. After the concept of compactness was isolated, the strategy was to see how that concept interacted with previous concepts about spaces. You connected compactness with limit points, with separation properties, and with products. You extended the definition of compactness to related covering properties—Lindelöf and countably compact. You looked for a local version of compactness and came up with paracompactness. The systematic exploration of covering properties was a great model for how to create and develop mathematical ideas.

Chapter 8

Continuity: When Nearby Points Stay Together

After defining a mathematical object or structure, a next natural challenge is to describe maps between such objects that respect or preserve the structures we have defined. For instance, in group theory we study homomorphisms, because a homomorphism is a function from one group to another group that respects the binary operations that are the heart of the concept of a group. When we consider functions between vector spaces in linear algebra, we study linear transformations, because linear transformations respect the linear structure that is the core of a vector space.

So now we want to describe topologically appropriate functions between topological spaces. Suppose f is a function from one topological space X to another topological space Y (denoted $f : X \to Y$). What properties should f have in order to respect the topologies on X and Y? Perhaps our first attempt would be to insist that f take each open set in X to an open set in Y. However, given our motivation for the whole subject of topology to generalize ideas we encountered from analysis, we might look at the most basic kind of continuous functions from calculus as a guide. The function $f(x) = x^2$ does not take open intervals around 0 to open sets in $\mathbb{R}$, so the restriction of taking open sets to open sets does not seem to be the right definition for continuity of functions between topological spaces. However, our experience with continuous functions on the real numbers gives us some indication of what features we want continuous functions between topological spaces to have. The definition of a continuous function from calculus starts with an open interval in the range (remember $\varepsilon > 0$?). Then it finds an open interval in the domain (remember $\delta > 0$?). That definition gives us the motivation for our definition of continuous functions between topological spaces.

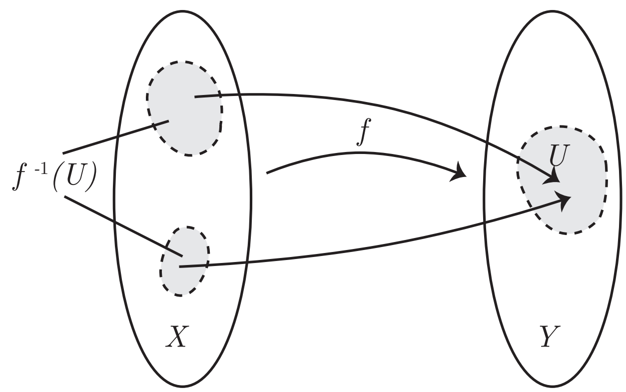


Figure 8.1: For a continuous function, inverse images of open sets are open.

8.1 Continuous Functions

Definition. Let X and Y be topological spaces. A function or map $f : X \to Y$ is a **continuous function** or **continuous map** if and only if for every open set U in Y, $f^{-1}(U)$ is open in X.

In other words, a continuous function is one in which inverse images of open sets are open. This definition does not look like the usual definition of continuous function you might have encountered in analysis. But perhaps the last of the following equivalent characterizations of continuity looks more like the definition of continuity from calculus.

Theorem 8.1. *Let X and Y be topological spaces and let $f : X \to Y$ be a function. Then the following are equivalent:*

1. *The function f is continuous.*
2. *For every closed set K in Y, the inverse image $f^{-1}(K)$ is closed in X.*
3. *For every limit point p of a set A in X, the image $f(p)$ belongs to $\overline{f(A)}$.*
4. *For every $x \in X$ and open set V containing $f(x)$, there is an open set U containing x such that $f(U) \subset V$.*

Equivalence (2) above says a continuous function is one in which the inverse images of closed sets are closed. Equivalence (3) essentially asserts that continuous functions preserve limit points. That summary is not quite accurate, because if a point p is a limit point of a set A and the whole domain space goes to a single point, for example, then that function would be continuous while $f(p)$ is not actually a limit point of $f(A)$ since $f(A)$ is an isolated point. However, $f(p)$ is in the closure of $f(A)$. With this rather trivial caveat excepted, the preservation of limit points characterizes continuous functions.

Effective Thinking Principle. *Bring to Mind All Equivalent Formulations.* After learning that a concept has equivalent formulations, get in the habit of consciously considering each one in order to select the most convenient option in a given situation.

In doing the next exercises, consciously consider all the various equivalences of continuity to find convenient formulations and/or to find alternative proofs.

Theorem 8.2. *Let X, Y be topological spaces and $y_0 \in Y$. The constant map $f : X \to Y$ defined by $f(x) = y_0$ is continuous.*

Theorem 8.3. *Let $X \subset Y$ be topological spaces. The inclusion map $i : X \to Y$ defined by $i(x) = x$ is continuous.*

Theorem 8.4. *Let $f : X \to Y$ be a continuous map between topological spaces, and let A be a subset of X. Then the restriction map $f|_A : A \to Y$ defined by $f|_A(a) = f(a)$ is continuous.*

Effective Thinking Principle. *Ask About Local Versions of Global Concepts.* After defining a concept for a whole space, investigate whether there is a local version of that concept.

Both from using good strategies of thinking and from knowing the definition of continuity from calculus, it is natural to seek a point by point version of continuity. As usual, looking at the various equivalences of continuity will help in defining what it means for a function between topological spaces to be continuous at a point.

Definition. Let $f : X \to Y$ be a function between topological spaces X and Y, and let $x \in X$. Then f is **continuous at the point** x if and only if for every open set V containing $f(x)$, there is an open set U containing x such that $f(U) \subset V$. Thus a function $f : X \to Y$ is continuous if and only if it is continuous at each point.

Equivalence (4) is also the most closely related to the calculus ε-δ definition of continuity for functions from $\mathbb{R}_{std}$ to $\mathbb{R}_{std}$. The topological space $\mathbb{R}_{std}$ has motivated many of the definitions of topology, so if the familiar concept of continuity in $\mathbb{R}_{std}$ did not correspond to our definition of continuity between topological spaces, then we would have to question the wisdom of our choice of definitions.

Effective Thinking Principle. *Generalizations Should Generalize.* After generalizing a concept, confirm that the generalization behaves properly when applied to the concept from which it sprang.

Theorem 8.5. *A function $f : \mathbb{R}_{std} \to \mathbb{R}_{std}$ is continuous if and only if for every point x in $\mathbb{R}$ and $\varepsilon > 0$, there is a $\delta > 0$ such that for every $y \in \mathbb{R}$ with $d(x, y) < \delta$, then $d(f(x), f(y)) < \varepsilon$.*

When 1^{st} countable spaces are involved, continuity can be described in terms of convergence; in particular, continuous functions preserve limits of sequences.

Theorem 8.6. *Let X be a 1^{st} countable space and Y be a topological space. Then a function $f : X \to Y$ is continuous if and only if for each convergent sequence $x_n \to x$ in X, $f(x_n)$ converges to $f(x)$ in Y.*

In many cases a continuous function is determined by its values on a dense set.

Theorem 8.7. *Let X be a space with a dense set D, and let Y be Hausdorff. Let $f : X \to Y$ and $g : X \to Y$ be continuous functions such that for every d in D, $f(d) = g(d)$. Then for all x in X, $f(x) = g(x)$.*

Hearkening back to our discussion of cardinality, you might now enjoy proving the following fact about the cardinality of continuous functions. This result should be contrasted with Theorem 2.29.

Theorem 8.8. *The cardinality of the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is the same as the cardinality of $\mathbb{R}$.*

We can create new continuous functions from old ones in many ways. For instance, compositions of continuous functions are continuous.

Theorem 8.9. *If $f : X \to Y$ and $g : Y \to Z$ are continuous, then their composition $g \circ f : X \to Z$ is continuous.*

We can paste two continuous functions together if the pieces are either both closed or both open.

Theorem 8.10 (pasting lemma)**.** *Let $X = A \cup B$, where A, B are closed in X. Let $f : A \to Y$ and $g : B \to Y$ be continuous functions that agree on $A \cap B$. Then the function $h : A \cup B \to Y$ such that $h = f$ on A and $h = g$ on B is continuous.*

Theorem 8.11 (pasting lemma)**.** *Let $X = A \cup B$, where A, B are open in X. Let $f : A \to Y$ and $g : B \to Y$ be continuous functions which agree on $A \cap B$. Then the function $h : A \cup B \to Y$ such that $h = f$ on A and $h = g$ on B is continuous.*

Exercise 8.12. *Is the pasting lemma true when A and B in the preceding theorems are arbitrary sets?*

If $f : X \to Y$ is a function and Y has a topology given by a basis or a subbasis, then it is sufficient to check the pre-images of basic or subbasic open sets to determine whether f is continuous.

Theorem 8.13. *Let $f : X \to Y$ be a function and let $\mathcal{B}$ be a basis for Y. Then f is continuous if and only if for every open set B in $\mathcal{B}$, $f^{-1}(B)$ is open in X.*

Theorem 8.14. *Let $f : X \to Y$ be a function and let $\mathcal{B}$ be a subbasis for Y. Then f is continuous if and only if for every open set B in $\mathcal{B}$, $f^{-1}(B)$ is open in X.*

8.2 Properties Preserved by Continuous Functions

> **Effective Thinking Principle.** *Investigate How Properties are Related by Transformations.* After defining structure-respecting transformations, investigate what properties are preserved by those functions.

Continuous functions preserve some of the properties of topological spaces that we have considered, but not others. A great way to develop mathematical ideas is to systematically consider which properties are preserved by continuous functions and which are not. The covering properties give us a promising start.

Theorem 8.15. *If X is compact, and $f : X \to Y$ is continuous and surjective, then Y is compact.*

Theorem 8.16. *If X is Lindelöf and $f : X \to Y$ is continuous and surjective, then Y is Lindelöf.*

Theorem 8.17. *If X is countably compact and $f : X \to Y$ is continuous and surjective, then Y is countably compact.*

Surjective continuous functions take dense sets to dense sets.

Theorem 8.18. *Let D be a dense set of a topological space X and let $f : X \to Y$ be continuous and surjective. Then $f(D)$ is dense in Y.*

Corollary 8.19. *Let X be a separable space and let $f : X \to Y$ be continuous and surjective. Then Y is separable.*

When we turn our attention to the separation properties, we meet with less success. In fact, the continuous images of Hausdorff, regular, or normal spaces need not preserve those separation properties. Since using quotient spaces (defined later in this chapter) may be the simplest way to generate those counterexamples, we suggest you defer the construction of those counterexamples until later in this chapter. Instead, let's use a strategy of developing mathematics that is often fruitful.

Effective Thinking Principle. *Add a Hypothesis.* When you can't prove a conjecture, create additional hypotheses that would allow you to prove it.

You will construct examples later that show you that the continuous image of a normal space may not be normal; however, if we follow our original intuition about how to define continuous functions, we find types of functions that do preserve normality. Recall that our first (and flawed) attempt at defining a function that respects the essential features of a topology was to consider functions that take open sets to open sets or perhaps closed sets to closed sets. So now let's consider functions that are continuous using the definition we settled on, but also have an additional property of taking open sets to open sets or closed sets to closed sets.

Definition. A function $f : X \to Y$ is **closed** if and only if for every closed set A in X, $f(A)$ is closed in Y. A function $f : X \to Y$ is **open** if and only if for every open set U in X, $f(U)$ is open in Y.

Every combination of continuity, openness, and closedness is possible.

Exercise 8.20. *1. Find an open function that is not continuous.*

2. Find a closed function that is not continuous.

3. Find a continuous function that is neither open nor closed.

4. Find a continuous function that is open but not closed.

5. Find a continuous function that is closed but not open.

Effective Thinking Principle. *Draw a Picture.* Draw a picture. Draw a picture!! DRAW A PICTURE!!! It is impossible to overemphasize the value of drawing a picture.

Closed continuous functions preserve normality. In proving this theorem and many of the theorems in this chapter and this book, drawing a picture can be extremely helpful. A picture is not a proof, but a picture can suggest relationships that you can then confirm in a proof.

Theorem 8.21. *If X is normal and $f : X \to Y$ is continuous, surjective, and closed, then Y is normal.*

Bases go to bases under continuous open maps.

Theorem 8.22. *If $\{B_\alpha\}_{\alpha\in\lambda}$ is a basis for X and $f : X \to Y$ is continuous, surjective, and open, then $\{f(B_\alpha)\}_{\alpha\in\lambda}$ is a basis for Y.*

Corollary 8.23. *If X is 2^{nd} countable and $f : X \to Y$ is continuous, surjective, and open, then Y is 2^{nd} countable.*

Some continuous functions are automatically closed.

Theorem 8.24. *Let X be compact and Y be Hausdorff. Then any continuous function $f : X \to Y$ is closed.*

The next theorem has as its conclusion that the image space of a function is 2^{nd} countable. Earlier you saw that the image of a 2^{nd} countable space under an open map is 2^{nd} countable. However, the map in the following theorem need not be open. So you will require more ingenuity to prove it.

Theorem 8.25. *Let X be compact and 2^{nd} countable and let Y be Hausdorff. If $f : X \to Y$ is continuous and surjective, then Y is 2^{nd} countable.*

8.3 Homeomorphisms

One of the most important reasons for introducing the concept of maps between objects and for requiring that the maps respect the underlying structure is to specify when two such objects are the "same". In group theory, we try to classify groups by determining which groups are actually the "same", i.e., related by a bijective homomorphism, called an isomorphism. In topology, the task is analogous.

Definition. A function $f : X \to Y$ is a **homeomorphism** if and only if f is a continuous bijection and the inverse map $f^{-1} : Y \to X$ is also continuous.

Definition. Two topological spaces, X and Y, are **homeomorphic** or **topologically equivalent** if and only if there exists a homeomorphism $f : X \to Y$.

Notice that a homeomorphism f provides both a bijection between sets *and* a bijection between the topologies on those sets, since open sets are preserved by f and f^{-1}. So all topological properties—properties that depend only on the topology (e.g., Hausdorff, regular, normal, compact, separable, and so on)—are preserved by a homeomorphism.

Theorem 8.26. *Being homeomorphic is an equivalence relation on topological spaces.*

When asked to show that two spaces are homeomorphic, the first thing to do is to find a desired bijection and then show it is continuous in both directions.

Exercise 8.27. *Let a and b be points in $\mathbb{R}^1$ with $a < b$. Show that (a, b) with the subspace topology from $\mathbb{R}^1_{std}$ is homeomorphic to $\mathbb{R}^1_{std}$.*

Checking continuity in both directions can be cumbersome, so it is helpful to find conditions on functions equivalent to being a homeomorphism that don't require checking the continuity of the inverse function.

Theorem 8.28. *If $f : X \to Y$ is continuous, the following are equivalent:*

a) f is a homeomorphism.

b) f is a closed bijection.

c) f is an open bijection.

Sometimes we can guarantee that a bijective function is a homeomorphism with apparently weaker conditions.

Theorem 8.29. *Suppose $f : X \to Y$ is a continuous bijection where X is compact and Y is Hausdorff. Then f is a homeomorphism.*

Effective Thinking Principle. *Check the Necessity of Hypotheses.* After you prove a theorem, check that all the hypotheses are necessary.

Exercise 8.30. *Construct some examples to show why the compactness and Hausdorff assumptions in the previous theorem are necessary.*

A homeomorphism between a space and a subset of another space is called an embedding.

Definition. A function $f : X \to Y$ is an **embedding** if and only if $f : X \to f(X)$ is a homeomorphism from X to $f(X)$, where $f(X)$ has the subspace topology from Y.

The previous theorem has a direct corollary about embeddings.

Corollary 8.31. *Let X be a compact space and let Y be Hausdorff. If $f : X \to Y$ is a continuous, injective map, then f is an embedding.*

8.4 Product Spaces and Continuity

In this section we explore how the concept of continuity interacts with the concept of product spaces.Product spaces have natural projection functions, and the continuity of those projection maps, as we shall see, characterizes the topology of the product space.

Definition. The **projection maps** $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are defined by $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$.

Theorem 8.32. *Let X and Y be topological spaces. The projection maps π_X, π_Y on $X \times Y$ are continuous, surjective, and open.*

In fact, the topology on the product space can be characterized as the coarsest topology that makes the projection maps continuous.

Theorem 8.33. *Let X and Y be topological spaces. The product topology on $X \times Y$ is the coarsest topology on $X \times Y$ that makes the projection maps π_X, π_Y on $X \times Y$ continuous.*

Exercise 8.34. *Find an example of X and Y that shows that the projection map $\pi_X : X \times Y \to X$ is not necessarily a closed map.*

Theorem 8.35. *Let X and Y be topological spaces. For every $y \in Y$, the subspace $X \times \{y\}$ of $X \times Y$ is homeomorphic to X.*

We can describe a criterion for when maps *into* a product space are continuous. This criterion is sometimes called the *universal mapping property of products*:

Theorem 8.36. *Let X, Y, and Z be topological spaces. A function $g : Z \to X \times Y$ is continuous if and only if $\pi_X \circ g$ and $\pi_Y \circ g$ are both continuous.*

Exercise 8.37. *What about maps* out of *a product space, i.e., $f : X \times Y \to Z$? Do you think f is continuous if it is continuous in each coordinate?*

The theorems above were stated for products of two spaces. Of course, all of those theorems can be extended to finite products. It turns out that those same theorems can be extended to infinite products as well. For the product of an arbitrary collection of spaces $\{X_\alpha\}_{\alpha\in\lambda}$, the projection functions are defined analogously to how they were defined for finite products.

Theorem 8.38. *Let $\prod_{\alpha\in\lambda} X_\alpha$ be the product of topological spaces $\{X_\alpha\}_{\alpha\in\lambda}$. The projection map $\pi_\beta : \prod_{\alpha\in\lambda} X_\alpha \to X_\beta$ is a continuous, surjective, and open map.*

The projection maps can be used to characterize the infinite product topology just as they were used to characterize finite products:

Theorem 8.39. *The product topology is the coarsest (smallest) topology on $\prod_{\alpha\in\lambda} X_\alpha$ that makes each projection map continuous.*

Similarly, the continuity of maps into infinite products can be determined by projecting onto each coordinate.

Theorem 8.40. *Let $\prod_{\alpha\in\lambda} X_\alpha$ be the product of topological spaces $\{X_\alpha\}_{\alpha\in\lambda}$ and let Z be a topological space. A function $g : Z \to \prod_{\alpha\in\lambda} X_\alpha$ is continuous if and only if $\pi_\beta \circ g$ is continuous for each β in λ.*

Exercise 8.41. *Let $\mathbb{R}^\omega$ be the countably infinite product of $\mathbb{R}$ with itself. Let $f : \mathbb{R} \to \mathbb{R}^\omega$ be defined by $f(x) := (x, x, x, ...)$. Then f is continuous if $\mathbb{R}^\omega$ is given the product topology, but not if given the box topology. (This strange result once again shows why the box topology would be a poor choice as the standard topology for infinite products.)*

Definition. Consider the following subsets of $\mathbb{R}$. Let $C_0 = [0, 1]$. Let $C_1 = [0, 1/3] \cup [2/3, 1]$, i.e., it is C_0 with its "middle third" removed. Let $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, which is obtained from C_1 by removing the middle thirds of each of its interval. Continue in this fashion, defining C_n to be C_{n-1} with the middle-thirds of each interval of C_n removed. Let

$$C = \cap_{i=1}^{\infty} C_i.$$

This space is called the **Cantor set** or the **standard Cantor set**.

Note that the Cantor set is an answer to Theorem 3.25.

Theorem 8.42. *The Cantor set is homeomorphic to the product $\prod_{n\in\mathbb{N}}\{0, 1\}$ where $\{0, 1\}$ has the discrete topology.*

8.5 Quotient Maps and Quotient Spaces

Now we'll learn another way to construct new topological spaces from old ones, in a way which corresponds with our intuition that new spaces can be formed by gluing old spaces together.

A sheet of paper X has a natural topology, namely, the subspace topology given by its embedding as a subset of the plane. This corresponds with our intuition—for any given point p on the sheet, the open sets around p describe a set of points "near" p on the sheet. Points in the interior of X have basic open sets that are disks. Points in the boundary of X have basic open sets that are half-disks.

Now imagine gluing a pair of parallel sides of X together to form a cylinder C. To be more formal, imagine $X = [0, 2] \times [0, 1]$ and for each $x \in [0, 2]$, we are **identifying** the pair of points $(x, 0)$ and $(x, 1)$, meaning we are considering these two points to be "identical" when glued together as a point on C. This gluing is sometimes notated by drawing arrows on the parallel sides of X that point in the same direction.

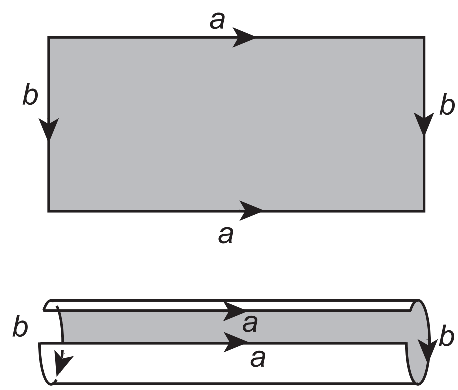


Figure 8.2: Sheet of paper with two sides identified forms a cylinder.

Note that after this gluing something has just happened that has changed the character of this space. For instance, the sequence of points $(1/2, 1/n)$ now converge to the point $p' = (1/2, 1)$ because it is now identified to $p = (1/2, 0)$, whereas this sequence did not converge to $(1/2, 1)$ before. Thus this new space C has a very natural topology that it derives from the topology of the sheet X, and we just need to figure out how to make this notion precise.

The first thing we can do is carefully describe the space C and then we can think of the gluing as a function $g : X \to C$. Let's represent the cylinder C as the set of points $(x, \sin\theta, \cos\theta)$ for $x \in [0, 2]$ and $\theta \in [0, 2\pi)$. Then the gluing can be described as a function $g : X \to C$ given by

$$g(x, y) = (x, \sin 2\pi y, \cos 2\pi y)$$

and it is apparent that $(x, 0)$ and $(x, 1)$ both get mapped to the same point $(x, 0, 1)$.

The second thing we observe is that the gluing function from our intuition is naturally continuous, since we have not torn the piece of paper in any way. So no matter how we define the topology on C, the gluing function g had better be continuous as well. The natural way to define the topology on C is to *define* U to be open in C if and only if the inverse image $g^{-1}(U)$ is open in X. Then the function g is automatically continuous. What consequence does this definition have on the topology of C? Well, we see that any open set around the point $g(p) = (0, 1, 1/2)$, for example, comes from open sets around $p = (0, 1/2)$ and $p' = (1, 1/2)$—in particular, half-disks around p and p' in X correspond to a full disk around $g(p)$ in C. The idea of a *quotient space* generalizes this idea beautifully.

Definition. Let X be a topological space, and let X^* be a partition of X into disjoint subsets whose union is X. Let $f : X \to X^*$ be the surjective map that carries each point of X to the element of X^* containing it, and define a topology on X^* in which a set U is open if and only if $f^{-1}(U)$

is open in X. The map f is called an **identification map**, because one can think of obtaining X^* by *identifying* all the elements in a partition class to a single point, and the space X^* is called an **identification space** of X under the map f.

Any equivalence relation $\sim$ on points of a space X yields a partition of X into equivalence classes, and the resulting identification space is denoted $X/\sim$.

Example. If X and Y are topological spaces that may or may not have points in common, we can topologize their union $X \cup Y$ as the identification space of the disjoint union $X \sqcup Y$. Note that $X \sqcup Y$ has an obvious topology: the open sets are generated as unions of open sets in X and in Y. The identification map $f : X \sqcup Y \to X \cup Y$ takes each point in $X \sqcup Y$ to the obvious point in $X \cup Y$, and produces a natural topology on $X \cup Y$: the open sets are ones whose intersection with X is open in X and whose intersection with Y is open in Y.

Exercise 8.43. *The cylinder C from our example above did not need to be embedded in $\mathbb{R}^3$ to be defined; it could have been defined as an identification space of $X = [0,1] \times [0,1]$, using the partition whose sets are either singletons or pairs:*

$$C^* = \Big\{\{(x,y)\} : x \in (0,1), y \in [0,1]\Big\} \cup \Big\{\{(0,y) \cup (1,y)\} : y \in [0,1]\Big\}.$$

What is the identification map $f : X \to C^*$? What is a basis for the topology on C^*?

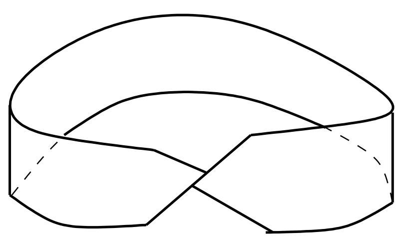


Figure 8.3: A Möbius band.

Exercise 8.44. *A **Möbius band** is obtained by taking a strip of paper X and gluing two opposite sides with a "twist". Sometimes this gluing is notated by drawing X with arrows on two parallel sides that point in opposite directions. Construct a Möbius band explicitly as an identification space of $X = [0,8] \times [0,1]$.*

Exercise 8.45. *A **torus** is the surface of a doughnut. Construct a torus explicitly as:*

1. *an identification space of C, the cylinder.*

2. *an identification space of $X = [0,1] \times [0,1]$.*

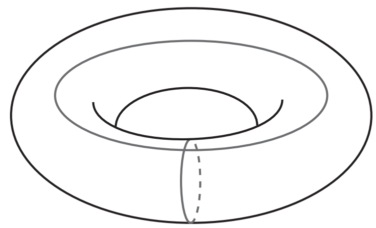


Figure 8.4: A torus.

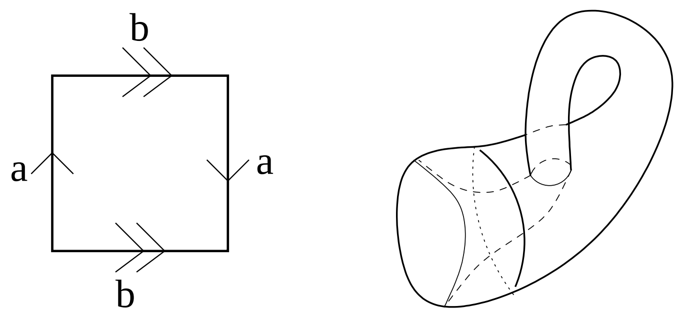


Figure 8.5: Gluing diagram for a Klein bottle.

3. *an identification space of* $\mathbb{R}^2$.

Example. Figure 8.5 shows how to construct a **Klein bottle** as an identification space of a square $X = [0,1] \times [0,1]$. If you try to physically glue the edges as shown, you will find the result can't be constructed in $\mathbb{R}^3$ without self-intersections. Notice how an advantage of the identification space construction is that it defines the topology on the Klein bottle without reference to an ambient space in which the Klein bottle must sit.

Exercise 8.46. *Describe the* 2*-dimensional sphere (the boundary of a* 3*-dimensional ball in* $\mathbb{R}^3$*) as an identification space of two discs in* $\mathbb{R}^2$ *by drawing a figure.*

We have seen how an identification space can be constructed from a partition of a topological space. This construction can be generalized.

Definition. Let $f : X \to Y$ be a surjective map from a topological space X onto a set Y. The **quotient topology** on Y with respect to f is the collection of all sets U such that $f^{-1}(U)$ is open in X.

Note that in the definition above, Y is simply a set and we are giving it a topology.

Theorem 8.47. *The quotient topology actually defines a topology.*

The next exercise characterizes the quotient topology in terms of the map f. Compare it to Theorem 8.39 about the product topology.

Theorem 8.48. *Let X be a topological space, Y be a set, and $f : X \to Y$ be a surjective map. The quotient topology on Y is the finest (largest) topology that makes f continuous.*

Definition. A surjective map $f : X \to Y$ between topological spaces is a **quotient map** and Y is a **quotient space** if and only if the topology on Y is the quotient topology with respect to f.

Note by definition that an identification space is a quotient space.

Theorem 8.49. *Let X and Y be topological spaces. A surjective, continuous map $f : X \to Y$ that is an open map is a quotient map.*

Theorem 8.50. *Let X and Y be topological spaces. A surjective, continuous map $f : X \to Y$ that is a closed map is a quotient map.*

Exercise 8.51. *Show with examples that not all quotient maps are open maps, and not all quotient maps are closed maps.*

Exercise 8.52. *Is $\pi : \mathbb{R}^2 \to \mathbb{R}$ defined by $\pi(x, y) = x$ a quotient map?*

The next theorem is called the *universal mapping property of quotient spaces*: it gives a criterion for whether maps out of a quotient space are continuous. Compare it to Theorem 8.36.

Theorem 8.53. *Let $f : X \to Y$ be a quotient map. Then a map $g : Y \to Z$ is continuous if and only if $g \circ f$ is continuous.*

Theorem 8.53 and judicious use of several previous theorems (e.g. Theorems 5.19, 8.15, 8.29, and 8.40) can simplify the proof of the next exercise, since the cylinder C^* as an identification space is also a quotient space.

Exercise 8.54. *Let the cylinders C^* and C be defined as at the beginning of this Section. Prove that C^* is homeomorphic to C by constructing a map $h : C^* \to C$ and showing it is a continuous bijection from a compact space into a Hausdorff space.*

We end this section with ways of building new topological spaces from old ones, using the quotient topology.

Definition. Let A and B be two disjoint spaces, with points $p \in A$ and $q \in B$. Then the **wedge product of A and B** (relative to p and q), denoted $A \vee B$, is defined as the quotient space $A \cup B / p \sim q$. In other words, we glue p to q.

Notice that the space resulting from the wedge process is in general dependent on the points we choose (even up to homeomorphism).

An important example of the wedge product is the **bouquet of n circles**. This space results when n circles are glued together at one point (so it is $n-1$ repeated applications of the wedge product). See Figure 8.6 (the space shown is, of course, homeomorphic to a bouquet of circles; if want to avoid distorting the circles, we must embed the space into $\mathbb{R}^3$). For obvious reasons, the bouquet of 2 circles is often called the figure eight.

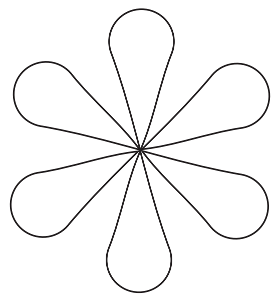


Figure 8.6: The bouquet of n circles.

Definition. Given a topological space X, consider the quotient space $X \times [0,1]$ such that all points $(x,0)$ are identified to a single point p.

Example. The cone over $\mathbb{S}^1$ is shown in Figure 8.7. Of course the word "cone" is commonly used to mean this object.

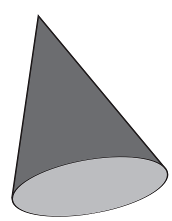


Figure 8.7: A cone.

Exercise 8.55. *Suppose X is a subspace of $\mathbb{R}^n$ for some n. View $\mathbb{R}^n$ as a subset of $\mathbb{R}^{n+1}$ in the usual way (that is, $\mathbb{R}^n$ is the space of the first n coordinates of $\mathbb{R}^{n+1}$ where the final coordinate is 0). Choose a point $x_0 \in \mathbb{R}^{n+1} - \mathbb{R}^n$. Let C be the subspace of R^{n+1} consisting of the union of all the line segments from x_0 to points in X. Show that C is homeomorphic to the cone over X as defined above, thus justifying the name "cone."*

We will see quotient spaces later in the chapter about the classification of 2-manifolds.

8.6 Urysohn's Lemma and the Tietze Extension Theorem

In this section, we will investigate an important relationship between normality of a space X and the existence of some continuous functions from X into $[0, 1]$ with the standard topology. That relationship is captured in theorems known as Urysohn's Lemma and the Tietze Extension Theorem.

The next lemma is used in the proof of Urysohn's Lemma. It may be useful to remember that the rationals are countable, so they can be dealt with sequentially.

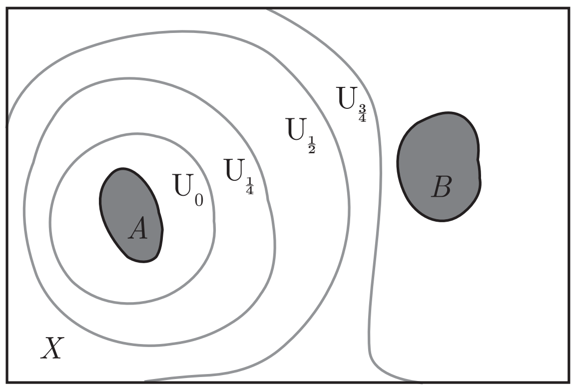


Figure 8.8: These sets may help you prove Urysohn's Lemma.

Lemma 8.56. *Let A and B be disjoint closed sets in a normal space X. Then for each rational $r \in [0, 1]$, there exists an open set U_r such that $A \subset U_0$, $B \subset (X - U_1)$, and for $r < s$, $\overline{U_r} \subseteq U_s$.*

The next theorem is called Urysohn's Lemma, although it is actually an important theorem in its own right. It is called a lemma, because it first appeared in a paper in which Urysohn used it to prove a theorem about the existence of metrics on certain kinds of spaces. We will discuss such metric spaces and see the theorem for which Urysohn's Lemma is a lemma in Chapter 10. Urysohn's Lemma essentially says that a space is normal if and only if disjoint closed sets can be "separated" by a continuous function, i.e., it takes one set to 0 and the other set to 1.

Urysohn's Lemma 8.57. *A topological space X is normal if and only if for each pair of disjoint closed sets A and B in X, there exists a continuous function $f : X \to [0, 1]$ such that $A \subset f^{-1}(0)$ and $B \subset f^{-1}(1)$.*

The proof of Urysohn's Lemma provides important insight into the relationship between continuous functions and nested open sets. In a profound way, it allows you to understand continuous functions at a deeper level, namely, by thinking about how the target space of a continuous function reflects its structure back into the domain space via the inverse images of sets. That statement is a bit ethereal; however, thinking carefully about it can give you a revealing understanding of continuity. In the case of continuous functions from a space X into $[0,1]$, continuity is exactly equivalent to the existence of the nested open sets described in Lemma 8.56. One consequence of that understanding is that it will allow you to prove the Tietze Extension Theorem.

Here's a lemma whose proof uses the Normality Lemma 5.29. Remember to draw a picture.

Lemma 8.58. *Let X be a normal space, and let A be a closed subset of X. Let $f : A \to [0,1]$ be a continuous function and let $r \in (0,1)$. Then there exist disjoint open sets U_r and V_r such that $f^{-1}([0,r)) \subset U_r$ and $f^{-1}((r,1]) \subset V_r$. Or equivalently, there exists an open set U_r such that $f^{-1}([0,r)) \subset U_r$ and $\overline{U_r} \cap f^{-1}((r,1]) = \varnothing$.*

Now use the lemma above to produce the nested sequence of open sets such as those in Lemma 8.56 that then allow you to define a function from a space X to $[0,1]$.

Tietze Extension Theorem 8.59. *A space X is normal if and only if for every closed set $A \subset X$ and continuous function $f : A \to [0,1]$, there exists a continuous function $F : X \to [0,1]$ such that $F(x) = f(x)$ for every $x \in A$.*

A function such as F in the Tietze Extension Theorem is said to **extend** f, meaning that F is defined on a larger set than f is defined on and they agree on the points on which f is defined.

> **Effective Thinking Principle.** *Understand Proofs, Not Just Statements, of Theorems.* Understanding the proofs of theorems rather than just the statements of theorems gives you power.

Your proof of the Tietze Extension Theorem may have used Lemma 8.56. However, alternative proofs can be created by applying the statement of Urysohn's Lemma repeatedly (rather than the proof of Urysohn's Lemma) to get a sequence of functions that converge to the desired function. As a hint to an alternative proof of the Tietze Extension Theorem, you might consider having the initial target interval be the interval $[-1,1]$ (instead of $[0,1]$), and think about applying Urysohn's Lemma to produce a function $g_1 : X \to [-1/3, 1/3]$ where the closed set $f^{-1}([-1,-1/3])$ goes to $-1/3$ and the closed set $f^{-1}([1/3,1])$ goes to $1/3$. Notice that for each $x \in A$, $|f(x) - g_1(x)| < 2/3$. So the function $f(x) - g_1(x) : A \to [-2/3, 2/3]$ is now a new function into an interval. You can now apply Urysohn's Lemma again in a similar way to produce a function $g_2 : X \to [-2/3 \times 1/3, 2/3 \times 1/3]$ such that for each point $x \in A$, $|f(x) - (g_1(x) + g_2(x))| < (2/3)^2$. The g_i functions are functions on all of X and the sum of the g_i functions becomes increasingly close to the function f on points of A. This strategy of proof uses insights about convergence. This approach to the Tietze Extension Theorem is nicknamed the 'Tricky 1/3's Proof.'

In some sense, the proof of the Tietze Extension Theorem developed by constructing appropriately nested open sets takes advantage of an understanding of continuity and an understanding of a proof of Urysohn's Lemma, specifically, of how the existence of a collection of nested open sets indexed by the rational numbers in $[0,1]$ creates a continuous function from X into $[0,1]$. The Tricky 1/3's Proof relies on a very clever way to repeatedly apply the statement of Urysohn's Lemma to produce a sequence of continuous functions that converge to the desired extension.

Effective Thinking Principle. *Extend Results.* After proving a theorem, see if it can be improved or modified.

The Tietze Extension Theorem above talks about extending maps from a closed subset of a space into a closed interval. A natural series of questions to ask is what other target spaces might have similar results. Here are some.

Theorem 8.60. *A space X is normal if and only if for every closed set $A \subset X$ and continuous function $f : A \to (0,1)$, there exists a continuous function $F : X \to (0,1)$ such that $F(x) = f(x)$ for every $x \in A$.*

Theorem 8.61. *A space X is normal if and only if for every closed set $A \subset X$ and continuous function $f : A \to [0,1)$, there exists a continuous function $F : X \to ([0,1)$ such that $F(x) = f(x)$ for every $x \in A$.*

Theorem 8.62. *A space X is normal if and only if for every closed set $A \subset X$ and continuous function $f : A \to [0,1] \times [0,1]$, there exists a continuous function $F : X \to [0,1] \times [0,1]$ such that $F(x) = f(x)$ for every $x \in A$.*

Theorem 8.63. *A space X is normal if and only if for every closed set $A \subset X$ and continuous function $f : A \to \prod_{\alpha \in \lambda} [0,1]_\alpha$, where each $[0,1]_\alpha$ is a copy of $[0,1]$ in the usual topology, there exists a continuous function $F : X \to \prod_{\alpha \in \lambda} [0,1]_\alpha$ such that $F(x) = f(x)$ for every $x \in A$.*

The Tietze Extension Theorem has some interesting consequences.

Theorem 8.64. *Let X be a normal space and let A be a closed subspace of X homeomorphic to $[0,1]$ with the usual topology. Then there exists a continuous function $r : X \to A$ such that for every $x \in A$, $r(x) = x$.*

A continuous map such as r above that takes a whole space to a subset leaving the points of the subset fixed is called a **retract** or a **retraction**.

Of course, the theorem is also true if the closed set A is homeomorphic to any of the other target spaces in the variations of the Tietze Extension Theorem.

With a little bit of cleverness you can prove the following:

Theorem 8.65. *Let X be a normal space and let A be a closed subspace of X homeomorphic to S^1 with the usual topology. Then there exists an open set U containing A and a continuous function $r : U \to A$ such that for every $x \in A$, $r(x) = x$.*

Exercise 8.66. *Think of (many) other possible alternatives to S^1 in the preceding theorem that would allow you to draw the same conclusion.*

Now let's turn our attention toward some features of closed sets that arise during our investigation of these maps from spaces into $[0, 1]$. If you have a continuous function $f : X \to [0, 1]$, then the inverse image of an interval, such as $f^{-1}([0, 1/2])$, is the intersection of a countable number of open sets. In that case, the intersection of the open sets $f^{-1}([0, 1/2 + 1/n))$ for n $\mathbb{N}$ would equal $f^{-1}([0, 1/2])$. That suggests the following definition.

Definition. A subset A of a space X is an F_σ set if and only if $A = \cap U_i$ where $i \in \mathbb{N}$ and each U_i is open. In other words, A is the intersection of countably many open sets.

If every closed set in space is an F_σ, then we have a stronger version of normality.

Definition. A space X is **perfectly normal** if and only if X is normal and every closed subset of X is an F_σ.

In perfectly normal spaces, the conclusion of Urysohn's Lemma can be strengthened to replace containment by equality.

Theorem 8.67. *Suppose X is perfectly normal. Then for each pair of disjoint closed sets A and B in X, there exists a continuous function $f : X \to [0, 1]$ such that $A = f^{-1}(0)$ and $B = f^{-1}(1)$.*

In perfectly normal spaces, separated sets can be separated by disjoint open sets.

Theorem 8.68. *Every perfectly normal space is completely normal.*

We conclude this section on Urysohn's Lemma and the Tietze Extension Theorem by seeing that many topological spaces are homeomorphic to subsets of products of intervals. The following theorem says that every normal, T_1 space can be embedded in a product of intervals, i.e., a cube of sufficiently high (possibly uncountable) dimension. Urysohn's Lemma will be useful.

Theorem 8.69. *Let X be a normal, T_1 space. Then X is homeomorphic to a subspace of $\prod_{\alpha\in\lambda}[0, 1]_\alpha$ for some λ, where each factor is the unit interval with the standard topology.*

In the proof of the above theorem you probably used Urysohn's lemma, but you really only needed a special case where one of the closed sets was a point. This insight suggests we could get by with a weaker condition than normality.

Definition. A space X is **completely regular** if and only if for each point p and closed set A with $p \notin A$, there is a continuous function $f : X \to [0,1]$ such that $f(p) = 0$ and for every point $y \in A$, $f(y) = \{1\}$.

Scholium 8.70. *A space X is completely regular and T_1 if and only if X can be embedded in $\prod_{\alpha\in\lambda}[0,1]_\alpha$ for some λ, where each factor is the unit interval with the standard topology.*

Note. A scholium is a theorem whose truth is an immediate consequence of the proof of a preceding theorem, but it does not follow from the statement of that preceding theorem. By contrast, a corollary is a consequence of the statement of a preceding theorem.

You may wonder whether there is a difference between completely regular and regular. It is not hard to see that completely regular spaces are regular, but not all regular spaces are completely regular (this fact stands in contrast to normality which was equivalent to the fact that disjoint closed sets can be separated by a continuous function). However, counterexamples are difficult to construct.

We conclude this section by seeing an example where open covers of spaces interact with themes involving the Tietze Extension Theorem. Many important spaces have desirable local properties. For example, in the second part of the book, we will discuss at length *surfaces*, or 2-manifolds, which are locally homeomorphic to a plane. The next theorem about the existence of *partitions of unity* is an important tool in other areas of mathematics for taking objects defined over some space X and breaking them into pieces defined locally in each set of a cover. You may find the Tietze Extension Theorem and the Incredible Shrinking Theorem 5.11 helpful to prove the following result.

Recall from the section on paracompactness that locally finite means that every point of X has a neighborhood that intersects only finitely many of the covering sets.

Theorem 8.71. *Given a locally finite open cover $\{U_\alpha\}_{\alpha\in\lambda}$ of a normal, T_1 space X, there is a collection of corresponding continuous functions $\phi_\alpha : X \to [0,1]$ such that (i) each ϕ_α is zero outside U_α, and (ii) the ϕ_α pointwise add to 1. The collection $\{\phi_\alpha\}_{\alpha\in\lambda}$ is called a **partition of unity**.*

So, for example, suppose you have a surface X and you want to define integration of a function $f : X \to \mathbb{R}$. You can take a locally finite open cover of X and use a partition of unity $\{\phi_\alpha\}$ to produce functions $f(x)\phi_\alpha(x)$, each defined on one set of the cover, then integrate each $f(x)\phi_\alpha(x)$ as a function over a piece of a plane.

8.7 Continuity—Functions that Know Topology

After defining any mathematical structure, a natural impulse is to map one such object to another in an appropriate way. In the case of topology, 'appropriate' was not the first thing you thought of, but it was the second. At first, considering inverse images of open sets to be open rather than making images of open sets to be open seemed backwards, but by now it is second nature.

Exploring continuity was a model of how mathematics is created. In previous chapters various properties of topological spaces were developed. So in this chapter it was natural to investigate how the properties of spaces played with maps from space to space. Many natural questions were of the form: "If I have a space with this property, will its image under a continuous function have that property too?" Other connections between properties and continuous functions led to defining new spaces via quotient maps and Urysohn's Lemma and the Tietze Extension Theorem, all of which we will see much more in the chapters to come.

This creation and development of the concept of continuity demonstrates why mathematics has such rich and boundless potential for unlimited growth.

Chapter 9

Connectedness: When Things Don't Fall into Pieces

The idea of a space being *connected* is one of the most readily apparent topological properties a space can have. Does Los Angeles connect to Anchorage by a set of freeways? That is a topological question. We may not be able to tell right away if a space is compact or Hausdorff, but we have an intuitive idea whether it is connected or not. In this chapter we will study how to capture this intuition. Before you continue, try to formulate a definition of what it means for a space being connected. Do *not* read ahead until you have written down at least one attempt at a definition.

Now let's look at some examples of spaces that might motivate definitions of connectedness and highlight potential issues.

Consider the space $(0,1)$ and the space $(0,1) \cup (2,3)$ in the subspace topology of $\mathbb{R}_{std}$. Intuitively $(0,1)$ looks connected and $(0,1) \cup (2,3)$ looks to be not connected, so we want our definition to agree with that intuition.

Less clear is the space $(4,5) \cup (5,6) \subset \mathbb{R}_{std}$. This space also seems to be in two pieces, so we want the definition of connected to declare this space to be not connected.

The next example may be more difficult to decide. Consider the closure of the *topologist's sine curve* in $\mathbb{R}^2_{std}$. The **topologist's sine curve** is defined as this set:

$$S = \left\{ \left(x, \sin\left(\frac{1}{x}\right) \right) | x \in (0,1) \right\}.$$

According to your definition, is the closure of S connected or not? A related space we will refer to for future examples is **the topologist's comb**:

$$C = \{(x,0) | x \in [0,1]\} \cup \bigcup_{n=1}^{\infty} \left\{ \left(\frac{1}{n}, y \right) \mid y \in [0,1] \right\}.$$

According to your definition, is $C \cup \{(0,1)\}$ connected or not?

It turns out that there are several reasonable definitions that capture different concepts associated with the intuitive idea of connectedness. One natural view of being connected is the idea of being able to "walk", so to speak, from any point to any other point without leaving the space. The closure of the topologist's sine curve fails this notion of connectedness since it is impossible to continuously move from a point on the wiggly side of the topologist's sine curve to a point on the limiting vertical line segment. This concept of "walkability" captures the idea of being *path connected* or *arcwise connected*, and we will study it in a later section.

However, there is a more basic idea of connectedness captured by the question of whether the space is in two disjoint pieces, where the pieces are open sets. This perspective gives the basic definition of connectedness for topological spaces.

9.1 Connectedness

Definition. Let X be a topological space. Then X is **connected** if and only if X is not the union of two disjoint non-empty open sets.

We defined the notion of *separated* sets earlier but we will remind you of the definition here.

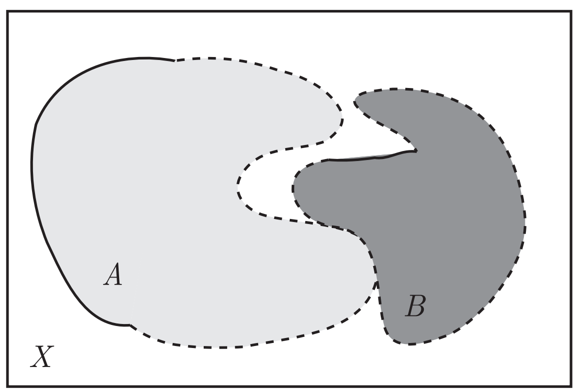


Figure 9.1: Separated sets A and B.

Definition. Let X be a topological space. Subsets A, B in X are **separated** if and only if $\overline{A} \cap B = A \cap \overline{B} = \varnothing$. Thus B does not contain any limit points of A, and A does not contain any limit points of B. The notation $X = A \mid B$ means $X = A \cup B$ and A and B are separated sets.

> **Effective Thinking Principle.** *Seek Equivalences.* After isolating a concept or a definition, seek equivalences with the goal of understanding the idea from different perspectives.

Theorem 9.1. *The following are equivalent:*

1. X *is connected.*
2. *there is no continuous function* $f : X \to \mathbb{R}_{std}$ *such that* $f(X) = \{0, 1\}$.
3. X *is not the union of two disjoint non-empty separated sets.*
4. X *is not the union of two disjoint non-empty closed sets.*
5. *the only subsets of* X *that are both closed and open in* X *are the empty set and* X *itself.*
6. *for every pair of points* p *and* q *and every open cover* $\{U_\alpha\}_{\alpha \in \lambda}$ *of* X *there exist a finite number of the* U_α*'s,* $\{U_{\alpha_1}, U_{\alpha_2}, U_{\alpha_3}, ..., U_{\alpha_n}\}$ *such that* $p \in U_{\alpha_1}$*,* $q \in U_{\alpha_n}$ *and for each* $i < n$*,* $U_{\alpha_i} \cap U_{\alpha_{i+1}} \neq \varnothing$.

Effective Thinking Principle. *Look at Examples to Make Definitions Meaningful.* After creating or encountering a definition, look at a range of specific examples to help you understand the definition more deeply.

Let's look at some examples to understand some features of the definition of connectedness and to start to develop an intuition about it.

Exercise 9.2. *Which of the following spaces are connected?*

1. $\mathbb{R}$ *with the discrete topology?*
2. $\mathbb{R}$ *with the indiscrete topology?*
3. $\mathbb{R}$ *with the finite complement topology?*
4. $\mathbb{R}_{LL}$*?*
5. $\mathbb{Q}$ *as a subspace of* $\mathbb{R}_{std}$*?*
6. $\mathbb{R} - \mathbb{Q}$ *as a subspace of* $\mathbb{R}_{std}$*?*

Since $\mathbb{R}_{std}$ is our paradigm for a connected space, our definition of connectedness had better make $\mathbb{R}_{std}$ connected. Although this fact appears intuitively obvious, to prove it you will need to use the least upper bound property of the reals, or something equivalent.

Theorem 9.3. *The space* $\mathbb{R}_{std}$ *is connected.*

Effective Thinking Principle. *Connect New Ideas with Old Ideas.* After isolating a new idea, consider previous concepts and constructions and explore how the new idea relates to them.

We have now defined connectedness. Let's first see how connectedness relates to basic set theoretic ideas such as subsets and unions; basic topological ideas such as closure; and familiar topological relationships such as products and continuous functions. The next several theorems begin that exploration.

Theorem 9.4. *Let A, B be separated subsets of a space X. If C is a connected subset of $A \cup B$, then either $C \subset A$ or $C \subset B$.*

The union of a collection of connected sets that have a point in common is connected. In fact, more generally:

Theorem 9.5. *Let $\{C_\alpha\}_{\alpha \in \lambda}$ be a collection of connected subsets of X and E be another connected subset of X such that for each α in λ, $E \cap C_\alpha \neq \varnothing$. Then $E \cup \left(\bigcup_{\alpha \in \lambda} C_\alpha\right)$ is connected.*

If you start with a connected subspace of a topological space and add limit points, the resulting subspace is connected.

Theorem 9.6. *Let C be a connected subset of the topological space X. If D is a subset of X such that $C \subset D \subset \overline{C}$, then D is connected.*

This theorem allows us to settle the question of the connectedness of the closure of the topologist's sine curve.

Exercise 9.7. *Show that the closure of the topologist's sine curve in $\mathbb{R}^2_{std}$ is connected.*

The following theorem can be used to produce connected subsets of a space.

Theorem 9.8. *Let X be a connected space, C a connected subset of X, and $X - C = A \mid B$. Then $A \cup C$ and $B \cup C$ are each connected.*

Next we consider how connectedness interacts with products and continuous functions. Products of connected spaces are connected.

Theorem 9.9. *For topological spaces X and Y, $X \times Y$ is connected if and only if each of X and Y is connected.*

Theorem 9.10. *For spaces $\{X_\alpha\}_{\alpha \in \lambda}$, $\prod_{\alpha \in \lambda} X_\alpha$ is connected if and only if for each α in λ, X_α is connected.*

Of course, this theorem confirms that the standard Euclidean spaces $\mathbb{R}^n$ are connected.

We have seen several reasons why the box topology is not the most useful concept to use in giving a topology to infinite products. Yet another reason is that the box topology of infinitely many connected spaces need not be connected.

Exercise 9.11. *Show that the box product of countably infinitely many copies of* $\mathbb{R}_{std}$ *is not connected.*

Connectedness is preserved by continuous functions.

Theorem 9.12. *Let* $f : X \to Y$ *be a continuous, surjective function. If* X *is connected, then* Y *is connected.*

Here's a theorem you may have proved before, but now you can prove it using connectedness.

Theorem 9.13. *(Intermediate Value Theorem) Let* $f : \mathbb{R}_{std} \to \mathbb{R}_{std}$ *be a continuous map. If* $a, b \in \mathbb{R}$ *and* r *is a point of* $\mathbb{R}$ *such that* $f(a) < r < f(b)$ *then there exists a point* c *in* (a, b) *such that* $f(c) = r$.

9.2 Cardinality, Separation Properties, and Connectedness

Now we explore which spaces with relative few points might be connected. This next theorem may be a challenge. It states that if you have only countably many points, and you have enough separation properties, then the space must be expressible as the union of two disjoint open sets.

Theorem 9.14. *Let* X *be a countable, regular,* T_1 *space with more than one point. Then* X *is not connected.*

Effective Thinking Principle. *Weaken Hypotheses and See What Happens.* One good habit for a mathematician is to see whether the hypotheses of a theorem can be weakened. Either a better theorem or an illuminating counterexample might result.

In the case of our theorem that countable, regular, T_1 spaces are not connected, trying to weaken the regularity hypothesis leads to the difficult challenge of constructing a countable, Hausdorff space that is connected. One such example is called *Bing's Sticky Foot Topology*. We will describe it here and ask you to verify that it is indeed connected and Hausdorff.

Example. **Bing's Sticky Foot Topology**: The points of this space consist of the points in the upper half plane (including the x-axis) with rational-rational coordinates. Basic open sets for this topology come in two types:

1. Subsets of the x-axis that are relatively open in $\mathbb{R}^1_{std}$ are basic open sets.

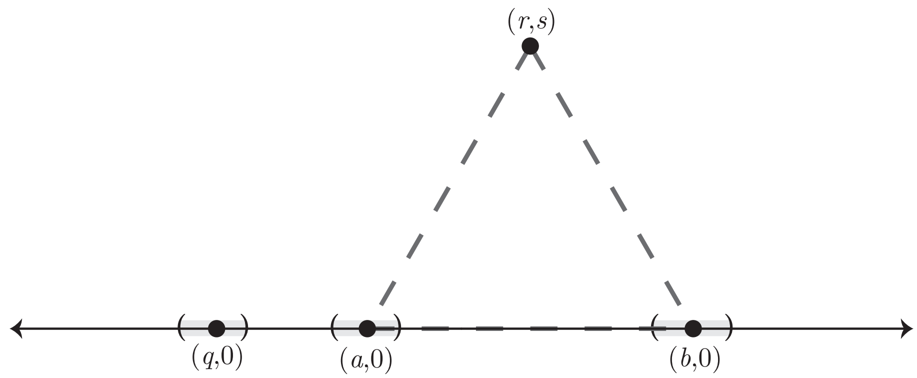


Figure 9.2: Basic open sets in Bing's Sticky Foot Topology.

2. To construct a basic open set containing the point (r, s) where $0 < s$, first find the two points on the x-axis $(a, 0)$ and $(b, 0)$ such that the three points (r, s), $(a, 0)$, and $(b, 0)$ are the vertices of an equilateral triangle. A basic open set containing the point (r, s) where $0 < s$ consists of (r, s) together with the rational points in open intervals in the x-axis about the points $(a, 0)$ and $(b, 0)$. If you draw the lines from (r, s) down to $(a, 0)$ and $(b, 0)$ and draw open intervals around $(a, 0)$ and $(b, 0)$ on the x-axis, you can see where the Sticky Foot name comes from.

Exercise 9.15. *Show that Bing's Sticky Foot Topology is a countable, connected, Hausdorff space.*

Bing's Sticky Foot Topology shows that the regularity hypothesis in Theorem 9.14 cannot be weakened. If we strengthen the regularity hypothesis to normality, then connected, T_1 spaces must have at least as many points as does $\mathbb{R}$. Urysohn's Lemma may be helpful in proving this result.

Theorem 9.16. *If X is a normal, T_1 space with more than one point and $|X| < |\mathbb{R}|$, then X is not connected.*

Feel free to skip the next exercise. It is not particularly important, but you might enjoy it. It points out that in fact, for $n \geq 2$, the space $\mathbb{R}^n$ remains connected even when some points are removed.

Exercise 9.17. *Let A be a countable subset of $\mathbb{R}^n$ for $n \geq 2$. Show that $\mathbb{R}^n - A$ is connected. In fact, if the cardinality of A is any cardinality less than the cardinality of $\mathbb{R}$, then $\mathbb{R}^n - A$ will still be connected. Actually, for any two points p and q in $\mathbb{R}^n - A$, p can be connected to q by two intersecting straight line segments in $\mathbb{R}^n - A$.*

9.3 Components and Continua

If a space is not connected, it is natural to think about the connected pieces that make it up. That impulse leads to the definition of a component of a space.

Definition. Let X be a space and $p \in X$. The **component** or *connected component* of p in X is the union of all connected subsets of X that contain p.

The basic facts about components are that they are connected, closed, maximal, and disjoint.

Theorem 9.18. *Each component of X is connected, closed, and not contained in any strictly larger connected subset of X.*

Theorem 9.19. *The set of components of a space X is a partition of X.*

> **Effective Thinking Principle.** *Consider Opposite Questions.* When studying a topic, it is often illuminating to consider the opposite questions.

We are undertaking an investigation of connected spaces, so for a moment let's think about the opposite question, namely, in how many ways can a space that is not connected be shown to be not connected. The investigation of connectedness is particularly rich and interesting in compact, Hausdorff spaces. So we turn our attention now to considering ways in which disconnected compact, Hausdorff spaces can be expressed as disjoint open sets. It turns out that compact, Hausdorff spaces can often be written as the union of two disjoint open sets in many ways. The theorem below states that if a compact, Hausdorff space is not connected and no component of the space intersects two given disjoint closed sets in the space, then the space can be written as the union of two disjoint open (and closed) subsets where one of the closed sets is in one half and the other closed set is in the other. In other words, in compact, Hausdorff spaces, basically unless a single component prevents such a separation, then such a separation is possible. The lemmas that precede the theorem may be useful in proving this theorem.

> **Effective Thinking Principle.** *Start With Simple Cases.* George Polya said, "If there is a hard problem you can't do, there is an easier problem you can't do. Find it."

We begin by aiming to prove if there is a compact, Hausdorff space with components A and B, then there is a separation of X into two disjoint closed subsets where A is in one and B is in the other. As a start, we notice an equivalent statement that may be easier to prove.

Lemma 9.20. *Let X be a topological space and let $\{H_\alpha\}_{\alpha\in\lambda}$ be the set of subsets of X that are both open and closed. Then the following are equivalent:*

1. *For every two components A and B of X, there exists a separation of X into two disjoint closed sets such that A is in one and B is in the other.*

2. *For every component A of X, $\bigcap\{H_\alpha \mid A \subset H_\alpha\} = A$.*

The reason that the second equivalence may be easier in the case of compact, Hausdorff spaces X is that we know that $\bigcap\{H_\alpha \mid A \subset H_\alpha\}$ is definitely a closed set that contains A. If that intersection were not just equal to A, then that intersection would be a closed subset that is not connected. So that intersection could be written as the union of two disjoint, non-empty closed sets, which would be closed in all of X. Since X is compact and Hausdorff, we know X is normal. When two disjoint closed subsets of a normal space arise, we certainly cannot resist the temptation to put them in disjoint open sets. Succumbing to that temptation is a great idea.

It is also a great idea to remember the following fact about compact spaces that was one of the early theorems in the chapter on compactness.

Lemma 9.21. *Let X be a compact space and let U be an open set in X. Let $\{H_\alpha\}_{\alpha\in\lambda}$ be closed subsets of X such that $\bigcap_{\alpha\in\lambda} H_\alpha \subset U$. Then there exist a finite number of the H_α's whose intersection lies in U.*

Perhaps these lemmas will allow you to prove that disjoint components of a compact, Hausdorff space can be put in different halves of a separation.

Lemma 9.22. *Let A and B be components of a compact, Hausdorff space X. Then $X = H \mid K$ where $A \subset H$ and $B \subset K$.*

In some sense, proving this case is a bit like establishing a Hausdorff-like condition where the components of X are viewed as points. That intuition could be formulated precisely as follows using the idea of identification spaces.

Theorem 9.23. *Let X be a compact, Hausdorff space. Let X^* be the partition of X into its components. Then the identification space X^* is a compact, Hausdorff space.*

Theorem 9.24. *Let A and B be closed subsets of a compact, Hausdorff space X such that no component intersects both A and B. Then $X = H \mid K$ where $A \subset H$ and $B \subset K$.*

As usual, after proving a theorem, we investigate whether all the hypotheses are necessary. The following example will demonstrate the necessity of the "compactness" hypothesis of Theorem 9.24.

Example. Let X be the subset of $\mathbb{R}^2$ equal to $([0,1] \times \bigcup_{i\in\omega_0}\{1/i\}) \bigcup\{(0,0),(1,0)\}$ with the subspace topology. Show that the conclusion to Theorem 9.24 fails when $A = \{(0,0)\}$ and $B = \{(1,0)\}$.

We next turn our attention toward compact, Hausdorff spaces that are connected. Since such spaces are in some sense generalizations of a closed interval in $\mathbb{R}$, they are called continua.

Definition. A **continuum** is a connected, compact, Hausdorff space.

We find that our investigation of non-connected, compact, Hausdorff spaces will be useful in understanding some of the structure of continua. First we prove that each component of an open subset of a continuum must extend out to the boundary of that open set.

Theorem 9.25. *Let U be a proper, open subset of a continuum X. Then each component of $\overline{U}$ contains a point of ∂U, the boundary of U. (Recall: $\partial U = \overline{U} - U$.)*

The next, similar, theorem is often referred to as the *"to the boundary" theorem*.

Theorem 9.26. *Let U be a proper, open subset of a continuum X. Then each component of U has a limit point on ∂U.*

We saw earlier that no regular, T_1 space with only a countable number of points could be connected. In the same spirit, the following theorem states that no continuum is the union of countably many disjoint closed sets.

Theorem 9.27. *No continuum X is the union of a countable number (> 1) of disjoint, non-empty closed subsets.*

As usual, we investigate the necessity of the hypotheses. In this case, we construct an example to demonstrate the necessity of the compactness hypothesis on X.

Exercise 9.28. *Show that in Figure 9.3, X is connected and is the union of a countable number of disjoint closed sets.*

The nested intersection of continua is a continuum. This innocuous-sounding result will allow us to construct some truly strange continua.

Theorem 9.29. *Let $\{C_i\}_{i\in\omega}$ be a collection of continua such that for each i, $C_{i+1} \subset C_i$. Then $\bigcap_{i\in\omega} C_i$ is a continuum.*

Theorem 9.30. *Let $\{C_\alpha\}_{\alpha\in\lambda}$ be a collection of continua indexed by a well-ordered set λ such that if $\alpha < \beta$, then $C_\beta \subset C_\alpha$. Then $\bigcap_{\alpha\in\lambda} C_\alpha$ is a continuum.*

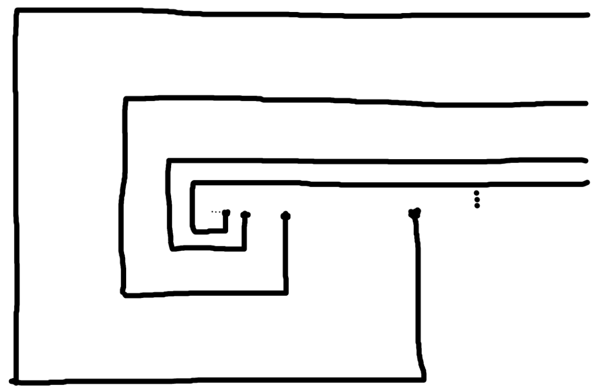


Figure 9.3: A connected space, the union of a countable number of disjoint sets.

Closed intervals suggested the name of continua, so it is reasonable to see whether every continuum has some features in common with a closed interval. The endpoints of a closed interval suggest an interesting theorem. If you delete an endpoint from a closed interval, the remainder of the interval remains connected. It turns out that every continuum has non-separating points.

Definition. Let X be a connected set. A point p in X is a **non-separating point** if and only if $X - \{p\}$ is connected. Otherwise p is a **separating** point.

Lemma 9.31. *Let X be a continuum, p be a point of X, and $X - \{p\} = H \mid K$. Then $H \cup \{p\}$ is a continuum and if $q \neq p$ is a non-separating point of $H \cup \{p\}$, then q is a non-separating point of X.*

This lemma will help you to prove the following theorem about the existence of non-separating points. Delightfully, the coming theorem confusingly involves four different kinds of "separation", as if our topology colleagues could not get enough usage out of the Latin root *separat*. We have a *separable* space (having a countable dense subset) which is Hausdorff (which is a *separation* property) and connected (that is, having no *separation*), and we wish to show it has non-*separating* points. What fun!

Theorem 9.32. *Let X be a separable continuum with more than one point. Then X has at least two non-separating points.*

The following theorem is a strict generalization of the preceding one, namely, it omits the separable hypothesis. To prove it, you can follow the same strategy as you probably used to prove the preceding theorem, but in this case, you may need to use transfinite induction.

Theorem 9.33. *Let X be a continuum with more than one point. Then X has at least two non-separating points.*

We must not abandon our investigation of continua without introducing you to a few of the exotic delights of this realm. The first one we will describe starts with the totally disconnected Cantor Set (see Section 8.4 for the definition). We then produce a connected set by joining pairs of points in the Cantor Set with semi-circles to create a fascinating example called the Knaster continuum or bucket-handle continuum.

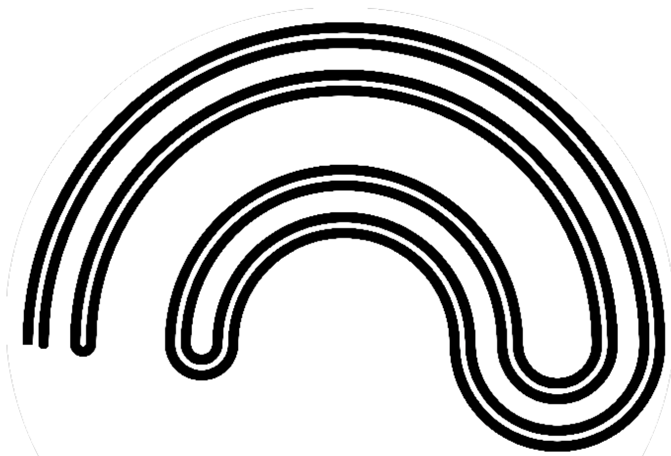


Figure 9.4: A sketch of the Knaster bucket-handle continuum.

Example. The **Knaster continuum** or **Bucket handle continuum** is a closed subset of $\mathbb{R}$ described as follows:

1. Begin with the Standard Cantor in $[0, 1]$. In the upper half plane add all the semi-circles centered at the point $(1/2, 0)$ that contain a point of the Cantor Set. So each of these semi-circles connects a pair of points of the Cantor Set, namely, every point of the Cantor Set in the interval $[0, 1/3]$ is paired with a point in the Cantor Set in the interval $[2/3, 1]$.

2. Next we add more collections of semi-circles in the lower half-plane, but centered at different points. The first collection of cup-shaped semi-circles are each centered at the point $(5/6, 0)$. These semi-circles pair points in the Cantor Set from the interval $[2/3, 7/9]$ with points in the Cantor set in the interval $[8/9, 1]$.

3. Now we proceed to the left. The second collection of cup-shaped semi-circles are each centered at the point $(5/18, 0)$, which is the midpoint of the interval $[2/9, 1/3]$. These semi-circles

pair points in the Cantor Set from the interval $[2/9, 7/27]$ with points in the Cantor set in the interval $[8/27, 1/3]$.

4. We proceed in this manner creating collections of cup-shaped semi-circles that pair points of the Cantor Set. Just to show that we can do the arithmetic, in general, for each $n \in \mathbb{N}$, we take all the cup-shaped semi-circles in the lower half plane centered at the point $(5/(2 \times 3^n), 0)$ that connect points in the Cantor Set in the interval $[6/3^{n+1}, 7/3^{n+1}]$ to the points in the Cantor Set in the interval $[8/3^{n+1}, 9/3^{n+1}]$.

5. The union of all the semi-circles above is the Knaster Continuum or bucket handle continuum.

One of the interesting properties about the bucket handle continuum is that it is impossible to write it as the union of two proper subcontinua. That property is sufficiently interesting to deserve a name.

Definition. A continuum X is **indecomposable** if and only if X cannot be written as the union of two proper subcontinua.

Theorem 9.34. *The bucket handle continuum is indecomposable.*

It is with sadness that we move on from a further discussion of indecomposable continua. We would love to cheerfully marvel at the homogeneity of the pseudo-arc or joyfully frolic in the Lakes of Wada. However, life is stern and life is earnest, so we must leave this verdant field for other equally rich delights.

We leave this section with a question that has been called the most interesting unsolved problem in plane topology. Suppose you have a continuum C in $\mathbb{R}^2$ that does not separate $\mathbb{R}^2$. Then does C have the fixed point property, that is, for every continuous function $f : C \to C$ is there a point $x \in C$ such that $f(x) = x$? This question has remained unanswered for nearly a century. Please solve it.

9.4 Path or Arcwise Connectedness

Perhaps one of the ways you thought of defining connectedness at the beginning of the chapter was the idea of being able to "walk" from any point of the set to any other point without leaving the set. This intuitive idea leads to the property called *path connectedness* or *arcwise connectedness*.

Definition. A *path* from x to y in a space X is a continuous map $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$. A space X is **arcwise connected** or **path connected** if every pair of points in X can be joined by a path in X.

The first thing to notice is that spaces that connect points by paths are definitely connected.

Theorem 9.35. *A path connected space is connected.*

However, not every space that is connected is path connected.

Example. The **flea and comb space** is the union of the topologist's comb with the point $(0, 1)$.

Exercise 9.36. *The flea and comb space is connected but not path connected.*

Exercise 9.37. *The closure of the topologist's sine curve is connected but not path connected.*

Path connectivity behaves well with respect to products.

Theorem 9.38. *The product of path connected spaces is path connected.*

Definition. Given a topological space X, define an equivalence relation on X by letting $x \sim y$ if there is a path connected subset of X containing both x and y. The equivalence classes are called the **path components** of X.

Exercise 9.39.

1. *What are the path components of the closure of the topologist's sine curve?*
2. *What are the path components of the closure of the topologist's comb?*

Exercise 9.40. *Must every non-empty open connected subset U of $\mathbb{R}^n$ be path connected?*

Connecting two points by a path just requires the existence of a continuous function, but you might prefer an embedded arc. For Hausdorff spaces, the existence of a path implies the existence of an embedded arc.

Theorem 9.41. *Let p and q be two points in a Hausdorff space X such that there exists a continuous function $f : [0, 1] \to X$ with $f(0) = p$ and $f(1) = q$. Then there exists an embedding $h : [0, 1] \to X$ with $h(0) = p$ and $h(1) = q$.*

9.5 Local Connectedness

Although the closure of the topologist's comb C is connected and path connected, if we look at a small neighborhood of the point $(0, 1)$, its intersection with C is neither connected nor path connected.

Definition. A space X is **locally connected at the point** p of X if and only if for each open set U containing p, there is a connected open set V such that $p \in V \subset U$. A space X is **locally connected** if and only if it is locally connected at each point.

Theorem 9.42. *The following are equivalent:*

1. *X is locally connected.*

2. *X has a basis of connected open sets.*

3. *For each $x \in X$ and open set U with $x \in U$, the component of x in U is open.*

4. *For each $x \in X$ and open set U with $x \in U$, there is a connected set C such that $x \in$ Int $C \subset C \subset U$.*

5. *For each $x \in X$ and open set U with $x \in U$, there is an open set V containing x and $V \subset$ (the component of x in U).*

Notice that a connected space need not be locally connected and a locally connected space need not be connected.

Exercise 9.43.

1. *Show that the closure of the topologist's comb is not locally connected.*

2. *Construct a space that is connected but not locally connected at any point.*

3. *Find an example of a space that is locally connected but not connected.*

As usual, one of the features we look for with a new property is how it behaves with respect to products.

Theorem 9.44. *The product of two locally connected spaces is locally connected.*

Local connectivity is about the only property where the box product behaves better than the standard product topology.

Exercise 9.45.

1. *Find an example of an infinite number of locally connected spaces where the infinite product space is not connected.*

2. *Prove that an arbitrary box product of locally connected spaces is locally connected.*

One of the standard questions we ask after defining a new property is whether it is preserved under continuous functions.

Theorem 9.46. *Let X be a locally connected space and let $f : X \to Y$ be a continuous, surjective, closed or open map. Then Y is locally connected.*

Exercise 9.47. *Construct an example of a locally connected space X and a continuous, surjective function $f : X \to Y$ such that Y is not locally connected.*

We first defined connected spaces and then we considered path connected spaces. So it is natural to consider the local version of path connectivity.

Definition. Let X be a topological space.

1. X is **locally path connected** or **locally arcwise connected at** p if and only if for each open set U containing p there is an open set V such that $p \in V \subset U$ such that each pair of points $x, y \in V$ can be joined by a path in U.

2. A space is **locally path connected** or **locally arcwise connected** if and only if it is locally path connected at each point.

Theorem 9.48. *A locally path connected space is locally connected.*

Notice in the definition of locally arcwise connected that the points in the open set V are connected by a path in U, but not by a path in V. That raises the question of why the definition did not insist that those paths must lie in V. The answer is that we could have made that definition.

Theorem 9.49. *The following are equivalent:*

1. *X is locally arcwise connected.*

2. *For each $x \in X$ and open set U with $x \in U$, there is an arcwise connected open set V such that $x \in V \subset U$.*

3. *X has a basis of connected, arcwise connected open sets.*

The closed interval is 2^{nd} countable, compact, Hausdorff, connected, and locally connected. That collection of properties defines spaces known as Peano continua.

Definition. A topological space X is a **Peano continuum** if and only X is compact, Hausdorff, 2^{nd} countable, connected, and locally connected.

Notice that we could have saved some words in the definition of a Peano continuum by defining it as a 2^{nd} countable, locally connected continuum. Then you could have practiced remembering that a continuum is a compact, connected, Hausdorff space.

To prove the following characterization of Peano continua, you will have an opportunity to remember and apply several previous theorems about what features of spaces are preserved by continuous functions. Enjoy.

Theorem 9.50. *A Hausdorff space X is a Peano Continuum if and only if it is the image of $[0, 1]$ under a continuous map.*

You may have noticed that we defined locally arcwise connected and it seems to have disappeared in the subsequent discussion, but it was hiding in the background the whole time.

Theorem 9.51. *Let $f : [0, 1] \to X$ be a continuous surjective map where X is Hausdorff. Then X is locally arcwise connected. Equivalently, every Peano Continuum is locally arcwise connected.*

If you look up the phrase Peano Continuum, you will probably read that a Peano Continuum is a locally connected metric continuum. But have no fears, by the end of the next chapter on metric spaces, you will see that the two definitions are equivalent.

Mathematics is an ever-growing enterprise. One famous unsolved problem in mathematics involves Peano Continua—or not. The question is whether the famous fractal known as the Mandelbrot Set is locally connected, that is, whether the Mandelbrot Set is a Peano Continuum.

9.6 Totally Disconnected Spaces and the Cantor Set

Effective Thinking Principle. *Opposite Extremes.* After creating a concept, explore its opposite as well.

This chapter is about connected spaces. So it is fitting that we also explore the opposite kinds of spaces, namely those that are the opposite of connected. As usual when we turn our minds toward an idea, several possibilities arise. One idea is the idea of a totally disconnected space.

Definition. A space X is **totally disconnected** if and only if every component of X is a single point.

Another concept of extreme disconnectedness is the idea of having a basis of sets that are both open and closed.

Definition. A space X is **0-dimensional** if and only if X has a basis of sets each of which is both open and closed.

We have two possible definitions for being the opposite of connected. It is natural to ask whether those two definitions are the same. The first thing to notice is that 0-dimensional, T_1 spaces are definitely totally disconnected.

Theorem 9.52. *Let X be a 0-dimensional, T_1 space. Then X is totally disconnected.*

However, it is possible to have spaces that are totally disconnected, but that are not 0-dimensional.

Exercise 9.53. *Create a Hausdorff space that is totally disconnected but is not 0-dimensional.*

The Cantor set is a famous 0-dimensional space with many interesting properties (see Section 8.4 for the definition).

Theorem 9.54. *The standard Cantor set is precisely those real numbers in $[0,1]$ that can be written using only 0's or 2's in their ternary (that is, base 3) expansion.*

Here are some fun facts about the Cantor set that you may enjoy exploring.

Exercise 9.55. *Show that every real number in $[0,2]$ is the sum of two numbers in the standard Cantor set.*

Exercise 9.56. *Let C be the Cantor set. Create a continuous function $f : C \to [0,1]$ that is surjective.*

Some of the interesting facts about the Cantor set relate to how it can be embedded in Euclidean spaces. Here are some examples. The first one shows that you could create an umbrella with a Cantor set.

Exercise 9.57. *Let C be the Cantor set. Create an embedding $h : C \to [0,1] \times [0,1]$ such that for every $x \in [0,1]$, $(\{x\} \times [0,1]) \cap h(C) \neq \varnothing$.*

In fact, you could use the Cantor set to hide from spying eyes.

Exercise 9.58. *Let C be the Cantor set. Create an embedding $h : C \to [-1,1] \times [-1,1] - \{(0,0)\}$ such that every ray from $(0,0)$ straight out to infinity intersects $h(C)$.*

A Cantor set can get in the way of every continuous function.

Exercise 9.59. *Let C be the Cantor set. Create an embedding $h : C \to [0,1] \times [0,1]$ such that for every continuous function $f : [0,1] \to [0,1]$, $G_f \cap h(C) \neq \varnothing$, where G_f is the graph of f.*

As we might expect, a Cantor set cannot separate points.

Exercise 9.60. *Let C be the Cantor set, let $h : C \to \mathbb{R}^2$ be an embedding, and let p and q be points in $\mathbb{R}^2 - h(C)$. Show that you can find a polygonal path from p to q in $\mathbb{R}^2 - h(C)$.*

Again on the intuitive side, every embedding of the Cantor set into $\mathbb{R}^2$ is the same up to a homeomorphism of the plane.

Theorem 9.61. *Let C be the standard Cantor set and let $h : C \to \mathbb{R}^2$ be an embedding. Then there exists a homeomorphism $H : \mathbb{R}^2 \to \mathbb{R}^2$ such that for every $x \in C$, $H(h(x)) = x$.*

Interestingly, the preceding theorem is not true for embeddings into higher dimensional Euclidean spaces. You may want to wait until you have learned some of the techniques from the last half of this book before you prove the following theorem.

Theorem 9.62. *Let C be the standard Cantor set. There exists an embedding $h : C \to \mathbb{R}^3$ such that no homeomorphism $H : \mathbb{R}^3 \to \mathbb{R}^3$ exists where $H(h(C)) = C$.*

You saw earlier that the Cantor set can be mapped onto $[0, 1]$. In fact, the Cantor set can be mapped onto every 2^{nd} countable, compact, Hausdorff space.

Theorem 9.63. *Let C be the standard Cantor set and let X be a 2^{nd} countable, compact, Hausdorff space. Then there exists a continuous, surjective function $f : C \to X$.*

The Cantor set is a fascinating space that illustrates the value of exploring opposites, in this case, the opposite of being connected.

9.7 Hanging Together—Staying Connected

One of the most productive strategies for creating mathematics and for understanding the world better is to start with some intuitive idea and then to create mathematical reflections, refinements, and extensions of that idea.

Connectedness is an idea with clear real-world meaning. Pinning it down topologically started with capturing the most basic topological reflection of connectivity, namely, not falling into two parts. Then the intuition about being able to travel from one place to another led to the idea of arcwise connected spaces. Then we generated additional extensions and refinements by employing the usual strategy of considering local versions of ideas.

The study of connectedness is an excellent example of how mathematical ideas can be created, refined, and extended.

Chapter 10

Metric Spaces: Getting Some Distance

As we mentioned early in this book, topology views the world flexibly—as if made of rubber—and asks: what features are preserved? We developed the ideas of open sets and topologies to capture in a flexible way the idea of points in a space being "close". And we did not need to refer to a notion of "distance" to specify this closeness! Using open sets, we have seen that we can formalize the ideas of convergence, connectedness, and continuity in a flexible way.

However, many of the spaces that we care about in this world do have a precise notion of distance attached to them. Scientists measure how close two points are. Astronomers talk about distance between stars. Genealogists ask how many generations separate two individuals. Biologists can say that two species are genetically close. While distances like these arise naturally from context, a distance is sometimes chosen for convenience: a mathematician can use many different measures of closeness for two objects. Topology can help us understand the relationships among various distance notions.

We will make precise this idea of a distance between points, and call it a *metric*, and then ask how our topological developments relate to a metric, when it exists. Spaces that have a metric are called *metric spaces*.

10.1 Metric Spaces

Metric spaces arise by generalizing the notion of the distance between two points in the familiar Euclidean spaces $\mathbb{R}^n$. The strategy is to look at that familiar idea of distance and cull from it central features, which then become the definition of a metric.

Definition. A **metric** on a set M is a function $d : M \times M \to \mathbb{R}_+$ (where $\mathbb{R}_+$ is the non-negative real numbers) such that for all $a, b, c \in M$, these properties hold:

1. $d(a, b) \geq 0$, with $d(a, b) = 0$ if and only if $a = b$;

2. $d(a, b) = d(b, a)$;

3. $d(a, c) \leq d(a, b) + d(b, c)$.

These three properties are often summarized by saying that a metric is **positive definite**, **symmetric**, and satisfies the **triangle inequality**.

A **metric space** (M, d) is a set M with a metric d.

Example. The function $d(x, y) = |x - y|$ is a metric on $\mathbb{R}$. This measure of distance is the **standard metric** on $\mathbb{R}$.

This metric can be generalized to $\mathbb{R}^n$ in many different ways, as the following examples show.

Exercise 10.1. *Verify that the following are all metrics on $\mathbb{R}^n$.*

1. *The **Euclidean metric** on $\mathbb{R}^n$ is defined by $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.*

2. *The **box metric** on $\mathbb{R}^n$ is defined by $d(\mathbf{x}, \mathbf{y}) = \max_i\{|x_i - y_i|\}$.*

3. *The **taxi-cab metric** on $\mathbb{R}^n$ is defined by $d(\mathbf{x}, \mathbf{y}) = \sum_i\{|x_i - y_i|\}$.*

Show that when $n \geq 2$, these metrics are different.

Effective Thinking Principle. *Explore Extremes.* After making a definition, explore extreme or unusual cases.

Example. On any set M, we can define the **discrete metric** as follows: for any $a, b \in M$, $d(a, b) = 1$ if $a \neq b$ and $d(a, a) = 0$. This metric basically tells us whether two points are the same or different.

Example. Here's a strange metric on $\mathbb{Q}$: for reduced fractions, let $d(\frac{a}{b}, \frac{m}{n}) = \max(|a - m|, |b - n|)$. Which rationals are "close" to one another under this metric?

The idea of a metric space is powerful, because it applies to more than just subsets of $\mathbb{R}^n$. For instance, function spaces are sets in which every point is a function.

Exercise 10.2. *Let X be a compact topological space. Let $\mathcal{C}(X)$ denote the set of continuous functions $f : X \to \mathbb{R}$. We can endow $\mathcal{C}(X)$ with a metric:*

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

and this distance is also sometimes denoted $\|f - g\|$. Check that d is a well-defined metric on $\mathcal{C}(X)$.

A space with a given metric has a natural topology induced by the metric. The standard metric on $\mathbb{R}$ and the Euclidean metric on $\mathbb{R}^n$ give rise to the standard topologies on these spaces in the following way.

Theorem 10.3. *Let d be a metric on the set X. Then the collection of all open balls*

$$\mathcal{B} = \{B(p,\epsilon) = \{y \in X | d(p,y) < \epsilon\} \textit{ for every } p \in X \textit{ and every } \epsilon > 0\}$$

forms a basis for a topology on X.

The topology generated by a metric d on X is called the *d-metric topology* for X.

Definition. A topological space $(X, \mathcal{T})$ is a **metric space** or is **metrizable** if and only if there is a metric d on X such that $\mathcal{T}$ is the d-metric topology. We sometimes write a metric space as (X,d) to denote X with the d-metric topology.

Note that previously we defined a metric space to be a set with a metric while here are defining a topological space to be a metric space under certain conditions. You may be worried about this variety of definitions for the same term, but do not be concerned. Since any set with a metric gives rise to a topology on that set induced by the metric, these two definitions are not in danger of conflicting. The idea of a metric on a set emphasizes a given metric, while the topological definition emphasizes the topology (for which the related metric may or may not be stated explicitly). This book concerns topology, so in what follows, when you read 'X is a metric space', you should think of X as a topological space induced by some metric.

Effective Thinking Principle. *What's the Same; What's Different.* When two mathematical concepts are related, explore the extent to which they correspond.

Any metric generates a topology, but different metrics can generate the same topology.

Exercise 10.4. *On $\mathbb{R}^n$, show that the Euclidean metric, box metric, and taxi-cab metric generate the same topology as the product topology on n copies of $\mathbb{R}_{std}$.*

Exercise 10.5. *Now find a metric on $\mathbb{R}^n$ that does* not *induce the product topology on n copies of $\mathbb{R}_{std}$.*

It is natural to ask what features of a metric can be altered while still generating the same topology.

Theorem 10.6. *For any metric space (X,d), there exists a metric $\bar{d}$ such that d and $\bar{d}$ generate the same topology, yet for each $x, y \in X$, $\bar{d}(x,y) < 1$.*

Thus every set is bounded under $\bar{d}$ even if not all sets are bounded under d. So boundedness of the metric is not a topological property.

Effective Thinking Principle. *Explore How New Ideas Relate to Old Ideas.* A standard and fruitful method for creating and learning mathematics is to explore how a new concept interacts with previous concepts.

Many of the coming theorems in this chapter arise as a natural exploration of how metric spaces are related to topological ideas developed in the previous chapters. We start by looking at subsets of metric spaces and learn that subspaces of a metric space are also metric spaces.

Theorem 10.7. *If X is a metric space and $Y \subset X$, then Y is a metric space.*

Metric spaces have every separation property we have introduced. A caution: there is a rather natural idea for how to prove the normality of a metric space that does not work. So it may be useful to consider the graphs of $y = 1/x$ (for $x > 0$) and $y = 0$ in the plane, which are disjoint closed sets, to make sure your proof would work in that case.

Theorem 10.8. *A metric space is Hausdorff, regular, and normal.*

In fact, metric spaces enjoy the stronger concepts of normality that were introduced earlier.

Theorem 10.9. *A metric space is completely normal and perfectly normal.*

In exploring the relationships of metric spaces to countability properties, we find many interesting connections.

Theorem 10.10. *A metric space is a 1^{st} countable space.*

The next theorem shows that in a metric space, many of the countability properties we learned earlier are equivalent.

Theorem 10.11. *In a metric space X, the following are equivalent:*

1. *X is separable,*
2. *X is 2^{nd} countable,*
3. *X is Lindelöf,*
4. *every uncountable set in X has a limit point,*

One route to proving this theorem is to show (1) implies (2) implies (3) implies (4) implies (1). The last implication may be the hardest. After you prove the theorem, ask yourself which of the implications relies on X being a metric space.

Theorem 10.11 can be used to infer that some spaces are not metrizable. For example, there is no metric on the upper half plane that will generate the sticky bubble topology $\mathbb{H}_{\text{bub}}$, because $\mathbb{H}_{\text{bub}}$ is separable but not 2^{nd} countable (Exercise 6.10). Of course, $\mathbb{H}_{\text{bub}}$ is not normal either. So there are many reasons why $\mathbb{H}_{\text{bub}}$ is not a metric space.

Exercise 10.12. *If you've read about the Souslin property in Section 6.4, then a fifth property can be added to the above theorem: a metric space X has the Souslin property if and only if it has the other properties mentioned in Theorem 10.11.*

Next we consider products of metric spaces and get off to a promising start. To prove the next result, your challenge is to define a metric on the product such that the metric topology on the product is the same as the product topology.

Theorem 10.13. *Let (X, d) and (Y, e) be metric spaces. Then $X \times Y$ is a metric space.*

In fact the countable product of metric spaces is a metric space.

Theorem 10.14. *Let $\{(X_i, d_i)\}_{i \in \omega_0}$ be a countable collection of metric spaces. Then $\prod_{i \in \omega_0} X_i$ is metrizable.*

Hint: Consider defining the metric on the product as $d(\mathbf{x}, \mathbf{y}) = \sup_i \frac{\bar{d}_i(x_i, y_i)}{i}$, where $\bar{d}(a, b) = \min\{d(a, b), 1\}$.

But uncountable products of spaces are never metrizable.

Exercise 10.15. *Show that if $\{X_\alpha\}_{\alpha \in \lambda}$ is an uncountable collection of non-degenerate spaces, then $\prod_{\alpha \in \lambda} X_\alpha$ is not metrizable.*

As usual, the box product does not behave well with respect to metrizability.

Exercise 10.16. *Consider the set $\mathbb{R}^\omega$ with the box topology, and show that it is not metrizable.*

When we turn our attention to the covering properties that involve finite subcovers, we see that in metric spaces, matters are simplified. In metric spaces, countable compactness and compactness are the same.

Theorem 10.17. *A metric space is compact if and only if it is countably compact.*

Recall that in a T_1 space, being countably compact is equivalent to every infinite subset having a limit point. Therefore, the previous theorem can be re-phrased as follows.

Theorem 10.18. *A metric space is compact if and only if every infinite subset of X has a limit point.*

In this first section of our exploration of metric spaces we simply marched through themes of topological spaces and saw how each one interacted with the metrizability. We considered subsets, separation properties, countability properties, products, and covering properties. Next we turn our attention to functions between metric spaces.

10.2 Continuous Functions Between Metric Spaces

For maps between metric spaces, the topological definition of continuity is equivalent to the ε-δ description of continuity that appeared in your calculus book. You'll want to use our earlier definition of continuous functions between topological spaces to prove this theorem.

Theorem 10.19. *A function f from a metric space (X, d_X) to a metric space (Y, d_Y) is continuous at the point x (in the topological sense) if and only if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $y \in X$, if $d_X(x, y) < \delta$, then $d_Y(f(x), f(y)) < \varepsilon$. The function f is continuous if and only if it is continuous at every point $x \in X$.*

Notice that in the characterization of continuity above, for a fixed $\varepsilon > 0$, we may be required to select different δ's at different points in the space. If we have a continuous function such that for each $\varepsilon > 0$ a fixed $\delta > 0$ works for every point in the space, then such a function has a stronger type of continuity called uniform continuity.

Definition. A function f from a metric space (X, d_X) to a metric space (Y, d_Y) is **uniformly continuous** if and only if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $x, y \in X$, if $d_X(x, y) < \delta$, then $d_Y(f(x), f(y)) < \varepsilon$.

Exercise 10.20. *Give an example of a continuous function from $\mathbb{R}^1$ to $\mathbb{R}^1$ with the standard topology that is not uniformly continuous.*

Theorem 10.21. *Let $f : X \to Y$ be a continuous function from a compact metric space to a metric space Y. Then f is uniformly continuous.*

Sometimes we create a sequence of continuous functions from one metric space to another with the goal of having those functions converge to a limiting continuous function. Converging pointwise is frequently not sufficient to guarantee convergence of the functions.

Exercise 10.22. *Find a sequence of continuous functions $f_i : [0, 1] \to [0, 1]$ ($i \in \mathbb{N}$) such that for each point $x \in [0, 1]$, the points $f_i(x)$ converge to a point p_x in $[0, 1]$ and yet the function $L : [0, 1] \to [0, 1]$ defined by $L(x) = p_x$ is not continuous.*

With examples like those you created in the previous exercise in mind, we define a type of convergence of a sequence of functions that requires more than merely pointwise convergence.

Definition. Let X be a topological space, and let (Y, d) be a metric space. A sequence of continuous functions $f_i : X \to Y$ **converges uniformly** if and only if for every $\epsilon > 0$, there is an integer M such that for every $x \in X$ and $m, n > M$, $d(f_m(x), f_n(x)) < \epsilon$.

Uniformly convergent continuous functions into a compact metric space will converge to a continuous function.

Theorem 10.23. *Let X be a topological space and let Y be a compact metric space. If a sequence of continuous functions $f_i : X \to Y$ converges uniformly, then $f : X \to Y$ defined by $f(x) = \lim f_i(x)$ for each $x \in X$ exists and is continuous.*

10.3 Lebesgue Number Theorem

The next theorem basically says that open covers of compact sets in a metric space can't just barely overlap. There must be some size larger than zero such that every point in the compact set is cushioned by at least that metric distance inside at least one of the elements of the open cover.

Lebesgue Number Theorem 10.24. *Let $\{U_\alpha\}_{\alpha\in\lambda}$ be an open cover of a compact set A in a metric space X. Then there exists a $\delta > 0$ such that for every point p in A, $B(p, \delta) \subset U_\alpha$ for some α. This number δ is called a* ***Lebesgue number*** *of the cover.*

One natural way to prove the Lebesgue Number Theorem is by contradiction. Among the proofs you discover, that approach should be one. However, we cannot resist suggesting two alternative methods as well. One method is to define a continuous function from A to the positive reals that basically measures the maximum size of an open ball around each point that lies in a single open set of the cover. Then remember that a continuous function of a compact set into the reals attains its minimum. Another approach is to realize that proving the theorem for open covers consisting of finitely many open balls is sufficient. Now consider replacing each open ball by an expanding union of slightly smaller open balls, thus replacing a finite cover by a related infinite cover. By compactness that new cover has a finite subcover that basically shrinks each ball in the original cover. Use that fact to get your Lebesgue number.

The Lebesgue Number Theorem has many applications, as we shall see in later chapters when we discuss paths in topological spaces and the fundamental group. The following exercise might help you imagine how it could be useful even when discussing covers of spaces that are not necessarily metric spaces.

Theorem 10.25. *Let $\gamma : [0,1] \to X$ be a **path**: a continuous map from $[0,1]$ into the space X. Given an open cover $\{U_\alpha\}$ of X, show that $[0,1]$ can be divided into N intervals of the form $I_i = [\frac{i-1}{N}, \frac{i}{N}]$ such that each $\gamma(I_i)$ lies completely in one set of the cover.*

10.4 Complete Spaces

In compact spaces, every infinite set has a limit point. Although not every infinite subset of the real line with the standard topology has a limit point, Cauchy sequences (defined below) *do* converge, which observation gives rise to the concept of a complete metric and a complete metric space.

Definition.

1. Let X be a metric space with metric d. A sequence $\{x_i\}_{i\in\mathbb{N}}$ of points in X is a **Cauchy sequence** if and only if for each $\epsilon > 0$, there is an integer M such that for all $m, n > M$, $d(x_m, x_n) < \epsilon$.
2. Let d be a metric for the metric space X. Then d is a **complete metric** for X if and only if every d-Cauchy sequence in X converges.
3. A metric space (X, d) is **complete** if and only d is a complete metric for X.
4. A topological space $(X, \mathcal{T})$ is **topologically complete** if and only if there is a complete metric d on X that generates the topology of X.

A primary example of a complete metric space is $\mathbb{R}^n$ with the usual metric.

Exercise 10.26.

1. *The space $\mathbb{R}^n$ is complete.*
2. *There is a metric that generates the standard topology on $\mathbb{R}^1$ that is not a complete metric.*

You saw in the exercise above that some metrics that generate the standard topology on $\mathbb{R}$ are complete and some metrics that generate that same topology are not complete. For compact metric spaces, the situation is different. All metrics on compact spaces are complete.

Theorem 10.27. *Let X be a compact metric space. Then every metric for X is a complete metric.*

One of the basic theorems about complete metric spaces can be described in two different ways–one way is to talk about the intersection of dense open sets and the second is to talk about the union of nowhere dense sets. Each one of these two equivalent statements is called the Baire Category Theorem.

Theorem 10.28 (The Baire Category Theorem)**.** *Let X be a complete metric space and $\{U_i\}_{i\in\mathbb{N}}$ be a collection of dense open sets. Then $\bigcap_{i\in\mathbb{N}} U_i$ is a dense set.*

Definition. A subset Y of a space X is **nowhere dense** if and only if $Int(Cl(Y)) = \varnothing$.

Theorem 10.29 (The Baire Category Theorem)**.** *Let X be a complete metric space. Then X is not the union of countably many nowhere dense sets.*

Effective Thinking Principle. *Find the Essence.* Seek the essential ingredients in a proof.

We saw that uniformly convergent continuous functions into a compact metric space converge to a continuous function. Only the completeness of the target space was necessary to draw that conclusion.

Theorem 10.30. *Let X be a topological space and let Y be a complete metric space. If a sequence of continuous functions $f_i : X \to Y$ converges uniformly, then $f : X \to Y$ defined by $f(x) = \lim f_i(x)$ for each $x \in X$ exists and is continuous.*

After defining a property, it is natural to ask which related spaces also have this property. Let's pursue that strategy for complete metric spaces.

Theorem 10.31. *If X and Y are complete metric spaces, then $X \times Y$ is complete.*

When we investigate subsets, we find some interesting results.

Theorem 10.32. *Every closed subset of a complete metric space is complete.*

Consider $(0, \infty)$, an open subset of $\mathbb{R}$. Notice that even though $\mathbb{R}$ is a complete metric space, $(0, \infty)$ is not complete under the standard metric, because the Cauchy sequence $1/2, 1/3, 1/4, ...$ does not converge to a point in this space. However, there is a complete metric for $(0, \infty)$ that generates the subspace topology on $(0, \infty)$, for instance,

$$d(a, b) = \sqrt{(b-a)^2 + (\frac{1}{b} - \frac{1}{a})^2}.$$

This metric is obtained by embedding $(0, \infty)$ as the graph of $f(x) = 1/x$ on $(0, \infty)$ in $\mathbb{R}^2$, which is a closed subset of $\mathbb{R}^2$, and using the standard Euclidean metric on $\mathbb{R}^2$. Thus $(0, \infty)$ is topologically complete using a metric that is not the same as the complete metric it inherited from $\mathbb{R}$. The next theorem generalizes this idea.

Theorem 10.33. *Let U be an open subset of a complete metric space X. Then U is topologically complete, that is, there is a complete metric on U that generates the relative topology of U.*

Hint: Just as with the example of $(0, \infty) \subset \mathbb{R}$, embed U as a closed subset of $X \times \mathbb{R}$.

The countable product of complete metric spaces admits a complete metric.

Theorem 10.34. *If $\{X_i\}_{i \in \mathbb{N}}$ is a collection of complete metric spaces, then $\prod_{i \in \mathbb{N}} X_i$ is complete.*

The preceding ideas may help if you are in the mood for the following two challenging theorems.

Theorem 10.35. *Let $\{U_i\}_{i \in \mathbb{N}}$ be a countable collection of open sets in a complete space X. Then $Y = \bigcap_{i \in \mathbb{N}} U_i$ is complete.*

Hint: Try to embed Y as a closed subset of the product of X with a countable number of copies of $\mathbb{R}$.

The converse of this theorem is also true, so we can present a characterization of which subsets of complete spaces are complete.

Theorem 10.36. *Let X be a complete space. Then $Y \subset X$ is complete if and only if there exists a countable collection of open sets $\{U_i\}_{i \in \mathbb{N}}$ such that $Y = \bigcap_{i \in \mathbb{N}} U_i$.*

10.5 Metric Continua

Definition. A **metric continuum** is a metric space that is also a continuum, that is, a metric continuum is a connected, compact metric space.

Definition. A **Peano Continuum** is a locally connected metric continuum.

Theorem 10.37. *A Hausdorff space X is a Peano Continuum if and only if X is the image of $[0, 1]$ under a continuous, surjective function.*

Theorem 10.38. *A Peano Continuum is path connected and locally path connected.*

Theorem 10.39. *An open, connected subset of a Peano Continuum is path connected.*

One of the goals of topology is to describe the essential features of familiar spaces that characterize those spaces. You may recall that every non-degenerate continuum has at least two non-separating points. The following theorem shows us that the unit interval is the only metric continuum with exactly two non-separating points.

Theorem 10.40. *Let X be a metric continuum with exactly two non-separating points. Then X is homeomorphic to $[0, 1]$.*

The next theorem characterizes the circle $\mathbb{S}^1$ as the only metric continuum where no point separates it, but every pair of points separates it.

Theorem 10.41. *Let X be a non-degenerate metric continuum where no point separates X but every pair of points separates X. Then X is homeomorphic to $\mathbb{S}^1$.*

The following theorem is known as the Kline Sphere Characterization Theorem. Don't work too hard on it, because it is very difficult.

Theorem 10.42. *Let X be a metric continuum with more than one point where no pair of points separates X, but every subset of X homeomorphic to $\mathbb{S}^1$ separates X. Then X is homeomorphic to $\mathbb{S}^2$.*

10.6 Metrizability

We saw above that a metric on a set gives rise to a topology. Then we saw some consequences of knowing that the topology on a space is generated by a metric. Now we ask a sort of reverse question: if $(X, \mathcal{T})$ is a topological space, when is it possible to find a metric on the set X such that the metric space topology is the same as the topology X already has?

Put more succinctly, when is a topological space metrizable? To show a space is not metrizable, we might show that one of the properties that every metric space enjoys does not hold for that space. At this point in our discussion, the only way to show that a space is metrizable is to produce a metric that generates the topology.

Exercise 10.43.

1. *Is the space $\mathbb{R}$ with the discrete topology metrizable?*
2. *Is the space $\mathbb{R}_{LL}$ metrizable?*

Since being a metric space is both a category of spaces and being metrizable is a property that a topological space may or not possess, you can now add a new row and a new column to your chart of properties and examples.

Exercise 10.44. *Take your chart of examples and properties and add metric space as an example and add metrizable as a property and fill in the chart.*

One of the basic questions about metrizability is to characterize metrizable spaces in terms of the topological features they have. Urysohn's Lemma is called a lemma, because it first appeared as a lemma to the following metrization theorem. The challenge to proving this theorem, or any metrization theorem, is to use the hypothesized topological features to define a metric on the space, that is, a distance between points, in such a way that the metric induces the given topology.

Urysohn's Metrization Theorem 10.45. *Every regular, T_1, 2^{nd} countable space is metrizable.*

Hint: First recall that we proved earlier that such a space is normal. Now find a way to use Urysohn's Lemma to embed the space in $\mathbb{R}^\omega$, which we know to be metrizable.

Urysohn's Metrization Theorem implies some facts about compact spaces.

Theorem 10.46. *Let X be a compact Hausdorff space that is 2^{nd} countable. Then X is metrizable.*

And putting together some insights about continuous functions, we can conclude that the images of compact metric spaces must be metrizable if the image is Hausdorff.

Theorem 10.47. *Let X be a compact metric space, Y be a Hausdorff space, and $f : X \to Y$ be a continuous, surjective function. Then Y is a compact metric space.*

One way to characterize compact metric spaces is as the continuous image of the Cantor set.

Theorem 10.48. *Let X be a Hausdorff space, and let C be the standard Cantor set. Then X is a compact metric space if and only if there exists a continuous surjective function $f : C \to X$.*

Urysohn's Metrization Theorem gives a good characterization of the metrizability of 2^{nd} countable spaces. You may recall that, using Urysohn's Lemma, every normal, T_1 topological space can be embedded in a product of $[0,1]$'s. The following theorem has a similar flavor of embedding a space in a product of intervals. In this case the space is a separable metric space and the product is a countable product.

Theorem 10.49. *Every separable metric space can be embedded in a countable product of $[0,1]$'s.*

If a metric space is not separable, then it cannot be embedded in a countable product of intervals. However, in the next section we will see that any metric space *can* be embedded in a countable product of hedgehogs!

10.7 Advanced Metrization Theorems

Metric spaces do not necessarily have a countable basis; however, they have bases that are the unions of countably many collections of appealing sets. The theorems in this section describe assumptions on bases for a topological space that imply that there exists a metric on the space that generates its topology. These metrization theorems were proved independently in the early 1950's by Bing, Nagata, and Smirnov. These theorems are similar, so they are often collectively referred to as the Bing-Nagata-Smirnov Metrization Theorem. Since their statements involve some technical terms, we explore a few ideas before stating the theorems in full.

We begin by defining a metric space called a *hedgehog*–named in reference to the many spines that stick out of that bristly beast.

Definition. Let $X = \bigcup_{\alpha \in \lambda} I_\alpha$ be the space created by taking a potentially uncountable disjoint union of unit intervals with the standard topology. Let H be the quotient space obtained from X by identifying all the 0 endpoints in all the intervals to a single point. The space H is called a **hedgehog.**

Exercise 10.50. *Show that a hedgehog is a metric space where the distance between two points can be described as taking the distance from one point to the* 0 *on its spine and then adding the distance out to the second point on the other point's spine.*

Theorem 10.51. *The countable product of hedgehogs is metrizable.*

One method to prove one direction of the Bing-Nagata-Smirnov Metrization Theorem below is to show that the hypotheses allow us to embed the space into a countable product of hedgehogs in a manner similar to how normal, T_1 spaces with a countable basis can be embedded in a countable product of intervals. Indeed our hypothesis about the basis is designed to allow us to produce such an embedding.

Definition.

1. A collection of subsets $\{E_\alpha\}_{\alpha\in\lambda}$ of a topological space X is a **discrete collection** if and only if for every point $x \in X$, there exists an open set U with $x \in U \subset X$ such that U intersects at most one E_α.

2. A basis $\mathscr{B}$ of a space X is a σ**-discrete basis** if and only if $\mathscr{B} = \bigcup_{i\in\mathbb{N}} B_i$ where each $\mathscr{B}_i = \{B_{i,\alpha}\}_{\alpha\in\lambda_i}$ is a discrete collection of open sets.

This first theorem uses the Normality Lemma to establish the normality of a regular, T_1 space with a σ-discrete basis.

Theorem 10.52. *A regular space with a σ-discrete basis is normal. In fact, given a discrete collection of closed sets $\{C_\alpha\}_{\alpha\in\lambda}$, there exists a discrete collection of open sets $\{U_\alpha\}_{\alpha\in\lambda}$ such that for each α, $C_\alpha \subset U_\alpha$.*

R.H. Bing proved his Metrization Theorem in 1951. It is an 'if and only if' statement. Both directions are difficult. The following theorem is one of those directions. Try to prove that the existence of a σ-discrete basis implies that the space can be embedded in a countable product of hedgehogs. We will give some guidance about how you might approach the other direction later.

Theorem 10.53. *A regular, T_1 space X with a σ-discrete basis is metrizable.*

To show that a metric space X has a σ-discrete basis, a good strategy is to fix a natural number n and to consider the open cover of X by $\frac{1}{n}$-balls and show that we can produce a countable number of discrete collections of open sets each of which is a refinement of the cover and such that the union of all the sets in all of the countable collection of sets covers X. Repeating this process for each natural number n generates our desired σ-discrete basis. This approach has an intermediate step of producing a countable number of collections of discrete closed sets that are refinements and whose union is the whole space.

To deal with the general case of any metric space, the cover of X by $\frac{1}{n}$-balls could well have an uncountable number of open sets in the cover. Let's learn a useful technique by considering a simpler case where we start with a countable open cover. So here is a practice lemma.

Lemma 10.54. *Let $\{U_i\}_{i\in\mathbb{N}}$ be a countable open cover of a metric space X. For each point $x \in X$ let $m(x)$ be the natural number such that $x \in U_i$ but $x \notin U_j$ for $j < i$. Then for every $n \in \mathbb{N}$ there exists a discrete collection of closed sets $\{C_{i,n}\}$ such that*

1. *for each i, $C_{i,n} \subset U_i$;*
2. *for each $x \in C_{i,n}$, $B(x, \frac{1}{n}) \subset U_i$;*
3. *for each i, $C_{i,n}$ does not intersect U_j for $j < i$; and*
4. *for each i, $C_{i,n}$ contains every point $x \in U_i$ for which $m(x) = i$ and for which $d(x, X - U_i) > \frac{1}{n}$.*

Then $\bigcup C_{i,n} = X$ and for each n, the collection of $\frac{1}{3n}$ neighborhoods of the $C_{i,n}$'s, that is,

$$\left\{ \bigcup_{x\in C_{i,n}} B(x, \frac{1}{3n}) \right\}_{i\in\mathbb{N}},$$

is a discrete collection of open sets.

The above lemma demonstrates a construction that also works for collections that are not countable. To deal with uncountable covers, we can use well-ordering. Recall that a set is well-ordered if and only it is totally ordered and every non-empty subset has a least element. The proof of the following lemma is basically the same as the proof of the lemma above. Notice that the function $m(x)$ in the statement below relies on the index set being well-ordered.

Lemma 10.55. *Let $\{U_\alpha\}_{\alpha\in\lambda}$ be an open cover of a metric space X where the index set λ is well-ordered. For each point $x \in X$ let $m(x)$ be the ordinal number α such that $x \in U_\alpha$ but $x \notin U_\beta$ for $\beta < \alpha$. Then for every $n \in \mathbb{N}$ there exists a discrete collection of closed sets $\{C_{\alpha,n}\}$ such that*

1. *for each α, $C_{\alpha,n} \subset U_\alpha$;*
2. *for each $x \in C_{\alpha,n}$, $B(x, \frac{1}{n}) \subset U_\alpha$;*

3. *for each α, $C_{\alpha,n}$ does not intersect U_β for $\beta < \alpha$; and*

4. *for each α, $C_{\alpha,n}$ contains every point $x \in U_\alpha$ for which $m(x) = \alpha$ and for which $d(x, X - U_\alpha) > \frac{1}{n}$.*

Then $\bigcup C_{\alpha,n} = X$ and for each n, the collection of $\frac{1}{3n}$ neighborhoods of the $C_{\alpha,n}$'s, that is,

$$\left\{ \bigcup_{x \in C_{\alpha,n}} B(x, \frac{1}{3n}) \right\}_{\alpha \in \lambda},$$

is a discrete collection of open sets.

You are now in a position where you can prove the other direction of Bing's Metrization Theorem, namely, proving that a metric space has a σ-discrete basis, thereby completing the proof of Bing's Metrization Theorem.

The Bing Metrization Theorem 10.56. *A regular, T_1 space X is metrizable if and only if X has a σ-discrete basis.*

The similar metrization theorem by Nagata and Smirnov characterizes a metric space as having a σ-locally finite basis rather than a σ-discrete basis. We will recall the definition of a collection of subsets of a space being locally finite, and define what it means for a collection to be σ-locally finite.

Definition.

1. A collection of subsets $\{E_\alpha\}_{\alpha\in\lambda}$ of a topological space X is a **locally finite collection** if and only if for every point $x \in X$, there exists an open set U with $x \in U \subset X$ such that U intersects at most a finite number of the E_α.

2. A basis $\mathscr{B}$ of a space X is a **σ-locally finite** if and only if $\mathscr{B} = \bigcup_{i\in\mathbb{N}} B_i$ where each $\mathscr{B}_i = \{B_{i,\alpha}\}_{\alpha\in\lambda_i}$ is a locally finite collection of open sets.

Once again, we begin by asserting that a regular space with a σ-locally finite basis is normal.

Theorem 10.57. *A regular space with a σ-locally finite basis is normal. In fact, given a discrete collection of closed sets $\{C_\alpha\}_{\alpha\in\lambda}$, there exists a discrete collection of open sets $\{U_\alpha\}_{\alpha\in\lambda}$ such that for each α, $C_\alpha \subset U_\alpha$.*

The Nagata-Smirnov Metrization Theorem 10.58. *A regular, T_1 space X is metrizable if and only if X has a σ-locally finite basis.*

Since a σ-discrete basis is a σ-locally finite basis, Bing's Metrization Theorem proves one direction of this theorem already, namely that a metric space has a σ-locally finite basis. So the only remaining challenge is to prove that a normal, T_1 space X with a σ-locally finite basis has a σ-discrete basis. Several steps are useful. The first follows the strategy of the lemma above.

Lemma 10.59. *Let X be a space with a σ-locally finite basis $\{\{B_{\alpha,n}\}_{\alpha\in\lambda_i}\}_{n\in\mathbb{N}}$. Let $\{U_\alpha\}_{\alpha\in\lambda}$ be a locally finite collection of open sets in a space X where the index set λ is well-ordered. (In the application, this collection of U_α's will be one of the locally finite collections of basis elements.) For each point $x \in \bigcup_{\alpha\in\lambda} U_\alpha$ let $m(x)$ be the ordinal number α such that $x \in U_\alpha$ but $x \notin U_\beta$ for $\beta < \alpha$. Then for every $n \in N$ there exists a discrete collection of closed sets $\{C_{\alpha,n}\}$ such that*

1. *for each α, $C_{\alpha,n} \subset U_\alpha$;*
2. *for each α, $C_{\alpha,n}$ does not intersect U_β for $\beta < \alpha$; and*
3. *for each α, $C_{\alpha,n}$ contains every point $x \in U_\alpha$ for which $m(x) = \alpha$ and for which $x \in \{B_{\alpha,n}\}$.*

Then

1. *for each n, $\{C_{\alpha,n}\}_{\alpha\in\lambda}$ is a discrete collection of closed sets;*
2. *$\bigcup_{\alpha\in\lambda;n\in\mathbb{N}} C_{\alpha,n} = \bigcup_{\alpha\in\lambda} U_\alpha$; and*
3. *for each n, there exists a discrete collection of open sets $\{V_{\alpha,n}\}_{\alpha\in\lambda}$ such that for each $\alpha \in \lambda$, $C_{\alpha,n} \subset V_{\alpha,n} \subset \overline{V_{\alpha,n}} \subset U_\alpha$.*

This lemma can be used to prove that a normal, T_1 space with a σ-locally finite basis has a σ-discrete basis. From this fact, the Nagata-Smirnov Metrization Theorem follows from Bing's Metrization Theorem above. As an historical note, originally Nagata and Smirnov independently constructed proofs of their theorem without going through Bing's Theorem.

10.8 Paracompactness of Metric Spaces

Recall that a space is paracompact if and only if it is Hausdorff and every open cover of X has a locally finite refinement. The challenge of proving paracompactness of a space is to keep in mind that there are three open covers involved: the given open cover, the locally finite refinement, and the set of open sets that demonstrate that the refinement is locally finite. So the key to proving paracompactness is often to produce the collection of sets that demonstrate the local finiteness at the same time you are producing the refinement itself.

The fact that every metric space has a σ-discrete basis makes proving the paracompactness of a general metric space quite similar to proving the paracompactness of a 2nd countable space. So let's begin by dealing with a 2nd countable space. This lemma statement is really designed to outline an approach to constructing a locally finite refinement. You may have developed these ideas back in Section 7.5.

Lemma 10.60. *Let $\{B_i\}_{i\in\mathbb{N}}$ be a countable basis of a regular space X. Let $\{U_\alpha\}_{\alpha\in\lambda}$ be an open cover of X. Let $\{C_i\}_{i\in\mathbb{N}}$ be the set of all B_i's such that each C_i lies in some U_α in the open cover. Then $\{C_i\}_{i\in\mathbb{N}}$ is an open refinement of the open cover $\{U_\alpha\}_{\alpha\in\lambda}$; however, it is not locally finite. Let $\{D_i\}_{i\in\mathbb{N}}$ be the set of all B_i's such that each $\overline{D_i}$ is a subset of some C_k. For each $i \in \mathbb{N}$ let $E_i = C_i - \bigcup\{\overline{D_j} | j < i$ and $\overline{D_j} \subset C_k$ for some $k < i\}$. Then $\{E_i\}_{i\in\mathbb{N}}$ is a locally finite refinement of $\{U_\alpha\}_{\alpha\in\lambda}$.*

Now you are ready to use the fact that metric spaces have a σ-discrete basis to prove that metric spaces are paracompact.

Theorem 10.61. *Metric spaces are paracompact.*

10.9 Going the Distance

The concept of a metric in a sense closed the circle on our exploration of point-set topology. We set out to take familiar spaces like $\mathbb{R}$ and describe essential features just using ideas about points and sets. In this chapter, we returned to the idea of distance that seems so basic to the concept of $\mathbb{R}$, and created the idea of a general metric space.

The impulse that guided most of our exploration of metric spaces was to take all the constructions and properties that we had developed for topological spaces in general and to investigate how those constructions and properties related to metric spaces. We then sought to understand which topological properties were the essential ingredients in deciding whether a given topological space actually could be generated by a metric.

Sadly, some people regard metric spaces as the only interesting topological spaces. Those people miss out on many joys that you experienced over the first chapters of this book. Please feel sorry for them—it's probably not their fault.

Part II

Algebraic and Geometric Topology

Chapter 11

Transition From Point-Set Topology to Algebraic and Geometric Topology: Similar Strategies, Different Domains

We view this book as partly about mathematics and partly about the practices of mind that lead us to create mathematics. The book is divided into two parts. The introduction to point-set topology in the preceding chapters illustrated the effectiveness of various ways of exploring the unknown.

One of the intriguing questions that you can ask about mathematics is whether mathematics is discovered or created. Perhaps the experience of developing point-set topology can give some nuance to that question. In some sense, every step of the process of creating concepts was laid bare. One could argue that the strategies that were employed to create each idea were rather straightforward and the results of pursuing the ideas that emerged were quite inevitable from the process.

That is, we started the whole exploration by taking concepts that were already familiar to us and then we set out to find the most fundamental essence that made the familiar concepts work. We thought about the real line and continuous functions and then we embarked on a journey of abstraction that included our investigating basic notions about sets. Thinking of size led us to develop the concept of cardinality. Thinking about essential requirements for capturing ideas of closeness in the real line from a set theoretic point of view rather than a distance point of view led us to create the idea of a topology.

Once we had isolated the idea of a topological space, many investigations naturally followed. The impulse to look at elemental features encouraged us to create the idea of a basis. The strategy of taking an idea and applying it to related objects led us to the notions of subspaces and product spaces.

We asked ourselves what features of spaces distinguish one topological space from another.

These questions led to a whole range of concepts such as the separation and covering properties. And then our impulse to consider size led us to describe ideas such as first and second countability.

One of the basic strategies for creating ideas in mathematics is to consider relationships between objects that we have created. That strategy led to the idea of continuity. Looking at intuitive concepts such as being connected led us to capture that notion with several variations of connectedness in topological spaces. Finally, we looked at the familiar idea of distance in the real numbers and circled back to see how the topological concepts we created are related to the intuitive idea of closeness created by distances. That led to the exploration of metric spaces.

11.1 Effective Thinking Principles—Strategies for Creating Concepts

The entirety of the mathematical developments we have seen so far are illustrations of the process of doing and creating mathematics. Along the way we recorded instances of these practices of discovery and exploration and listed them as Effective Thinking Principles. These methods of effective thinking and creation of ideas served us well in the first part of this book and will be employed again in the second part.

Here we gather and organize some of the themes captured in the Effective Thinking Principles to help us reflect on practices of mind that led to the mathematical creations in the first part of this book and that will lead us onto new ideas as we turn our attention to the algebraic and geometric themes of the second part of this book.

Principles of Effective Thinking—How to Create Ideas

1. **Find the Essence**. Seek essential ingredients in a proof or concept. Often isolating essential features opens up new worlds of insight.

2. **Start With Simple Cases**. George Polya said, "If there is a hard problem you can't do, there is an easier problem you can't do. Find it."

3. **Create Examples**.

   (a) When learning any definition or concept in mathematics, one helpful step is to construct several examples that illustrate the meaning of the definition or concept. Specifically, find or create examples that manifest the differences among related concepts or definitions.

   (b) To understand theorems more deeply, look at or create examples that reveal the implications of the theorem in specific cases.

(c) Explore Extremes. After making a definition, explore extreme or unusual cases.

4. **Draw a Picture**.
   Draw a picture. Draw a picture!! DRAW A PICTURE!!! It is impossible to overemphasize the value of drawing a picture.

   Part of the value of drawing a picture is to see features in your picture that you did not intentionally put there. For example, if you draw a triangle, the sum of the angles will equal 180 degrees. When you drew the triangle, you did not make sure it had 180 degrees. A picture can suggest insights that will help with your argument. Get in the habit of drawing pictures!

5. **Extend Insights and Theorems**.
   The best source of new ideas is old ideas. One of the most effective strategies for helping to understand and create mathematics is to look systematically for the extent to which an insight or a theorem can be extended or not.

   (a) Explore Limits of Theorems. When you discover a theorem, explore its limitations and possible extensions by systematically weakening the hypotheses and checking to see whether the theorem is still true and strengthening the conclusion to see whether you can deduce more than you originally thought. In each case, create examples that demonstrate the necessity of hypotheses and the limits of conclusions.

   (b) Add Hypotheses—Strengthen Conclusions. Given a theorem, see if strengthening the hypotheses in various ways allows you to draw stronger conclusions.

   (c) Weaken Hypotheses and See What Can Still Be Deduced. The hypotheses of a theorem may all be necessary to draw the stated conclusion; however, perhaps an interesting weaker conclusion can be drawn with weakened hypotheses.

   (d) Identify Essence of Hypotheses. To understand theorems better and to improve them if possible, identify exactly what aspects of the hypotheses were actually used in the proof.

   (e) Consider Analogies of Previous Results. After developing variations of previous concepts, look at previous results and see what analogous results hold.

   (f) Explore How New Ideas Relate to Old Ideas. A standard and fruitful method for creating and learning mathematics is to explore how a new concept interacts with previous concepts.

6. **Understand Proofs, Not Just Statements, of Theorems**. Understanding the proofs of theorems rather than just the statements of theorems gives you power.

These are just a few strategies for learning and creating ideas that we encountered so far. We hope that one outcome of your experience with this book is that you come to see that solving problems, proving theorems, and creating insights is more method than magic. Employing systematic methods of inquiry reliably leads to new ideas.

11.2 Onward: to Algebraic and Geometric Topology

In the first part of this book, we pursued the set-theoretic essence of topological spaces and continuous functions between them. We saw how we can distinguish among topological spaces by seeing whether they have different topological properties, such as separation properties.

However, we live in what appears to be a Euclidean space. So it is natural for us to want to distinguish differences among objects that we see around us—objects such as a sphere and a torus. The question becomes: how would we capture the difference between such objects? These spaces both have the nicest general topological features, such as having a countable basis and having all the separation properties we could desire. So when we seek to make distinctions among such objects, we must develop a new collection of techniques that allow us to distinguish one object from another.

The second part of this book leads us to discover techniques by which we will be able to tell the topological difference between a torus and a Klein bottle, or the difference between Euclidean spaces of dimension 15 and dimension 16.

You might wonder why we need such techniques—after all, maybe you intuitively understand why these spaces are different. However, as you have seen and shall continue to see, the truth sometimes surprises us. Things we thought were different turn out to be topologically identical—and vice versa—some pairs of spaces may seem the same, but turn out to be subtly different.

Another motivation for developing more nuanced methods of distinguishing spaces arises when spaces may be presented to us in a manner that challenges our ability to understand them totally–for example, our universe. We may imagine that every point in the universe has a neighborhood much like the neighborhoods we inhabit, but even if that were true, we would still be left with the challenge of describing the global structure of the universe. For instance, suppose someone handed you a map of a space broken up into pieces, much like an atlas of the globe might present us with a lot of maps of small parts of the world without any one map showing the whole. We might be instructed about how the local maps fit together piece by piece, but we might not be told what the overall space looks like. For example, suppose our atlas were an atlas of parts of a torus rather than parts of the surface of the Earth. Could we use the combinatorial information about overlapping patches to understand the overall space in some way?

Or suppose someone handed us a graph of dots and edges describing some complex network of relationships. Can we use topology to try to understand the "shape" of that network? And much recent work in data science has been devoted to understanding "the shape of data" using methods from algebraic topology: by putting balls of increasing radii around the data points and watching how the homology of the resulting space changes over time.

The idea of viewing spaces as being made up of a finite number of simple building blocks is a big theme in topology, and this reduction to a finite number of pieces often leads us to categorize the topology we will explore with an adjective: *combinatorial* or *algebraic* or *geometric*, depending on what techniques are brought to the fore. For instance, in algebraic topology, we develop techniques for associating algebraic objects, such as groups, with topological spaces. In combinatorial topology, we consider how combinatorial properties of a space that is broken into pieces shed light on its topology. And geometric topology emphasizes the detailed visual characteristics of the objects we are studying.

11.3 Manifolds and Complexes: Building Locally, Studying Globally

In this section, we will not formally define anything, but instead will give you an intuitive sense of some of the spaces that we will study in the second part of this book.

Euclidean spaces are abstractions and extensions of geometric features that arise from our common experience of the world. When we look at a tabletop, we can imagine a perfectly smooth surface that extends forever in every direction. That abstraction is the Euclidean plane. If we imagine the space in a room extended indefinitely in all directions, we are envisioning Euclidean 3-space. Higher dimensional Euclidean spaces are natural generalizations and extensions of those common spaces. Since we appear to inhabit these Euclidean spaces, we naturally raise questions about them or spaces inside them or spaces related to them.

Spaces that are locally Euclidean are probably the most frequently studied spaces in all of topology. A space that is locally the same as Euclidean space is called a *manifold*. Manifolds are extremely important in many branches of mathematics as well as in many sciences. Here are a couple of examples: A 2-sphere, $\mathbb{S}^2$, is the surface of a ball in 3-space. Every point on a 2-sphere has an open set around it that is homeomorphic to an open disk in the plane. Likewise, a torus, the boundary of a doughnut, has the property that every point has an open set around it homeomorphic to an open disk in the plane. So a 2-sphere and a torus are examples of 2-manifolds.

We are comfortable with the idea of a 3-manifold since we seem to live in one. Indeed, naively speaking, our universe appears to be a 3-manifold (or a 4-manifold if one wants to consider time); that is, we can imagine that every point in the universe locally looks like a room. However, we don't know whether the totality of the universe might look more like an abstraction of a tabletop or a sphere or a 3-dimensional torus or something else. Perhaps we will be able to deduce the global topological type of the universe as we gain knowledge in cosmology and mathematics.

One of the big impulses in studying manifolds is to attempt to classify them, that is, to present an organized list of manifolds of a given dimension that identifies every manifold of that dimension as a specific member of that list based on recognizable criteria. We will classify 2-manifolds in that way in the next chapter. The success and clarity with which compact, connected 2-manifolds are classified is, unfortunately, not available in higher dimensions.

Another class of spaces we could think about studying are the ones created by assembling simple pieces from Euclidean geometry to create almost any object you can imagine. The simple pieces are called simplexes or simplices. A single point is a 0-simplex; a line segment is a 1-simplex; a triangle is a 2-simplex; a tetrahedron is a 3-simplex; and the pattern continues. A simplex is a rectilinear subset of $\mathbb{R}^n$ that can be described in a simple way using linear algebra. These simplices can be put together to create complexes. Finite complexes are simply finite unions of simplices in $\mathbb{R}^n$.

The advantage to working with finite complexes is that they are made of a finite number of simple pieces that fit together neatly. So they can often be analyzed using combinatorial methods such as counting vertices and edges and so on, or by using inductive methods.

11.4 The Homeomorphism Problem

Perhaps the most basic question in topology is, "Given two topological spaces, how can we tell whether or not they are homeomorphic?" For instance, if you look at a square with parallel edges glued pairwise in the same direction is that the same as a hexagon with parallel edges glued pairwise in the same direction? (Surprisingly they are.) If you take a 2-dimensional sphere and cross it with an interval, do you get a 3-dimensional sphere? (You do not.) In the chapters to come, we will develop tools that help to distinguish one space from another.

A property of a space that is preserved under homeomorphisms is called an **invariant**. For example, normality is an invariant: if one space is normal and another is not normal, then the spaces are definitely not homeomorphic. One of the fundamental strategies for proving that two spaces are **not** homeomorphic is to identify an invariant on which the two spaces differ. The problem with the invariant properties that were introduced in the first part of this book is that they tend not to be refined enough to distinguish spaces that are nice subsets of Euclidean spaces, such as a torus versus a sphere.

So we seek distinguishing characteristics of spaces that capture various appealing geometrical differences, such as holes. Intuitively, we have a sense of what a hole is. It appears that a torus has a different number of holes than a sort of double doughnut has. The problem is that we need a definition of 'hole' such that the number or type of 'holes' becomes an invariant under homeomorphisms.

Several different strategies arise for capturing the intuition of 'holes' and their analogues in higher dimensions. Some strategies involve associating a group or groups with a space, where the group becomes more complicated depending on the how holey the space is.

Developing these tools allows us to answer many appealing questions and to prove some of the most satisfying, fundamental theorems in topology. For example, in the next chapter we will classify all compact, connected 2-manifolds in such a way that we can look at such an object, compute a couple of invariants, and determine exactly what 2-manifold we are looking at up to homeomorphism.

Another category of theorem we will prove is fixed point theorems. If you take any continuous function of a ball to itself, there will be some point that gets mapped to itself. A similar theorem states that if you squash a beach ball on the ground, then two points that started as antipodal will be squashed onto the same point. This type of theorem can be proved by employing the invariants that we will use to distinguish spaces and investigating how they interact with continuous functions.

Another type of theorem that we will prove are theorems about geometric separation. It is intuitively obvious that any embedded simple closed curve in the plane separates the plane into two pieces; however, that theorem, known as the Jordan Curve Theorem, is surprisingly challenging to prove. However, the tools we create will allow us to prove not only that theorem but also analogous theorems in all dimensions.

11.5 Same Strategies, Different Flavors

The first part of this book explored set-theoretic topology. That exploration was driven by strategies of thinking that led to the creation of new ideas and increasingly refined insights. Those same strategies of concept creation can equally well be turned toward challenges of understanding the topology of geometric objects that we see around us everyday or abstractions of them.

The chapters ahead have a different flavor to them compared with the flavor of the ideas presented in the first part of this book. Now you will be asked to use geometric insights with confidence.

Here is an example that has no special significance in itself; it is intended to illustrate the kind of geometric thinking we will encourage in the chapters ahead, a type of thinking that will force you to think very concretely about simple objects (namely triangles) in $\mathbb{R}^n$.

Suppose you have a triangle σ_1 in $\mathbb{R}^n$ and you connect each vertex of σ_1 to the center of the opposite side. Then you will divide σ_1 into exactly six subtriangles. If you did that same process for each of those six subtriangles, you would divide σ_1 into exactly 36 subtriangles. Suppose now you have two triangles, σ_1 and σ_2 in $\mathbb{R}^n$ that share an edge, but are otherwise disjoint. Suppose you now divide each of σ_1 and σ_2 in to 36 subtriangles as described before. Then the center point of the common edge is a vertex of exactly 8 triangles whose union is homeomorphic to a disk.

The reasoning involved in the above example has a concrete, geometrical flavor, in contrast to the far more abstract reasoning about sets that is involved in the set-theoretical theorems in the first part of this book. In addition to this concrete, geometrical reasoning, the second part of the book will also include making connections between topological spaces and algebraic entities such as groups. That algebraic reasoning is yet another flavor of mathematical analysis. We hope you embrace the variety of methods of thinking that lie ahead.

You will soon be proving some of the most famous theorems in topology—classification theorems, fixed point theorems, and theorems describing the geometry of our world and its beautiful abstractions and extensions.

Chapter 12

Classification of 2-Manifolds: Organizing Surfaces

One of the disparaging insults you can lob at someone who does not know topology is that they *can* tell the difference between a coffee cup and a donut. But we topologists know they are the same. The fact that a donut (if sufficiently elastic) can be stretched and deformed until it looks like a coffee cup is an insight that some people do not instantly see. But knowing the material in this chapter would help them see why their boundary surfaces are homeomorphic.

Definition. A topological space is a **surface** or a 2**-manifold** if and only if it is a separable metrizable space where every point has a neighborhood homeomorphic to an open disk in $\mathbb{R}^2$.

Thus a surface is locally 2-dimensional. A basic question of this chapter is a classification question. The first step we take is to focus our attention on a sub-collection of 2-manifolds, namely, those that are compact and connected. The classification question is the following challenge: Can we create a list that contains all the possible compact, connected surfaces? How can we tell when two such surfaces are the same or different topologically? In this chapter, we will successfully discover a satisfying way to categorize all compact, connected, 2-manifolds.

12.1 Examples of 2-Manifolds

Effective Thinking Principle. *Look at Examples.* Looking at specific examples helps to develop intuition and understanding.

Before classifying them, it will be instructive to look at some examples of 2-manifolds.

The 2-sphere is the simplest example of a compact 2-manifold. A 2-sphere, denoted $\mathbb{S}^2$, is any space homeomorphic to the set of all points in $\mathbb{R}^3$ that are a unit distance from the origin. That

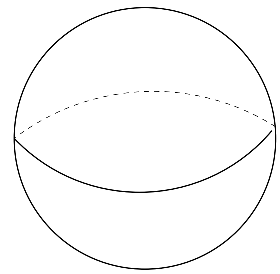


Figure 12.1: A 2-sphere is the boundary of a ball.

is, a 2-sphere is homeomorphic to the boundary of a ball. Similarly, a 1-sphere, denoted $\mathbb{S}^1$, is any space homeomorphic to the set of all points in $\mathbb{R}^2$ that are a unit distance from the origin, that is, a circle.

The next simplest compact 2-manifold is the boundary of a donut, called a **torus**. The torus is denoted $\mathbb{T}^2$ and, being the surface of a donut, it is a delicious example of a 2-manifold. See Figure 12.2.

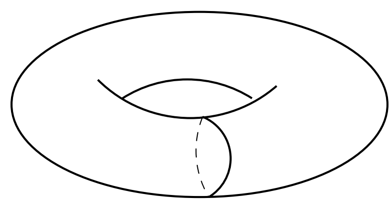


Figure 12.2: A torus.

Exercise 12.1. *Show that the torus T^2 is homeomorphic to $\mathbb{S}^1 \times \mathbb{S}^1$.*

Are the torus and sphere homeomorphic? While your intuition may tell you no, without some kind of proof you can't be sure there isn't some secret homeomorphism that you haven't thought of that magically turns one into the other.

Consider another pair of examples. Take a rectangular sheet of infinitely thin rubber and glue one pair of parallel edges together (forming a tube), then glue the other pair of parallel edges together. With a little effort, you can visualize this object and see that it is a torus. Now consider a hexagonal sheet of rubber, and successively glue the 3 pairs of parallel edges together. It is far less obvious that the space you get from this construction is also a torus—it's not even obvious the space is a 2-manifold. Thus in this instance, two spaces that look quite different or are described quite differently are actually homeomorphic.

In spite of such difficulties, we will see in this chapter how to classify all compact, connected 2-manifolds. This classification entails the challenge of distinguishing various spaces that are identical locally but differ globally. A preliminary step in that challenge is to recognize that any compact 2-manifold is homeomorphic to an object that is constructed from finitely many flat triangles. Objects made of finitely many simple pieces allow us to use combinatorial techniques to understand their global properties.

The major visual difference between the 2-sphere and the 2-torus is that the latter has a 'hole.' This observation leads us to consider 2-manifolds constructed in the same way as the torus, but with more holes. For example, the **double torus** is shown in Figure 12.3.

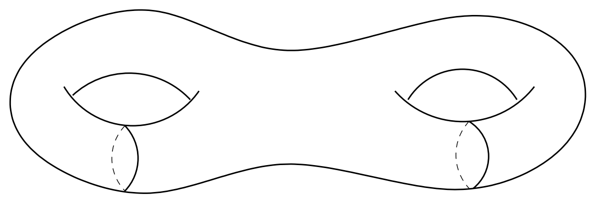


Figure 12.3: The double torus or surface of genus 2.

You can similarly construct surfaces with arbitrarily many holes. Some examples are shown in Figures 12.4 and 12.5. Notice that with the triple torus, all the holes are shown "in a line" whereas for the quadruple torus, the holes are arranged in a circle.

Exercise 12.2. *For a given number of holes, demonstrate that the n-holed torus where the holes are lined up is homeomorphic to an n-holed torus where the holes are arranged in a circle.*

For exercises like this that ask you to demonstrate a geometric homeomorphism, we are not asking you to define a formal homeomorphism—no equations are expected. Rather, it suffices to describe a process by which you would systematically distort one figure to look like the other figure—something like drawing the frames that make up an animated cartoon. Or perhaps you will think of some way to describe a visual one-to-one correspondence.

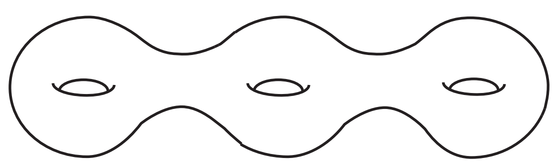


Figure 12.4: The triple torus or surface of genus 3.

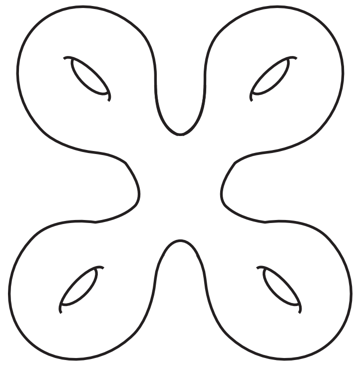


Figure 12.5: The quadruple torus or surface of genus 4.

So far all the 2-manifolds we have seen can be embedded in $\mathbb{R}^3$ (which is certainly advantageous for visualization). The 2-manifold shown in Figure 12.6, known as the **Klein bottle** and denoted $\mathbb{K}^2$, suggests that embeddability in $\mathbb{R}^3$ may not be possible for every 2-manifold. While the Klein bottle cannot be embedded in $\mathbb{R}^3$ (we will prove this fact in the chapter on homology), it can be embedded in $\mathbb{R}^4$. This embedding into $\mathbb{R}^4$ can be visualized from the figure by using the fourth dimension to avoid the self-intersection.

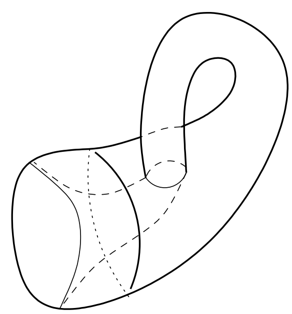


Figure 12.6: The Klein bottle.

Another 2-manifold that cannot be embedded in $\mathbb{R}^3$ is known as the **projective plane** (or **real projective** 2**-space**) and denoted $\mathbb{R}\mathrm{P}^2$. It is the space of all the lines in $\mathbb{R}^3$ that pass through the origin. That is, each straight line through the origin in $\mathbb{R}^3$ is a point of $\mathbb{R}\mathrm{P}^2$. The basis for the topology is the collection of open cones with the cone point at the origin.

Exercise 12.3. *1. Show that $\mathbb{R}\mathrm{P}^2 \cong \mathbb{S}^2/\langle x \sim -x\rangle$, that is, the projective plane is homeomorphic to the 2-sphere with diametrically opposite points identified.*

2. Show that $\mathbb{R}\mathrm{P}^2$ is also homeomorphic to a disk with two edges on its boundary (called a ***bigon****),* *identified as indicated in Figure 12.7.*

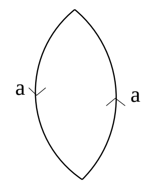


Figure 12.7: $\mathbb{RP}^2$.

3. *Show that the Klein bottle can be realized as a square with certain edges identified.*

12.2 The Classification of 1-Manifolds

Effective Thinking Principle. *Start with an Easier Question.* When faced with a difficult challenge, don't do it. Instead, find and do a related, easier challenge.

The central theorem for this chapter is a classification theorem for 2-manifolds. As a warm up to that result, we will ask you to state and prove a classification theorem for compact, connected 1-manifolds.

Definition. A topological space is a 1**-manifold** if and only if it is a separable metrizable space where every point is in an open set homeomorphic to an open interval in $\mathbb{R}^1$.

Theorem 12.4. *Suppose M is a compact, connected 1-manifold. Then M is triangulable. That is, M is homeomorphic to a subset C of $\mathbb{R}^n$ consisting of a finite collection of straight line segments where any two segments of C are either disjoint or meet at an endpoint of each.*

Exercise 12.5. *Provide a complete classification of compact, connected 1-manifolds. That is, describe a collection of topological spaces such that every compact, connected 1-manifold is homeomorphic to one member of the collection.*

The next exercise just allows disconnected 1-manifolds.

Exercise 12.6. *Provide a complete classification of compact 1-manifolds.*

12.3 Triangulability of 2-Manifolds

The totality of this short section merely asserts that for now we will be accepting without proof the following fact. This theorem is difficult to prove, so we defer a discussion of its proof until the Chapter 15.

Theorem 12.7. *Every compact 2-manifold is triangulable, that is, it is homeomorphic to a subset C of $\mathbb{R}^n$ consisting of a finite collection $T = \{\sigma_i\}_{i=1}^k$ of rectilinear triangles (a fancy word for a rectilinear triangle is a 2-simplex) where each pair of triangles are disjoint or they meet in one vertex of each or they share a single edge. Since the space C is homeomorphic to a 2-manifold, each edge of each triangle making up C is shared by exactly two triangles, and around each vertex is a circle of triangles whose union is a disk.*

In the sequel, we will sometimes refer to sets of the edges of triangles that make up C, so we have a name for them:

If $T = \{\sigma_i\}_{i=1}^n$ is the collection of 2-simplices that make up a 2-manifold C in $\mathbb{R}^n$, then the 1-skeleton of T is the set of all the edges of the triangles in T.

12.4 The Classification of 2-Manifolds

This section presents a proof to the classification theorem for compact, connected, triangulated 2-manifolds. The basic strategy of this proof is to show that removing an open disk from a compact, connected, triangulated 2-manifold gives us a space homeomorphic to a (closed) disk with some number of bands attached to its boundary in a specified way. The number of bands, and how they are attached then gives us a classification of the surface.

The disk we will remove is a neat neighborhood of some of the edges of the triangles that make up the 2-manifold. In order to create an appropriately small neighborhood that captures the intuitive idea of just taking a little neighborhood around those edges, we define a barycentric subdivision.

For the entirety of the proof of the classification theorem, we encourage you to think very concretely about these 2-manifolds. Think of them as physical objects that you could hold and touch. The 2-manifold is made of flat triangles. You know what a triangle is. A barycentric subdivision just divides each single triangle into exactly 6 pieces.

Definition. Let M^2 be a 2-manifold with triangulation $T = \{\sigma_i\}_{i=1}^k$. The **barycentric subdivision** T' of T is the collection $T' = \{\sigma_i'\}_{i=1}^{6k}$ of triangles obtained by taking each 2-simplex (that is, triangle) in the collection of triangles T and dividing it into exactly six sub-triangles by drawing straight line segments from the center of each side to the opposite vertex.

Notice that the union of the triangles in T' is exactly the same as the union of triangles in T, there are just 6 times as many of them. Often we do this barycentric subdivision process twice. That process creates the second barycentric subdivision of T, denoted T''. Notice that there are 36 times as many triangles in T'' as there are in T, but the underlying subset of $\mathbb{R}^n$ is exactly the same.

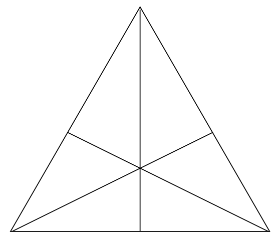


Figure 12.8: Barycentric subdivision of a 2-simplex.

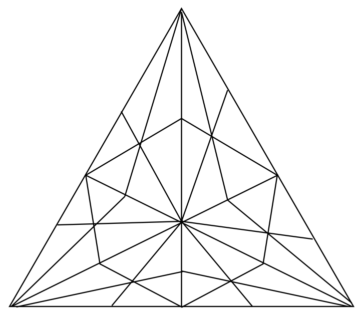


Figure 12.9: Second barycentric subdivision of a 2-simplex.

Now we can specify what we mean by a neat neighborhood of some collection of edges of T.

Definition. Let M^2 be a 2-manifold in $\mathbb{R}^n$ with triangulation $T = \{\sigma_i\}_{i=1}^k$. Let A be the union of some of the vertices and edges of the triangles in T. The **regular neighborhood** of A, denoted $N(A)$, equals $\bigcup\{\sigma''_j \mid \sigma''_j \in T''$ and $\sigma''_j \cap A \neq \varnothing\}$ (recall that T'' is the second barycentric subdivision of T).

Exercise 12.8. *The boundary of a tetrahedron is naturally triangulated with a triangulation T consisting of four 2-simplexes, having six edges and four vertices.*

1. *On the boundary of a tetrahedron draw the first and second barycentric subdivisions of T.*
2. *Locate the edges of the four triangles in T.*
3. *Draw the regular neighborhood of the union of all the edges of T.*
4. *Draw the regular neighborhood of a single edge of a triangle in T.*

We can create a torus by taking a square and identifying the top edge to the bottom edge and the right side to the left side. We can triangulate the torus with 18 triangles as shown, and you could literally create a physical model of such a triangulated torus that you could hold in your hand.

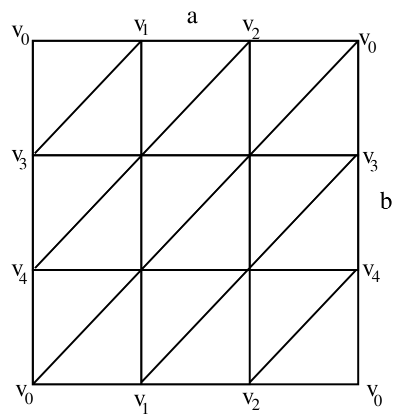


Figure 12.10: The torus triangulated.

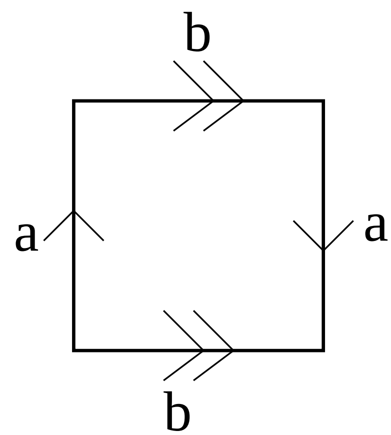


Figure 12.11: A Klein bottle is obtained by gluing the edges as indicated.

Similarly, we can create a Klein bottle by taking a square and identifying the top edge to the bottom edge without a twist and then identifying the right side to the left side with a half twist, that is, the top of the right edge is identified with the bottom of the left edge and vice versa. Again the Klein bottle can be triangulated using 18 triangles. In the case of the Klein bottle, you cannot create a triangulated Klein bottle that lives in Euclidean 3-space; however, you can create such a Klein bottle in $\mathbb{R}^4$. Regardless of its inability to be created in 3-space, the pre-identification diagram for the Klein bottle can be drawn on a piece of paper, as you see in Figure **??**.

Exercise 12.9. *In the second barycentric subdivisions of a triangulation of the torus (Figure* **??***), find regular neighborhoods of various subsets of the edges.*

It may be useful to remember two definitions from graph theory: the definition of a *graph* and the definition of a *tree*. Recall that a graph (V, E) is a set of vertices V and a set E of pairs of those

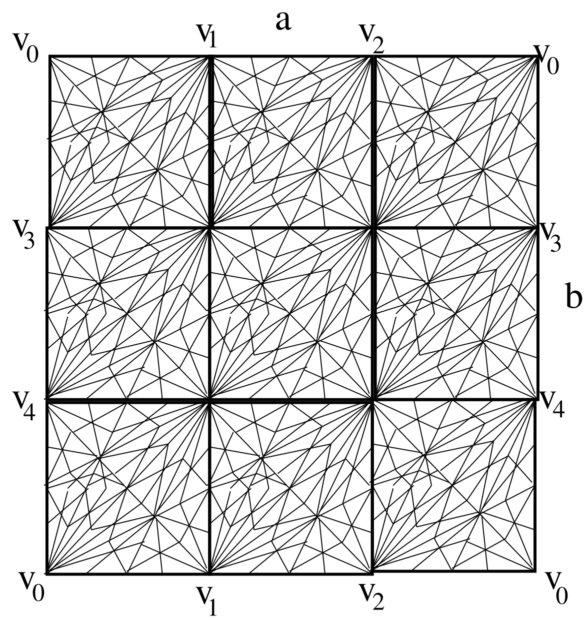


Figure 12.12: The torus triangulated by its second barycentric subdivision.

vertices, which are the edges of the graph. A tree is a graph (V, E) that has no circuits, that is, the graph has no sequence of edges $\{v_1, v_2\}$, $\{v_2, v_3\}$, $\{v_3, v_4\}$, ...$\{v_{k-1}, v_k\}$,$\{v_k, v_1\}$ where no two edges in the sequence are the same edge.

Exercise 12.10. *Consider the triangulation of the torus in Figure 12.4. Describe those graphs created from edges in the* 1*-skeleton of* T *that have regular neighborhoods homeomorphic to a disk.*

Effective Thinking Principle. *Generalize Insights.* After noticing something in an individual instance, see whether that same insight applies to more general cases.

Your observation in the exercise above can be generalized.

Theorem 12.11. *Let* M^2 *be a compact, triangulated* 2*-manifold with triangulation* T*. Let* S *be a tree whose edges are* 1*-simplices in the* 1*-skeleton of* T*. Then* $N(S)$*, the regular neighborhood of* S*, is homeomorphic to* $\mathbb{D}^2$.

When we look at a triangulated 2-manifold it is natural to connect the centers of adjacent 2-simplices. That naturally leads to the concept of the dual 1-skeleton.

Definition. Let T be a triangulation of a 2-manifold M. Consider two adjacent 2-simplices in T, σ_1 and σ_2. There are two edges of 2-simplices in the first barycentric subdivision of T that create a path from the barycenter of σ_1 to the barycenter of σ_2. One edge goes from the barycenter of σ_1 to the midpoint of the common edge of σ_1 and σ_2, and the second edge goes from that midpoint to the barycenter of σ_2. The union of those two edges is viewed as a single 'edge' in the dual 1-skeleton of T. The **dual 1-skeleton of** T consists of the set of all such 'edges' between adjacent 2-simplices of T.

The next theorem encourages you to do some drawing and observing. Draw a picture of a few triangles that might be part of the triangulation of a 2-manifold. Then draw the second barycentric subdivision of them and highlight a tree consisting of a few 'edges' in the dual 1-skeleton of the original triangulation. Now shade in all the 2-simplices in the second barycentric subdivision that touch your tree and see why their union must be a disk. That look at the local geometry of your triangulated 2-manifold will allow you to see why the following theorem is true.

Theorem 12.12. *Let M^2 be a compact, triangulated 2-manifold with triangulation T. Let S be a tree equal to a union of 'edges' in the dual 1-skeleton of T. Then $\cup\{\sigma''_j \mid \sigma''_j \in T'' \text{ and } \sigma''_j \cap S \neq \varnothing\}$ is homeomorphic to $\mathbb{D}^2$.*

In the next theorem you will show that a tree in the 1-skeleton of a triangulated 2-manifold cannot disconnect the dual 1-skeleton.

Theorem 12.13. *Let M^2 be a connected, compact, triangulated 2-manifold with triangulation T. Let S be a tree in the 1-skeleton of T. Let S' be the subgraph of the dual 1-skeleton of T whose 'edges' do not intersect S. Then S' is connected.*

The following two theorems state that M^2 can be divided into two pieces, one a disk D_0, and the other a disk D_1 with bands (the H_i's) attached to it. You might think about creating D_0 by fattening up a maximal tree in the 1-skeleton of the triangulation of M^2 and then creating D_1 by finding an appropriate tree in the dual 1-skeleton and fattening it up. Just observe that the only things that could be left over must be disjoint disks, the H_i's.

Theorem 12.14. *Let M^2 be a connected, compact, triangulated 2-manifold. Then $M^2 = D_0 \cup D_1 \cup \left(\bigcup_{i=1}^{k} H_i\right)$ where D_0, D_1, and each H_i is homeomorphic to $\mathbb{D}^2$, $\operatorname{Int} D_0 \cap D_1 = \varnothing$, the H_i's are disjoint, $\bigcup_{i=1}^{k} \operatorname{Int} H_i \cap (D_0 \cup D_1) = \varnothing$, and for each i, $H_i \cap D_1$ equals 2 disjoint arcs each arc on the boundary of each of H_i and D_1.*

Theorem 12.15. *Let M^2 be a connected, compact, triangulated 2-manifold. Then:*

1. *There is a disk D_0 in M^2 such that $M^2 - (\operatorname{Int} D_0)$ is homeomorphic to the following subset of $\mathbb{R}^3$: a disk D_1 with a finite number of disjoint strips, H_i for $i \in \{1, \ldots n\}$, attached to boundary of D_1 where each strip has no twist or a $1/2$ twist. (See Figure 12.13.)*

2. *Furthermore, the boundary of the disk with strips, $D_1 \cup \left(\bigcup_{i=1}^{k} H_i\right)$, is connected.*

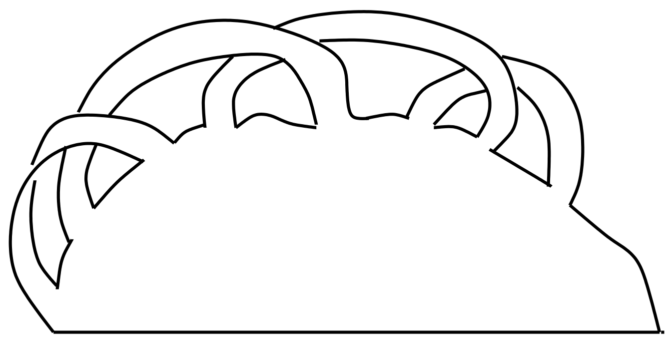


Figure 12.13: A disk with four handles attached.

As you work on the following exercise, remember that the boundary of D_0 (a simple closed curve) is equal to the boundary of $D_1 \cup \left(\bigcup_{i=1}^{k} H_i\right)$.

Exercise 12.16. *In the conclusion of the previous theorem, any strip H_i divides the boundary of D_1 into two arcs, e_i^1 and e_i^2, where H_i is* not *attached, that is, the two arcs that make up $(D_1 \bigcap H_i)$ are disjoint from the two arcs e_i^1 and e_i^2 except at their endpoints. Show that if a strip H_j is attached to D_1 with no twists, then there must be a strip H_k that is attached to both e_j^1* and *e_j^2.*

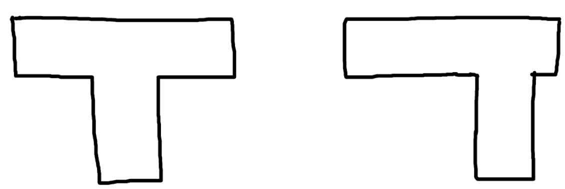


Figure 12.14: Distorting a T-shaped object by a homeomorphism.

Here is a little geometry. Suppose you have a rectangle with another rectangle hanging down as pictured in Figure 12.14. Then that T shaped object is homeomorphic to a distorted T where the stem of the T is moved to appear to be attached a bit to one side. This type of distortion can be used to slide strips over strips to put our disk with strips into a more orderly arrangement.

Theorem 12.17. *Let M^2 be a connected, compact, triangulated 2-manifold. Then there is a disk D_0 in M^2 such that $M^2 -$* Int *D_0 is homeomorphic to a disk D_1 with strips attached as follows: first come a finite number of strips with $1/2$ twist each of whose attaching arcs are consecutive along* Bd *D_1, and next come a* *finite number of pairs of untwisted strips, each pair with attaching arcs entwined as pictured with the four arcs from each pair consecutive along* $\mathrm{Bd}\, D_1$.

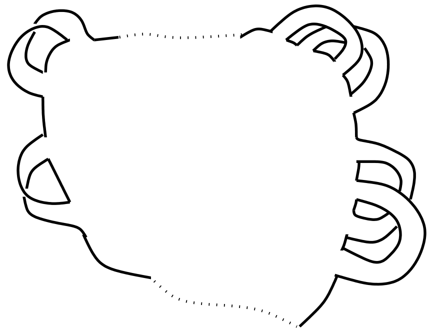


Figure 12.15: Twisted strips and entwined strips.

The next theorem abandons general cases for a moment, and, instead, asks you to show the topological equivalence of two specific sets, X and Y: each set is a disk with three strips attached. You could physically make X and Y if you wished to.

Theorem 12.18. *Let X be the union of a disk with three strips attached as follows: a disk E_0 with one strip attached with a $1/2$ twist with its attaching arcs consecutive along* $\mathrm{Bd}\, E_0$ *and one pair of untwisted strips with attaching arcs entwined as pictured with the four arcs consecutive along* $\mathrm{Bd}\, E_0$. *Let Y again be a union of a disk with three strips attached, but the three are attached differently. The set Y consists of a disk E_1 with three strips with a $1/2$ twist each whose attaching arcs are consecutive along* $\mathrm{Bd}\, E_1$. *Then X is homeomorphic to Y.*

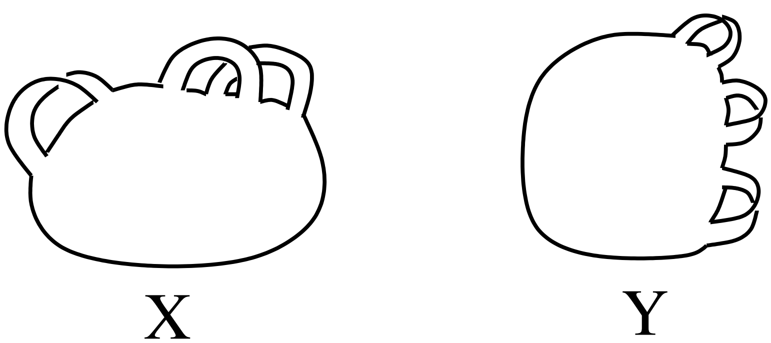


Figure 12.16: These spaces are homeomorphic.

Effective Thinking Principle. *Simplify When Possible.* If a potentially complicated situation can be made simpler or more orderly, do so.

Our drive toward organization leads us to create neater arrangements of our strips. The following result demonstrates that upon the removal of a single disk, the remainder of a compact, connected, 2-manifold is an easily described subset of $\mathbb{R}_3$.

Theorem 12.19. *Let M^2 be a connected, compact, triangulated 2-manifold. Then there is a disk D_0 in M^2 such that $M^2 - \operatorname{Int} D_0$ is homeomorphic to one of the following:*

a) a disk D_1,

b) a disk D_1 with k $\frac{1}{2}$-twisted strips with consecutive attaching arcs, or

c) a disk D_1 with k pairs of untwisted strips, each pair in entwining position with the four attaching arcs from each pair consecutive.

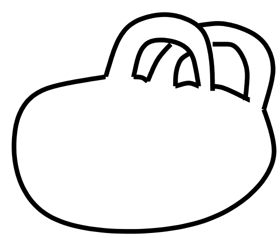


Figure 12.17: Entwining pair of strips.

It will turn out that the patterns of strips that arise in this classification result correspond in a natural way to tori and projective planes. To capture that connection, we can re-phrase this classification theorem more memorably by defining a method for combining two 2-manifolds to produce a third 2-manifold.

12.5 The Connected Sum

In this section we will develop the concept of a *connected sum*, an important method for creating a new manifold out of others. The connected sum will allow us to state the classification theorem of compact 2-manifolds more succinctly.

Definition. Let K_T and L_S be two connected, compact, triangulated 2-manifolds in $\mathbb{R}^n$. Choose 2-simplices $\sigma_1 \in T$ and $\sigma_2 \in S$. Select a homeomorphism h from the boundary of σ_1 to the boundary of σ_2 that is linear on each of the three edges. Consider the topological space created by taking the sets $K - \operatorname{Int}\sigma_1$ union $L - \operatorname{Int}\sigma_2$ and identifying their boundaries via h (where the interior of a 2-simplex is defined to be the 2-simplex except for its edges and vertices). The resulting surface is the **connected sum** of M_1 and M_2, denoted $M_1 \# M_2$.

Theorem 12.20. *Suppose M_1 and M_2 are compact, triangulated, connected 2-manifolds and M is a connected sum of M_1 and M_2 (that is, we can select triangulations of M_1 and M_2, apply the process above and arrive at a space homeomorphic to M). Then M is a compact, connected, triangulable 2-manifold.*

Notice that, given two 2-manifolds, the definition of connected sum depends on several choices: we have to select a triangulation of each manifold, a 2-simplex from each, and a homeomorphism between their boundaries. In fact, (up to homeomorphism, of course), the resulting surface does not depend on our choices (as long as we assume each manifold is connected, which we have). For now, we will just accept the well-definedness of connected sum as fact.

You should note that '#' is commutative and associative.

Exercise 12.21. *Suppose M is a compact, connected, triangulated 2-manifold. What is $\mathbb{S}^2 \# M$?*

Exercise 12.22. *Sketch $\mathop{\#}\limits_{i=1}^{n} \mathbb{T}^2$.*

We can now re-state our classification result in terms of connected sums. Your challenge in proving the next theorem is to explain the geometry of how the connected sum construction relates to the adding of more strips to a disk.

Theorem 12.23 (Classification of compact, connected 2-manifolds)**.** *Any connected, compact, triangulated 2-manifold is homeomorphic to the 2-sphere $\mathbb{S}^2$, a connected sum of tori, or a connected sum of projective planes.*

At this juncture, you have demonstrated that any compact, connected, triangulated 2-manifold is homeomorphic to one of the connected sums listed above; however, you have not yet shown that these connected sums are all different from one another. We will undertake that challenge in future sections of this chapter, but before doing so, we wish to guide you to discover an alternative method for demonstrating that any compact, connected, triangulated 2-manifold is homeomorphic to a sphere, a connected sum of tori, or a connected sum of projective planes.

12.6 Polygonal Presentations of 2-Manifolds

In the previous sections we noticed that it was convenient to look at a torus as a square with the right and left hand sides identified and with the top and bottom sides identified. Likewise, a Klein bottle was similarly described as a rectangle with various points on the boundary identified. Viewing 2-manifolds as quotient spaces of polygons is a useful way to visualize and analyze 2-manifolds.

Let's consider the process by which a torus can be shown to be homeomorphic to a polygon with edges identified in pairs. See Figure 12.18. By cutting along two particular curves in the torus, we can 'unroll' the torus to get a square in the plane with the property that if we identify pairs of edges, we see that the quotient space after that identification is the torus we started with—the identification simply repairs the cuts we made in the first place.

We can carry out a similar process in Figure 12.19 on the double torus to get an octagon such that if appropriate pairs of its edges are identified in pairs, we reconstruct the double torus. These presentations have the major advantage that they lie in the plane, which in some cases make them easier to work with visually. We will refer to these quotient manifestations as **polygonal presentations**.

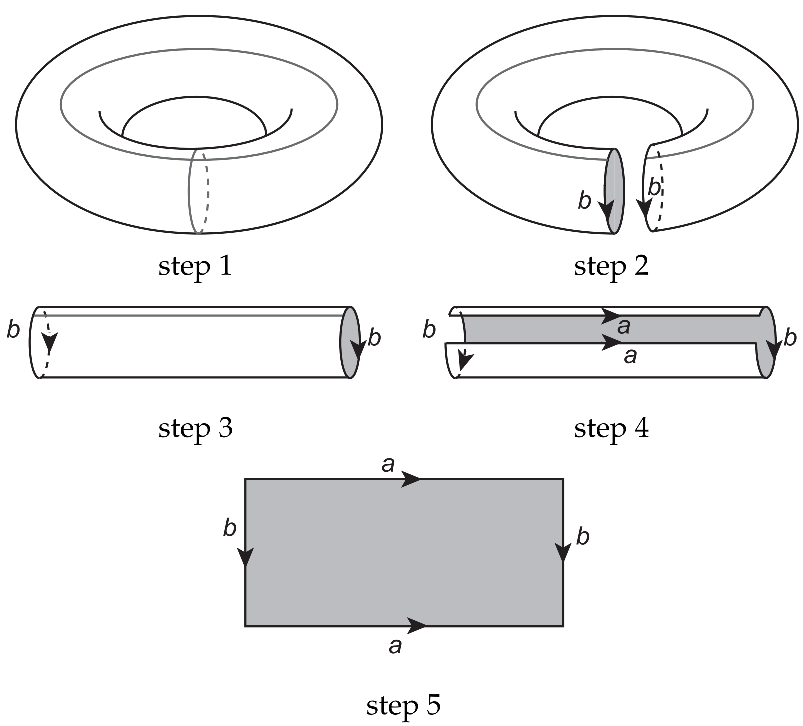


Figure 12.18: Finding a polygonal presentation for the torus.

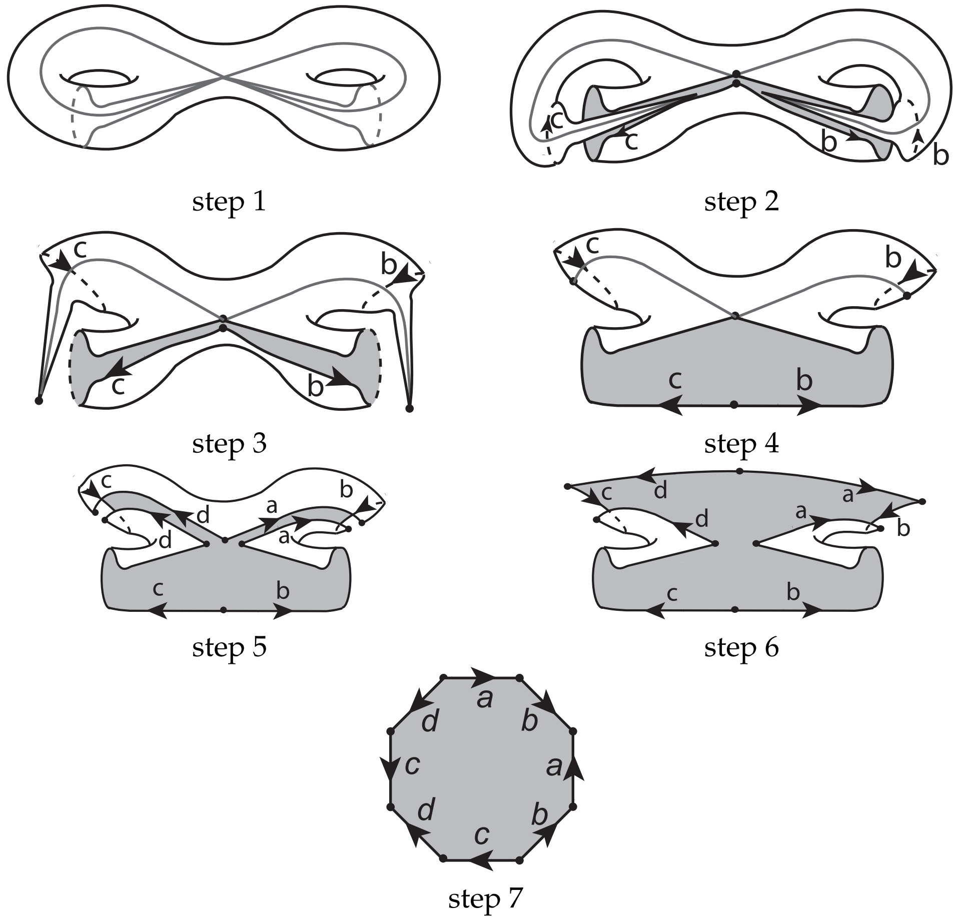
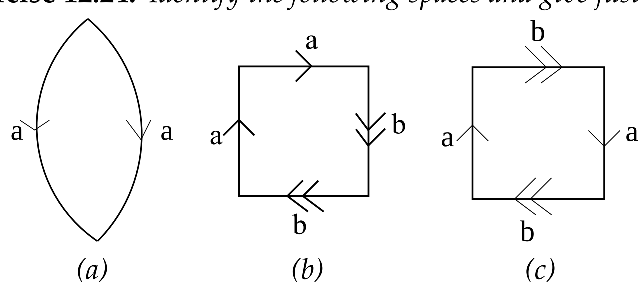


Figure 12.19: Finding a polygonal for the double torus.

Exercise 12.24. *Identify the following spaces and give justification.*



All the spaces above have their edges identified in pairs. In fact, you will prove next that if you start with a polygon and identify its edges in pairs, the quotient space will always be a 2-manifold. We will reserve the term **polygonal presentation** for the quotient space obtained by starting with a polygon in the plane and identifying the edges in pairs. For simplicity, we will assume all our identifications are linear on each edge.

Theorem 12.25. *Let P be a polygonal presentation. Then P is a 2-manifold.*

We end this section with a simple notation for polygonal presentations. Suppose we have a polygonal presentation. If we assign a unique letter to each pair of edges that are glued together (as was done in our figures) and we read the letters as we follow the edges along the boundary of the disk (starting at a certain edge) going clockwise, we get a "word" made up of these letters. However, to completely specify the gluing, we need to know not only which edges are glued together, but in which orientation. To keep track of orientations, we will write the letter alone if the orientation of the corresponding arrow points clockwise and we will write the letter with a -1 superscript if the arrow points counterclockwise. For example, by the process in Figure 12.18, $aba^{-1}b^{-1}$ corresponds to a torus. Exercise 12.24 demonstrates that different words can generate the same 2-manifold.

Theorem 12.26. *Suppose M is a compact, connected, triangulable 2-manifold. Then M is homeomorphic to a polygonal presentation.*

The converse of the last theorem also holds. In the next theorem the compactness and connectedness are rather easy, and you have already proved that a polygonal presentation is a 2-manifold. Please show the triangulability directly, that is, without appealing to the fact that every 2-manifold is triangulable.

Theorem 12.27. *Let P be a polygonal presentation. Then P is a compact, connected, triangulable 2-manifold.*

We will now proceed to discover an alternative proof that every compact, connected, triangulated 2-manifold is homeomorphic to a sphere, a connected sum of tori, or a connected sum of projective planes.

12.7 Another Classification of Compact 2-Manifolds

You have now shown that every compact, connected, triangulated 2-manifold is homeomorphic to a polygon with edges identified pairwise. Hence, one way to classify these manifolds is to classify polygonal presentations. The strategy is to begin with a given polygonal presentation and show that it yields the same 2-manifold as increasingly neater polygonal presentations.

In the next sequence of theorems you will prove that any polygonal presentation is homeomorphic either to $\mathbb{S}^2$ or to a canonical polygonal presentation for $\mathop{\#}\limits_{i=1}^{n} \mathbb{T}^2$ and $\mathop{\#}\limits_{i=1}^{n} \mathbb{RP}^2$.

The first theorem shows that an edge identified to an adjacent edge in the opposite direction can be erased.

Theorem 12.28. *Let $Abb^{-1}C$ be a string of $2n$ letters where each letter occurs twice, neglecting superscripts (and there is at least one pair other than b and b^{-1}). Then the 2-manifold obtained from the word $Abb^{-1}C$ is homeomorphic to that obtained from AC.*

The next theorem will require your using a strategy of proof that you may find useful in many of the theorems to come. Namely, think about taking a given polygonal presentation simply cutting the polygon into two polygons by a straight cut from one vertex to another. Record the fact that you could recover your original 2-manifold by identifying the two exposed edges that you have just created along the cut. Now suppose one of the edge pairs in the original presentation has one of them on one of the two pieces and the other one on the other piece. Then you could pick up one piece and attach make the identification of those two edges. Understand why you now have a new polygonal presentation of the same 2-manifold. By judicious choices about where to cut and glue, you can successfully prove the following theorems.

Theorem 12.29. *Suppose P is a polygonal presentation not homeomorphic to $\mathbb{S}^2$. Then there is a homeomorphic polygonal presentation where all the vertices are in the same equivalence class, that is, all the vertices are identified to each other.*

The theorems that follow are sequential in that you incrementally add virtues to the polygonal presentation as you proceed.

Theorem 12.30. *Suppose P is a polygonal presentation not homeomorphic to $\mathbb{S}^2$. Then P is homeomorphic to a polygonal presentation where all the vertices are identified and for every pair of edges with the same orientation, the two edges of that pair are consecutive.*

Theorem 12.31. *Suppose P is a polygonal presentation not homeomorphic to $\mathbb{S}^2$. Then P is homeomorphic to a polygonal presentation where all the vertices are identified, every pair of edges with the same orientation are consecutive, and all other edges are grouped in disjoint sets of two intertwined pairs following the pattern $aba^{-1}b^{-1}$.*

The following theorem is very concrete. It says if a particular pattern occurs in the word, that pattern can be replaced with another pattern.

Theorem 12.32. *If A and C are (possibly empty) words, then the polygonal presentation $Aaba^{-1}b^{-1}ccC$ is homeomorphic to that represented by $AddeeffC$.*

We can now put every polygonal presentation into one of three categories.

Theorem 12.33. *Any compact, connected, triangulated 2-manifold M is homeomorphic to the polygonal presentation given by one of the following words: aa^{-1}, $a_1a_1 \dots a_na_n$ (where $n \geq 1$) or $a_1a_2a_1^{-1}a_2^{-1} \dots a_{n-1}a_na_{n-1}^{-1}a_n^{-1}$ (where $n \geq 2$ is even).*

This classification result talks about words in polygonal presentations. It would be nice to see the relationship between these words and the more geometrically appealing idea of connected sum. That is the connection explored in the next exercise. To answer the next exercise, you may want to recall that the word $aba^{-1}b^{-1}$ corresponds to the torus and the word $aba^{-1}b^{-1}cdc^{-1}d^{-1}$ corresponds to the double torus.

Exercise 12.34. *Suppose that we have two compact, connected 2-manifolds represented by the words w_1 and w_2, respectively. Suppose in addition that w_1 and w_2 have no letters in common. What can you say about the 2-manifold corresponding to the concatenated word w_1w_2 in terms of the connected sum?*

Recall that the word $aba^{-1}b^{-1}$ is a polygonal presentation for the torus and the word aa^{-1} is a polygonal presentation for the projective plane.

Exercise 12.35. *Re-state Theorem 12.32 above in the case that A and C are empty, in terms of connected sum.*

We can now re-phrase our classification scheme of polygonal presentations in terms of connected sum. Fortunately, we get the same result as we arrived at with our previous approach to the classification question.

Theorem 12.36 (Classification of compact, connected 2-manifolds)**.** *Any compact, connected, triangulated 2-manifold is homeomorphic to exactly one of the following:*

1. $\mathbb{S}^2$,
2. *a connected sum of n tori, and*
3. *a connected sum of n projective planes.*

Once again, we have not yet shown that a given compact, connected 2-manifold appears only once on this list. Pinning that down requires a few more steps.

12.8 Orientability

One of the challenges with the classification of 2-manifolds is to show why similar 2-manifolds such as the Klein bottle and the torus are different topologically. Indeed, perhaps you have a sense about the twisted nature of the Klein bottle that suggests that the Klein bottle is different from the torus in a way that the 2-sphere is not different from the torus. The purpose of this section is to formulate this twisted difference precisely by way of an invariant known as *orientability*.

Orientability is one of the many concepts in topology (or in mathematics in general) that is easy to understand on an intuitive level but somewhat difficult to pin-down rigorously. Loosely speaking, a connected surface is orientable if we can, near each point, assign a 'clockwise' direction so that this direction varies continuously and consistently as we move along the surface. Such a choice of direction is called an *orientation* of the manifold.

Exercise 12.37. *Describe heuristically a strategy by which you would define a consistent clockwise direction on the standard embedding of the 2-sphere in $\mathbb{R}^3$. What is the relevant property?*

We will use the vertices of a 2-simplex to make this definition precise. We begin by describing what we mean by an orientation of an edge and a 2-simplex.

Definition. Two orderings of the vertices v_0, v_1, v_2 of a 2-simplex are said to be **equivalent** if they differ by an even permutation. (e.g., $\{v_0, v_1, v_2\} \sim \{v_1, v_2, v_0\}$ and $\{v_0, v_1, v_2\} \not\sim \{v_1, v_0, v_2\}$). This equivalence relation produces precisely two equivalence classes of orderings of vertices of a 2-simplex. The equivalence class of an ordering $\{v_0, v_1, v_2\}$ will be denoted $[v_0 v_1 v_2]$.

In the case of an edge with vertices v_0 and v_1, there are two orderings of the two vertices.

Definition. An **orientation** of an edge or a 2-simplex is a choice of one of the two equivalence classes of orderings of its vertices. Then the chosen one is called *positive* and the other one is called *negative*, and if σ is an **oriented simplex**, its negatively oriented counterpart is denoted $-\sigma$.

Note that an orientation of a 1-simplex can be represented by drawing an arrow along the edge in one of two directions. Likewise an orientation of a 2-simplex can be represented by drawing either a clockwise arrow or a counterclockwise arrow on it and declaring that to be the positive direction.

If we choose an orientation of a 2-simplex, then there are associated orientations on each of its edges.

Definition. If $[v_0 v_1 v_2]$ is the positive orientation of a 2-simplex, then the **induced orientation** on its three edges are what you expect by going around the boundary of the triangle, namely: $[v_0 v_1]$, $[v_1 v_2]$, and $[v_2 v_0]$ are positive.

Exercise 12.38. *Show that the induced orientation is well defined; in other words, that it is independent of the choice of positive equivalence class representative for the original 2-simplex.*

We can now define what we mean by an orientable, triangulated 2-manifold. A triangulated 2-manifold is orientable if it is possible to select orientations for each 2-simplex in such a way that neighboring 2-simplices have compatible orientations. The concept of 'compatible' comes from the following observation. If you draw two triangles in the plane that share an edge and orient them both in a counterclockwise ordering (for example) then the shared edge has induced orientations from the two triangles that are opposite. In other words, when the orientations on both triangles are the same, the induced orientations on a shared edge are opposite. This realization gives rise to the following definition.

Definition. A triangulated 2-manifold, (M^2, T), is **orientable** if an orientation can be assigned to each 2-simplex in T in such a way that given any 1-simplex in the triangulation, the two 2-simplices that share it as a face induce opposite orientations. Otherwise, (M^2, T) is **non-orientable**. A choice of (consistent) orientation for each 2-simplex in T is called an **orientation** of M^2.

Since our definition of orientability depends heavily on a triangulation, it is a little bit tricky to show that orientability is a topological invariant. What really needs to be shown are two things: (1) if you subdivide the triangles in a triangulation, you get the same answer about orientability; and (2) for any two triangulations of the same 2-manifold, there are subdivisions of each that make the two subdivisions isomorphic. We will not undertake that challenge until Chapter 15. For now, please assume those facts and prove the following equivalences of orientability.

Theorem 12.39. *Show that the following are equivalent for a 2-manifold M.*

1. *Every triangulation of M is not orientable (that is if K_T is a simplicial complex to which M is homeomorphic, K_T is not orientable).*
2. *M admits a triangulation that is not orientable.*
3. *M admits a triangulation that contains a collection of simplices whose union is homeomorphic to the Möbius band.*
4. *M admits an embedding of a Möbius band.*
5. *There is a map $F : \mathbb{S}^1 \times [0,1] \to M$ such that $F(\cdot, t)$ is an embedding for each t and such that $F(\cdot, 1) = F(r(\cdot), 0)$, where r is a reflection map of $\mathbb{S}^1$ about some line through its center.*

Either the third or the fourth statement above will allow us to conclude that orientability is a topological invariant after we have proved that 2-manifolds can be triangulated.

Theorem 12.40. *Let $M_1, \ldots, M_n$ be connected, compact, triangulated 2-manifolds. Let M be a connected sum of $M_1, \ldots, M_n$. Then M is orientable if and only if M_i is orientable for each $i \in \{1, \ldots, n\}$.*

Exercise 12.41. *State and prove which compact, connected, triangulated 2-manifolds are orientable and which are not.*

12.9 The Euler Characteristic

In this section, we will define the *Euler characteristic*. This number, which is associated to a triangulation of a 2-manifold, is vastly important in many applications (which should be clear after this section). In particular, it will allow us to deepen significantly our understanding of compact 2-manifolds. The formulation is surprisingly simple.

Definition. Suppose K_T is a compact, triangulated 2-manifold with triangulation T. Let V be the number of vertices, E be the number of edges, and F be the number of 2-simplices in T. Then the Euler characteristic of K_T is

$$\chi(K_T) = V - E + F.$$

Notice that the sum defining Euler characteristic is always well-defined since a triangulation must have finitely many simplices.

The Euler characteristic is a topological invariant. The work needed to show this fact is of similar character to showing that 2-manifolds are triangulable, so we will defer a discussion of its being an invariant until Chapter 15 when we discuss triangulability of compact 2-manifolds. For now please just assume that the Euler characteristic is a topological invariant.

In any case, the computation of the Euler characteristic of a 2-manifold does not depend on its particular triangulation. For this reason, we will typically use the notation $\chi(K)$ rather then $\chi(K_T)$. Of course to calculate the Euler characteristic, a triangulation must be chosen.

Exercise 12.42. *Calculate the Euler characteristic of the following spaces.*

1. $\mathbb{S}^2$
2. $\mathbb{T}^2$
3. $\mathbb{K}^2$
4. $\mathbb{R}\mathrm{P}^2$

We will now demonstrate the usefulness of the Euler characteristic.

Lemma 12.43. *Suppose M_1 and M_2 are compact 2-manifolds. If $M_1 \# M_2$ is* any *choice for the connect sum of M_1 and M_2, then $\chi(M_1 \# M_2) = \chi(M_1) + \chi(M_2) - 2$.*

Exercise 12.44.

1. *Calculate the Euler characteristic of $\overset{n}{\underset{i=1}{\#}} \mathbb{R}\mathrm{P}^2$.*
2. *Calculate the Euler characteristic of $\overset{n}{\underset{i=1}{\#}} \mathbb{T}^2$.*

Theorem 12.45. *The combination of Euler characteristic and orientability is a complete invariant of compact, connected* 2*-manifolds.*

The Euler characteristic and orientability makes the identification of 2-manifolds almost trivial in many instances.

Exercise 12.46. *Identify the following* 2*-manifolds as a sphere, a connected sum of* n *tori (specifying* n*), or a connected sum of* n *projective planes (specifying* n*).*

a. $\mathbb{T}\#\mathbb{RP}$

b. $\mathbb{K}\#\mathbb{RP}$

c. $\mathbb{RP}\#\mathbb{T}\#\mathbb{K}\#\mathbb{RP}$

d. $\mathbb{K}\#\mathbb{T}\#\mathbb{T}\#\mathbb{RP}\#\mathbb{K}\#\mathbb{T}$

Exercise 12.47. *Identify the surface obtained by identifying the edges of the decagon as indicated in Figure 12.20.*

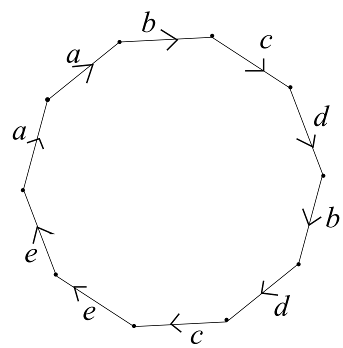


Figure 12.20: The decagon with edges identified in pairs.

12.10 Manifolds with Boundary

An important generalization of the concept of a 2-manifold is a type of space called a **manifold with boundary**. We first give several important examples of 2-manifolds with boundary.

The first example is a space we have already met: the disk. The next is the annulus, which is a disk with a smaller (concentric) open disk removed (see Figure 12.21).

Another 2-manifold with boundary, appropriately named the 'pair of pants', is shown in Figure 12.22

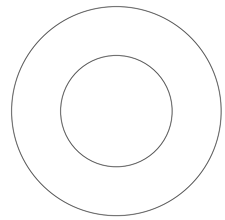


Figure 12.21: The annulus.

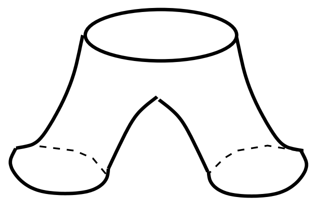


Figure 12.22: The pair of pants.

A disk with handles attached also qualifies. A disk with two intertwined handles attached is shown in Figure 12.23.

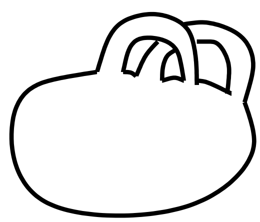


Figure 12.23: Disk with intertwined handles.

A famous example of a 2-manifold with boundary is a space known as the Möbius band. This space can be constructed by taking a strip of paper, putting a half twist in, and taping the ends together. It is shown in Figure 12.24.

Exercise 12.48. *Notice that the edge (or boundary) of a Möbius band is a simple close curve. Construct a space by gluing a disk to the Möbius band along their respective boundaries. Show that this space is homeomorphic to the projective plane.*

We will now give the precise definition of a manifold with boundary.

Definition. The set

$$\{(x_1, x_2) \in \mathbb{R}^2 | x_2 \geq 0\}$$

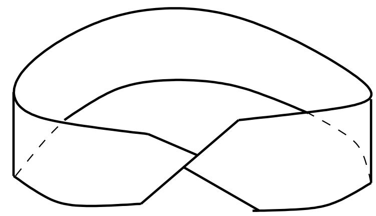


Figure 12.24: The Möbius band.

in $\mathbb{R}^2$ is called 2**-dimensional upper half-space** and is denoted $\mathbb{R}^2_+$.

Definition. A 2**-dimensional manifold with boundary** or 2**-manifold with boundary** is a second countable, Hausdorff space, M, such that for each $p \in M$, there is a neighborhood U of p that is homeomorphic to an open set in $\mathbb{R}^2_+$.

Notice that this definition is indeed a generalization since an (appropriate) open 2-ball is still an open subset of $\mathbb{R}^2_+$.

Any number of 2-manifolds with boundary can be created by starting with a 2-manifold and making it holey by punching some holes (interiors of disks) in it.

12.11 Classifying 2-Manifolds: Going Below the Surface of Surfaces

In this chapter we took on the challenge of analyzing surfaces, that is, topological spaces that locally look like the plane. These surfaces are appealing objects including the sphere, the torus, double and triple tori, the sensuous Klein bottle, and infinitely many more.

One of the basic impulses of mathematics and perhaps of life in general is to organize related objects in satisfying ways. In the case of compact, connected 2-manifolds, we saw that every such surface actually was constructed from just three basic building blocks—the sphere, the torus, and the projective plane. At first, we might well have guessed that putting together a torus and a projective plane via the connected sum, for example, might have produced a 2-manifold that could not be created by just using projective planes alone; however, we saw that every connected, compact 2-manifold was in fact either a sphere or it could be creating by taking the connected sum of only tori or only projective planes. It is quite satisfying to find that this whole category of surfaces are generated by such a short list of elemental building blocks.

One feature of surfaces that we just barely saw small glimpses of was the aesthetic richness of artistic renderings of these beautiful ideas. We leave it to you to open your eyes to finding graceful and elegant visions of surfaces in nature and in art.

Chapter 13

Fundamental Group: Capturing Holes

In the last chapter, we considered how to distinguish surfaces by manipulating the spaces and putting them in some canonical form to compare them. In this chapter, we'll consider how to distinguish spaces by observing how features *inside* them can be deformed, such as paths and loops. This focus will lead naturally to the question of whether a space has a 'hole' in it.

Whether a space has an obvious hole in it is an important idea that pops up in unexpected places. For instance, consider this famous theorem about polynomials:

Theorem (Fundamental Theorem of Algebra). *A polynomial $p(z) = a_n z^n + ... + a_1 z + a_0$ with complex coefficients and degree $n > 1$ has at least one root.*

This theorem does not seem at first to be a topological theorem! However, we can get our first hint that topology plays a role by looking at a special case.

Exercise 13.1. *A polynomial $p(x) = a_n x^n + ... + a_1 x + a_0$ with real coefficients where $a_n \neq 0$ and n is odd has at least one real root.*

One strategy for showing the existence of a root is to ask whether the graph of p crosses the x-axis. By observing the values of the polynomial on the endpoints of an interval $[-K, K]$ when K is sufficiently large, you can see that the existence of a root follows from the Intermediate Value Theorem.

Now let's ask what happens when you consider an arbitrary polynomial with complex coefficients. We can proceed by analogy. The polynomial p is now a function from the complex plane to the complex plane, so instead of considering what p does to a very large interval on the real line, consider what p does to the circular curve C on the boundary of a large disk around the origin in the complex plane.

If you know a little complex arithmetic, you'll recall that to multiply two complex numbers, you multiply the distances from the origin and add the angles. So the polynomial z^n takes the circle C centered at the origin to a circle that wraps around the origin n times. For a sufficiently large z, the leading term $a_n z^n$ dominates all the lower degree terms, so the image of C under p will be a really large curve that wraps around the origin n times. What does this fact imply about the image of the disk that C bounds? Can the image of the disk avoid the origin if its boundary wraps around the origin non-zero number of times?

Consider the disk as the union of concentric circles together with a center point. If the center point does not itself map to zero, the continuous image of a tiny circle will map to a tiny loop that does not wind around origin. By following the image of circles of increasing radius under the map p, we see a continuous deformation between a loop that does not wind around the origin to the image of C which winds around the origin n times. Must this deformation cross the origin eventually? We now see how this question about algebra has turned into a topological question about deformations of loops in the complex plane.

13.1 Invariants and Homotopy

Given two topological spaces, how can we tell whether they are homeomorphic? To prove they are homeomorphic, we would need to prove that there is a specific map that witnesses the homeomorphism. To prove they are not homeomorphic, we might seek a property preserved by homeomorphism (called an *invariant*) that one space possesses but the other does not. For instance, properties from point-set topology such as compactness, connectedness, metrizability, countability, separability, etc. can be useful in this regard.

However, consider the 2-dimensional sphere and the torus (the surface of a donut). Both spaces are metrizable, connected, compact, second-countable, etc. so they are 'nice' from the point of view of point-set topology. Nevertheless, it is intuitively clear that they are not topologically equivalent. To distinguish spaces like these, we will need to construct different invariants. We saw that orientability and the Euler characteristic were useful invariants for distinguishing surfaces. But now we turn our attention to the idea of associating a group with a space.

Effective Thinking Principle. *Make Intuition Precise.* One of the best methods for creating ideas is to take an intuitive idea and make it precise.

Consider the most apparent difference between the sphere and the torus. Intuitively, we would say the torus contains something we would describe as a "hole," whereas the sphere does not. The **fundamental group**, first developed by Henri Poincaré at the turn of the 20$^{\text{th}}$ century, is one way to make our intuitive notion precise.

Unlike the invariants of point-set topology, which are typically properties that a space may or may not possess (like compactness, metrizability, separability, etc.) or a number (like the number of connected components), the fundamental group is an algebraic group. One would expect this more complex invariant to carry more information about a space, and it often does.

An annulus, that is, the region between two concentric circles, captures some of the basic ideas of the fundamental group. An annulus has the shape of a race track. Consider a path on the annulus that begins and ends at the same point (which we later call the *base point*). A path that goes half-way and comes back should somehow be different from a path that goes around the annulus. One of those paths goes around the hole and the other doesn't. A path that goes half way around, comes back, and then goes half around the other way before coming back is qualitatively different from the path that goes all the way around. If you were racing in a car, going around the path is a good way to try to win the race, whereas just going back and forth near the starting line will probably not lead to victory. A path that goes twice around should be different from going around once, as should a path that goes all the way around in the opposite direction. The fundamental group is a structure that effectively measures these differences (and similarities) and in this way the fundamental group captures the intuitive idea that an annulus has a hole.

To construct the fundamental group, we must first study the concept of *homotopy*. Suppose that $f, g : X \to Y$ are (continuous) maps from one topological space to another. Loosely speaking, we say that these maps are *homotopic* if we can continuously deform the image of f to the image of g, all the while remaining inside Y.

Definition. Let X and Y be topological spaces, and let $f, g : X \to Y$ be continuous functions. Then f is **homotopic** to g (written $f \simeq g$) if and only if there is a continuous map $F : X \times [0,1] \to Y$ such that these equations hold:

$$F(x,0) = f(x)$$
$$F(x,1) = g(x)$$

for all $x \in X$. The map F is called a **homotopy** between f and g.

We often denote the second argument of a homotopy $F(x,t)$ by the letter t and we can think of t as a time parameter. Then the homotopy $F : X \times [0,1] \to Y$ between two maps f and g can be viewed a continuous 1-parameter family of maps $F_t : X \to Y$ that deforms $F_0 = f$ at time 0 into $F_1 = g$ at time 1.

For instance, if X is a square, then a continuous function $f : X \to \mathbb{R}^3$ could specify some image of the square in $\mathbb{R}^3$. The function f is sort of like the position of a flag (if the flag were allowed to cross through itself). Two continuous functions of X, f and g, are homotopic if there is

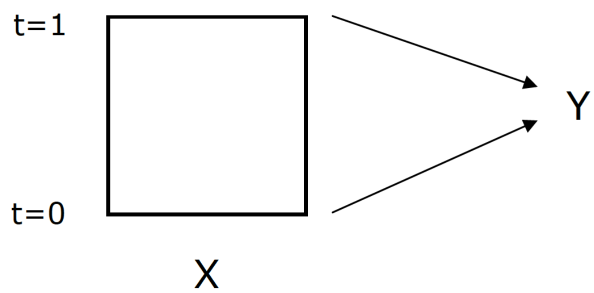


Figure 13.1: A homotopy maps $X \times [0,1]$ into another space Y. Diagrams like this can be helpful to visualize homotopies.

a continuous way to deform one image at time 0 into the other at time 1.

Any two continuous functions $f, g : X \to \mathbb{R}^3$ are homotopic using the **straight-line homotopy** in $\mathbb{R}^3$:

$$F(x,t) = (1-t)f(x) + tg(x).$$

Note that F is continuous, equal to f at time 0, and equal to g at time 1. It is called a straight line homotopy because for a fixed x, the image $F(x,t)$ moves in a straight line from $f(x)$ to $g(x)$. The straight-line homotopy is available to us because any two points in $\mathbb{R}^3$ can be connected by a straight line.

In contrast, suppose we consider two continuous functions f and g from a circle $\mathbb{S}^1$ into the punctured plane $\mathbb{R}^2 - (0,0)$. Similarly to our investigation into the Fundamental Theorem of Algebra, suppose f takes $\mathbb{S}^1$ to a circle that goes around $(0,0)$, for example, f could be the identity map. And suppose g takes $\mathbb{S}^1$ to a circle that does not go around $(0,0)$, for example, suppose $g(\mathbb{S}^1)$ is a unit circle centered at the point $(1,1)$. Then there is no homotopy between f and g in $\mathbb{R}^2 - (0,0)$. The straight line homotopy does not work, because for some point $x \in \mathbb{S}^1$, the straight line between $f(x)$ and $g(x)$ will cross $(0,0)$. Proving that there is such a point is not obvious, but you will prove it soon enough.

Definition. A function $f : X \to Y$ whose image is a single point is called a **constant map**. A map is said to be **null homotopic** if and only if it is homotopic to the constant map.

Effective Thinking Principle. *Seek Refinements.* Once you have created an idea, seek refinements and variations.

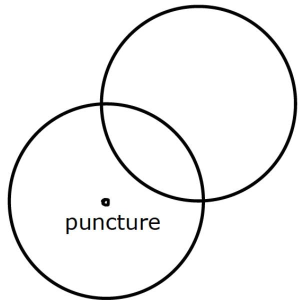


Figure 13.2: These circles are images of two maps from $\mathbb{S}^1$ into the punctured plane that are not homotopic.

Often we will want our maps to be homotopic in such a way that during the transition from one map into the other, certain function values remain fixed throughout. This idea is captured in the next definition.

Definition. Given topological spaces X and Y with $S \subset X$, two continuous functions $f, g : X \to Y$ are **homotopic relative to** S if and only if there is a continuous function $H : X \times [0, 1] \to Y$ such that

$$\begin{array}{rclcll} H(x,0) & = & f(x) & & & \text{for all } x \in X \\ H(x,1) & = & g(x) & & & \text{for all } x \in X \\ H(x,t) & = & f(x) & = & g(x) & \text{for all } x \in S \text{ and } t \in [0,1] \end{array}$$

Functions f and g being homotopic relative to S is denoted $f \simeq_S g$.

In other words, H is a homotopy that leaves the image of every point in S fixed throughout the entire process.

Theorem 13.2. *Given topological spaces X and Y with $S \subset X$, homotopy relative to S is an equivalence relation on the set of all continuous functions from X to Y. In particular, if $S = \varnothing$, homotopy is an equivalence relation on the set of all continuous functions from X to Y.*

To establish the conditions for an equivalence relation in this theorem, you'll need to use given homotopies to construct other homotopies. You may find the Pasting Lemma (from Chapter 8) useful.

We will be most interested in homotopies between paths. Recall that a **path** in a space X is a continuous function from $[0, 1]$ into X. Two paths will be considered equivalent if we can deform one into the other without moving the end points. In particular, equivalent paths must have the same starting point and the same ending point.

Definition. Two paths α, β are **equivalent**, denoted $\alpha \sim \beta$, if and only if α and β are homotopic relative to $\{0, 1\}$. The equivalence class of paths containing α is denoted by $[\alpha]$ (See Figure 13.3).

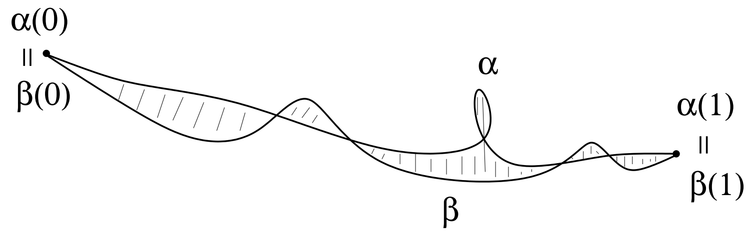


Figure 13.3: Path equivalence.

The physical idea of walking from point a to point b to point c yields the natural idea of how to combine paths.

Definition. Let α, β be paths with $\alpha(1) = \beta(0)$. Then their **product**, denoted $\alpha \cdot \beta$, is the path that first moves along α, followed by moving along β, defined explicitly by:

$$\alpha \cdot \beta(t) = \begin{cases} \alpha(2t), & 0 \leq t \leq \frac{1}{2} \\ \beta(2t-1), & \frac{1}{2} < t \leq 1 \end{cases} .$$

Notice the need to speed up in order to trace out both the paths α and β during the 1 unit of time allotted for a path.

We can show this notion of product is well-defined on equivalence classes. To do this, you will need to construct a specific homotopy. Some may find it helpful to draw a diagram of the domain of the homotopy $[0, 1] \times [0, 1]$ and marking on the domain where the various points go.

Theorem 13.3. *If α, α', β, and β' are paths in a space X such that $\alpha \sim \alpha'$, $\beta \sim \beta'$, and $\alpha(1) = \beta(0)$, then $\alpha \cdot \beta \sim \alpha' \cdot \beta'$.*

Thus products of paths can be extended to products of equivalence classes by defining

$$[\alpha] \cdot [\beta] := [\alpha \cdot \beta].$$

The previous theorem shows that this product is well-defined.

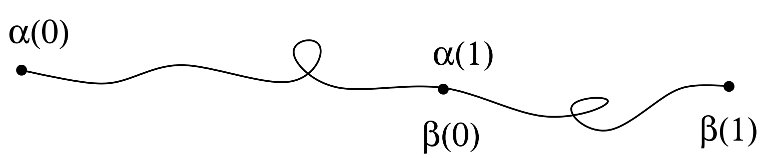


Figure 13.4: Path Product.

Effective Thinking Principle. *Details Bring Understanding.* Personally going through the details of fundamentals brings a clarity of understanding that merely nodding at expected results can never bring.

Next, we can show that this product has the associative property. Once again, you will need to construct an explicit homotopy.

Theorem 13.4. *Given paths α, β, and γ where the following products are defined, then $(\alpha\cdot\beta)\cdot\gamma \sim \alpha\cdot(\beta\cdot\gamma)$ and $([\alpha]\cdot[\beta])\cdot[\gamma] = [\alpha]\cdot([\beta]\cdot[\gamma])$.*

If we think of a path α as taking us from $\alpha(0)$ to $\alpha(1)$, then tracing out that same image in reverse yields a natural inverse.

Definition. Let α be a path, then its **path inverse** α^{-1} is the path defined by $\alpha^{-1}(t) = \alpha(1-t)$.

If we take a path and then take its inverse, the combined path is equivalent to not moving at all.

Definition. Let X be a topological space. A map $e_{x_0} : [0,1] \to X$ that sends each point of $[0,1]$ to single point x_0 is called a **constant path**.

Theorem 13.5. *Let α be a path with $\alpha(0) = x_0$. Then $\alpha \cdot \alpha^{-1} \sim e_{x_0}$, where e_{x_0} is the constant path at x_0.*

Paths that begin and end at the same point will be of special interest, because they will capture the idea of finding 'holes' in the space.

Definition. Let X be a topological space. A path $\alpha : [0,1] \to X$ is called a **loop** or a **closed path based at** x_0 if and only if $\alpha(0) = \alpha(1) = x_0$.

The concept of path equivalence applies to loops as well, since loops are paths that begin and end at the same point. Thus two loops α and β based at a point x_0 are equivalent if there is a homotopy between α and β holding the endpoints fixed at x_0.

Definition. Let X be a topological space. A loop α based at x_0 is **homotopically trivial** or is a **trivial loop** if and only if α is equivalent to the constant path at x_0.

We now have the ingredients to associate a group with a topological space X. This group has been designed to capture the idea of holes in the space by looking at the space of all loops in X and regarding as equivalent loops that are homotopic to one another.

Definition. Let $x_0 \in X$, a topological space. Then the set of equivalence classes of loops based at x_0 with binary operation $[\alpha]\cdot[\beta] = [\alpha\cdot\beta]$ is a called the **fundamental group of X based at x_0** and is denoted $\pi_1(X, x_0)$. The point x_0 is called the **base point** of the fundamental group.

Theorem 13.6. *The fundamental group $\pi_1(X, x_0)$ is a group. The identity element is the class of homotopically trivial loops based at x_0.*

The fundamental group is defined for a space X with a specified base point selected. However, for many spaces the choice of base point is not significant, because the fundamental group computed using one base point is isomorphic to the fundamental group using any other point. In particular, path connected spaces exhibit this independence of base points.

Theorem 13.7. *If X is path connected, then $\pi_1(X, p) \cong \pi_1(X, q)$ for any points $p, q \in X$.*

Because of this isomorphism, we sometimes just write $\pi_1(X)$ for the fundamental group of a path connected space X. However, to actually exhibit elements of this group, we must choose a base point. Also, the proof shows that this isomorphism is not *canonical*, meaning that there is not one obvious choice of isomorphism. There could be different isomorphisms depending on which path is chosen from p to q.

Corollary 13.8. *Suppose X is a topological space and there is a path between points p and q in X. Then $\pi_1(X, p)$ is isomorphic to $\pi_1(X, q)$.*

Technically, a loop is a path whose endpoints are mapped to the same place, but intuitively, a loop is a map from the unit circle $\mathbb{S}^1$ into the space. That intuition is formalized using the following wrapping map.

Definition. The map $\omega : \mathbb{R}^1 \to \mathbb{S}^1$ defined by $t \mapsto (\cos 2\pi t, \sin 2\pi t)$ is called the **standard wrapping map** of $\mathbb{R}^1$ to $\mathbb{S}^1$.

Exercise 13.9. *Let α be a loop into a topological space X. Then $\alpha = \beta \circ \omega|_{[0,1]}$ where ω is the standard wrapping map and β is some continuous function from $\mathbb{S}^1$ into X. This relationship gives a correspondence between loops in X and continuous maps from $\mathbb{S}^1$ into X.*

So we may think of a loop as a map from a circle when it is useful for us to do so. This description also allows us to state a useful characterization of triviality of a loop.

Theorem 13.10. *Let X be a topological space and let p be a point in X. Then a loop $\alpha = \beta \circ \omega|_{[0,1]}$ (where ω is the standard wrapping map and β is a continuous function from $\mathbb{S}^1$ into X) is homotopically trivial if and only if β can be extended to a continuous function from the unit disk $\mathbb{D}^2$ to X.*

Effective Thinking Principle. *Look at Examples.* Specific examples often give understanding and insight that general theorems do not convey.

Let's compute some fundamental groups. We begin with some spaces that have trivial fundamental groups.

Theorem 13.11. *Show the following (1 denotes the trivial group):*

1. $\pi_1([0,1]) \cong 1$.
2. $\pi_1(\mathbb{R}^n) \cong 1$ *for* $n \geq 1$.
3. $\pi_1(X) \cong 1$, *if* X *is a convex set in* $\mathbb{R}^n$.
4. $\pi_1(X) \cong 1$, *if* X *is a cone.*
5. $\pi_1(X) \cong 1$ *if* X *is a star-like space in* $\mathbb{R}^n$ *(a subset* X *of* $\mathbb{R}^n$ *is called* ***star-like*** *if there is a fixed point* $x_0 \in X$ *such that for any* $y \in X$*, the line segment between* x_0 *and* y *lies in* X*; a five pointed 'star' is an example of a star-like space that is not convex).*

Exercise 13.12. *Show the following:*

1. $\pi_1(\mathbb{S}^0, 1) \cong 1$ *where* $\mathbb{S}^0$ *is the zero-dimensional sphere* $\{-1, 1\}$*, the set of points unit distance from the origin in* $\mathbb{R}^1$.
2. $\pi_1(\mathbb{S}^2) \cong 1$.
3. $\pi_1(\mathbb{S}^n) \cong 1$ *for* $n \geq 3$.

Definition. A path-connected topological space with trivial fundamental group is said to be **simply connected**.

The next space, shown in Figure 13.5, is called the Hawaiian earring. It consists of an infinite sequence of circles, each of radius half the last, that all intersect and are tangent at a single point.

Exercise 13.13. *Show that the cone over the Hawaiian earring is simply connected. Can you generalize your insight?*

Of course, the fundamental group would not serve a useful purpose if all spaces were simply connected. Our first example of a space with a non-trivial fundamental group is the circle. The following theorem will require some significant work to prove. You may find the Lebesgue Number Theorem to be pertinent, see Theorem 10.25.

Theorem 13.14. *1. Any loop* $\alpha : [0,1] \to \mathbb{S}^1$ *with* $\alpha(0) = 1$ *can be written* $\alpha = \omega \circ \widetilde{\alpha}$*, where* $\widetilde{\alpha} : [0,1] \to \mathbb{R}^1$ *satisfies* $\widetilde{\alpha}(0) = 0$ *and* ω *is the standard wrapping map.*

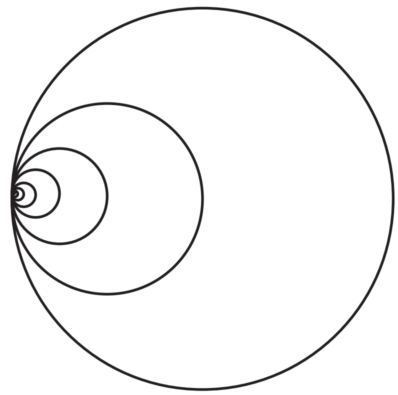


Figure 13.5: The Hawaiian earring.

2. *If* $\alpha : [0,1] \to \mathbb{S}^1$ *is a loop, then* $\widetilde{\alpha}(1)$ *is an integer.*

3. *Loops* α_1 *and* α_2 *are equivalent in* $\mathbb{S}^1$ *if and only if* $\widetilde{\alpha_1}(1) = \widetilde{\alpha_2}(1)$.

4. $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$.

To add to the spaces whose fundamental groups we can compute, let us now look at the Cartesian products of spaces and observe that the fundamental group of a product of topological spaces is just the product of the fundamental groups of the factors. (The group structure on a product of groups does the obvious thing, performing products in each factor.)

Theorem 13.15. *Let* (X, x_0), (Y, y_0) *be path connected spaces. Then*

$$\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \times \pi_1(Y, y_0)$$

via the canonical map that takes a loop γ *in* $X \times Y$ *to* $(p \circ \gamma, q \circ \gamma)$ *where* $p : X \times Y \to X$ *and* $q : X \times Y \to Y$ *are the projection maps.*

Exercise 13.16. *Find:*

1. $\pi_1(X)$ *where* X *is a solid torus.*

2. $\pi_1(\mathbb{S}^2 \times \mathbb{S}^1)$

3. $\pi_1(\mathbb{S}^2 \times \mathbb{S}^2 \times \mathbb{S}^2)$

4. $\pi_1(X)$, *where* X *is a direct product of* k_n *copies of* $\mathbb{S}^n$, *with* $k_n = 0$ *for* n *sufficiently large.*

We can understand the fundamental group of a torus more precisely than the above exercise indicates.

Exercise 13.17. *The fundamental group of the torus $\pi_1(\mathbb{T}^2)$ is $\mathbb{Z}^2$. Moreover, if μ is a meridian and λ is a longitude, then $\{[\mu], [\lambda]\}$ is a $\mathbb{Z}$-basis for $\pi_1(\mathbb{T}^2)$.*

13.2 Induced Homomorphisms and Invariance

Effective Thinking Principle. *Structures, Then Maps.* After defining a mathematical concept, investigate how the concept behaves under transformations.

A standard technique in mathematics is to explore how the structure of one mathematical object is transported to another object via a map. Since topological spaces are mapped to each another via continuous functions, we will study the effect of a continuous function on the fundamental group.

Definition. Let $f : X \to Y$ be a continuous function. Then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ defined by $f_*([\alpha]) = [f \circ \alpha]$ is called the **induced homomorphism on fundamental groups**.

Exercise 13.18. *Check that for a continuous function $f : X \to Y$, the induced homomorphism f_* is well-defined (that is, the image of an equivalence class is independent of the chosen representative). Show that it is indeed a group homomorphism.*

Our next theorem shows that the induced homomorphisms obey the so-called **functorial properties**. The notation $f : (X, x_0) \to (Y, y_0)$ means $f : X \to Y$ is a function that satisfies $f(x_0) = y_0$.

Theorem 13.19. *The following are true:*

1. *If $f : (X, x_0) \to (Y, y_0)$ and $g : (Y, y_0) \to (Z, z_0)$ are continuous maps, then $(g \circ f)_* = g_* \circ f_*$.*
2. *If $id : (X, x_0) \to (X, x_0)$ is the identity map, then $id_* : \pi_1(X, x_0) \to \pi_1(X, x_0)$ is the identity homomorphism.*

The fundamental group would have almost no use from a topological point of view if different manifestations of the same space could have distinct fundamental groups. Fortunately, topologically equivalent spaces have isomorphic fundamental groups.

Theorem 13.20. *If $h : X \to Y$ is a homeomorphism then*

$$h_* : \pi_1(X, x_0) \to \pi_1(Y, h(x_0))$$

is a group isomorphism. Thus homeomorphic path-connected spaces have isomorphic fundamental groups.

In other words, the fundamental group of a path connected space is a topological invariant, and hence we can establish that two path connected spaces are not homeomorphic if we can show that they have different (non-isomorphic) fundamental groups. Thus, the fundamental group helps to distinguish among spaces, but it is not a *complete invariant*, meaning there are spaces that are not homeomorphic that have isomorphic fundamental groups, as you have already seen.

Our association of an algebraic group to a topological space is extremely useful, because we can use algebra to answer certain topological questions and, as we shall see later, we can also use topology to answer algebraic questions. The following result is an example of the first technique.

Theorem 13.21. *Fix a torus with chosen meridian μ and longitude λ. Suppose $p, q \in \mathbb{Z}$. Then there is a homeomorphism of the torus to itself which takes a representative of the class $q[\mu] + p[\lambda] \in \pi(\mathbb{T}^2)$ to μ if and only if p and q are relatively prime.*

As a consequence, the construction of a 3-manifold known as a lens space $L(p, q)$ can only be constructed when p and q are relatively prime. Lens spaces are discussed in Section 13.5.

13.3 Homotopy Equivalence and Retractions

Theorem 13.22. *If $f, g : (X, x_0) \to (Y, y_0)$ are continuous functions and f is homotopic to g relative to x_0, then $f_* = g_*$.*

Definition. Two spaces X and Y are **homotopy equivalent** or have the same **homotopy type** if there exist continuous maps $f : X \to Y$ and $g : Y \to X$ such that

$$g \circ f \simeq id_X \text{ and } f \circ g \simeq id_Y$$

where id_X denotes the identity on X and id_Y denotes the identity on Y. The function g is a **homotopy inverse** of f. Spaces X and Y being homotopy equivalent is denoted $X \sim Y$. The functions f and g are called **homotopy equivalences**.

Lemma 13.23. *Homotopy equivalence of spaces is an equivalence relation.*

Homotopy equivalent spaces have isomorphic fundamental groups. One of the challenges of the proof is to deal with the base points.

Theorem 13.24. *If $f : X \to Y$ is a homotopy equivalence and $y_0 = f(x_0)$, then $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ is an isomorphism. In particular, if $X \sim Y$, then $\pi_1(X) \cong \pi_1(Y)$.*

An important special type of homotopy equivalence is the following.

Definition. Let $A \subset X$. A continuous function $r : X \to A$ is a **strong deformation retraction** if and only if there is a homotopy $R : X \times [0,1] \to X$ such that

$$\begin{array}{rcll} R(x,0) & = & x & \text{for all } x \in X \\ R(x,1) & = & r(x) & \text{for all } x \in X \\ R(a,t) & = & a & \text{for all } a \in A \text{ and } t \in [0,1]. \end{array}$$

If $r : X \to A$ is a strong deformation retraction, then A is a **strong deformation retract** of X.

Thus a strong deformation retraction is homotopic to the identity map and leaves points of A fixed throughout the homotopy.

Exercise 13.25. *Show that for $n \geq 0$, $\mathbb{R}^{n+1} - \{0\}$ can be strong deformation retracted onto $\mathbb{S}^n$.*

Lemma 13.26. *If A is a strong deformation retract of X, then A and X are homotopy equivalent.*

We can use these insights to prove one special case of the Invariance of Domain Theorem.

Theorem 13.27. *$\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}^n$ for any $n \neq 2$.*

The theta space , as its name reflects, is a space shaped like the Greek letter theta. It is shown in Figure 13.6.

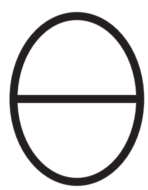


Figure 13.6: The theta space.

Exercise 13.28. *Let x and y be two points in $\mathbb{R}^2$. Show that $\mathbb{R}^2 - \{x,y\}$ strong deformation retracts onto the figure eight. In addition, show that $\mathbb{R}^2 - \{x,y\}$ strong deformation retracts onto a theta space.*

Theorem 13.29. *If $r : X \to A$ is a strong deformation retraction and $a \in A$, then $\pi_1(X,a) \cong \pi_1(A,a)$.*

Exercise 13.30. *Calculate the fundamental group of the following spaces.*

1. *An annulus.*

2. *A cylinder.*

3. *The Möbius Band.*

4. An open 3-ball with a diameter removed.

The fundamental group of a subspace does not necessarily inject into the fundamental group of the ambient space.

Exercise 13.31. *Find an example of a space X with a subspace A such that if $i : A \to X$ is the inclusion map, $i_* : \pi_1(A) \to \pi_1(X)$ is not injective.*

Nevertheless, a relaxation of the strong deformation retract condition gives us a situation under which we can conclude that i_* is injective.

Definition. Let $A \subset X$. A continuous function $r : X \to A$ is a **retraction** if and only if for every $a \in A$, we have $r(a) = a$. If $r : X \to A$ is a retraction, then A is a **retract** of X.

Theorem 13.32. *Let A be a retract of X via the inclusion $i : A \hookrightarrow X$ and retraction $r : X \hookrightarrow A$. Then for $a \in A$, $i_* : \pi_1(A, a) \to \pi_1(X, a)$ is injective and $r_* : \pi_1(X, a) \to \pi_1(A, a)$ is surjective.*

As a consequence, we can use the fundamental group to prove an intuitively plausible fact.

Theorem 13.33 (No Retraction Theorem for $\mathbb{D}^2$)**.** *There is no retraction from $\mathbb{D}^2$ to its boundary.*

The No Retraction Theorem for $\mathbb{D}^2$ can be used to produce another proof of the Brouwer Fixed Point Theorem for $\mathbb{D}^2$.

Theorem 13.34 (Brouwer Fixed Point Theorem for $\mathbb{D}^2$)**.** *Let $f : \mathbb{D}^2 \to \mathbb{D}^2$ be a continuous map. Then there is some $x \in \mathbb{D}^2$ for which $f(x) = x$.*

Definition. A topological space X is **contractible** if and only if the identity map on X is null homotopic (that is, if the identity map is homotopic to a constant map).

Lemma 13.35. *A space is contractible if and only if it is homotopy equivalent to a point.*

Theorem 13.36. *A contractible space is simply connected.*

Theorem 13.37. *A retract of a contractible space is contractible.*

Consider the house with two rooms , is shown in Figure 13.7. In words, this space resembles two hollow cubes, one stacked on the other (so that the top of the lower one is the bottom of the other). There is a tube running from the 'roof' of the complex into the lower cube and running from the 'floor' into the higher cube. To each tube is a attached a flange which connects it to the side, top, and bottom of the cube which it is inside (these are shown in grey).

Corollary 13.38. *The house with two rooms is contractible.*

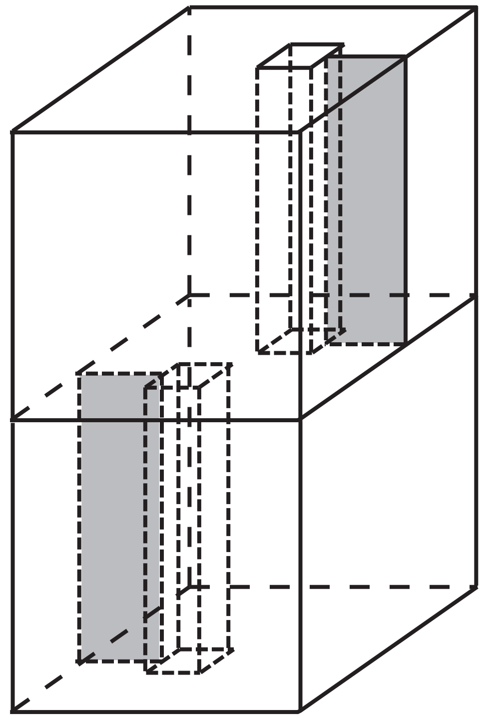


Figure 13.7: The house with two rooms.

Another iconic topological creation is the Dunce's Hat. This space is a quotient space constructed by taking a triangle and identifying each side to the other two by first gluing two sides together to form what would look like a real dunce's hat and then identifying the seam to the circle that is on the base of the hat by rolling it around the base. The figure shows the identification. The Dunce's Hat is contractible, but it is very hard to see how the contraction actually works.

Corollary 13.39. *The Dunce's Hat is contractible.*

13.4 Van Kampen's Theorem

At this stage, we have few tools for computing fundamental groups. So the question now becomes: How can we compute the fundamental group of spaces that are more complex than $\mathbb{S}^1$, such as many of the spaces we met in the previous chapters? So far, we can compute the fundamental group of a space in essentially three ways:

1. directly, that is by using an argument based on the specific geometry of the space (such as we did for $\mathbb{S}^1$ and many simply connected spaces);

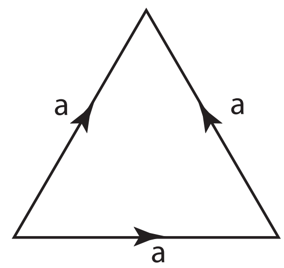


Figure 13.8: The dunce's hat.

2. by writing the space as a Cartesian product of spaces whose fundamental groups we know; and
3. by showing the space is homotopy equivalent to a space whose fundamental group we know.

> **Effective Thinking Principle.** *Complexity from Pieces.* Frequently, viewing objects as constructed from simpler pieces allows us to deal with far greater complexity than we could otherwise fathom.

To expand the number of spaces whose fundamental groups we can effectively compute, we will think of a space as the union of two subspaces and investigate how we can use knowledge about the fundamental groups of the two subspaces in order to deduce the fundamental group of the whole space. The result that describes that relationship among fundamental groups of subsets and the fundamental group of the whole space is called Van Kampen's Theorem.

> **Effective Thinking Principle.** *Start with Simple Cases.* Simple cases often exhibit essential ideas that unlock complex cases.

We'll begin with a special case that brings up several issues that will be important in deducing both the statement and the proof of the general case. This special case of Van Kampen's Theorem occurs when our space X is the union of two open sets U and V and $U \cap V$ is simply connected.

Theorem 13.40. *Let $X = U \cup V$, where U and V are open and path connected and $U \cap V$ is path-connected, simply connected, and non-empty. Then $\pi_1(X)$ is isomorphic to the free product of $\pi_1(U)$ and $\pi_1(V)$, that is, $\pi_1(X) \cong \pi_1(U) * \pi_1(V)$.*

Question 13.41. *Let X be the bouquet of n circles. What is $\pi_1(X)$?*

As usual, whenever we prove a theorem, we explore the necessity of the various hypotheses.

Exercise 13.42. *Find a path-connected space X with open, path-connected subsets U and V of X such that $X = U \cup V$ such that U and V are both simply connected, but X is not simply connected. Conclude that the hypothesis that $U \cap V$ is path connected is necessary.*

In proving the special case of Van Kampen's Theorem above, you probably observed that when a space is written as the union of two open sets that overlap in a path-connected way, each loop in the whole space is homotopic to a finite product of loops, each of which lies entirely in one of the two open sets. Let's formalize that observation in the following lemma.

Lemma 13.43. *Let $X = U \cup V$, where U and V are open and $U \cap V$ is path connected, and let $p \in U \cap V$. Then any element of $\pi_1(X, p)$ has a representative $\alpha_1\beta_1\alpha_2\beta_2 \cdots \alpha_n\beta_n$, where each α_i is a loop in U based at p and each β_i is a loop in V based at p.*

One of the hypotheses in the special case of Van Kampen's Theorem above is that two subsets whose union is X are open sets. The following example demonstrates the necessity of the hypothesis that U and V be open. Recall that the cone over a Hawaiian earring has trivial fundamental group.

Theorem 13.44. *Let X be a wedge of two cones over two Hawaiian earrings, where they are identified at the points of tangency of the circles of each Hawaiian earring, as in Figure 13.9. Then $\pi_1(X) \ncong 1$.*

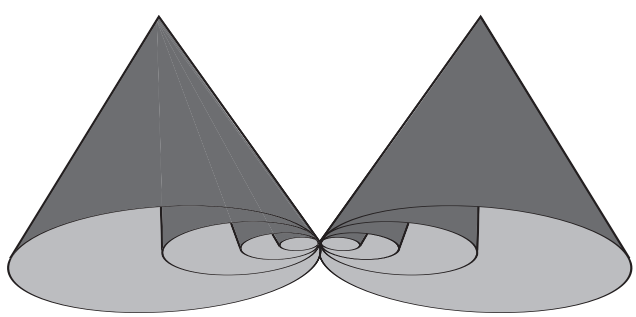


Figure 13.9: A wedge of two cones over Hawaiian earrings (note that there would actually be infinitely many little 'cones').

Van Kampen's Theorem takes a situation where a space is the union of two open subsets and shows how the fundamental groups of the two open subsets are combined to create the fundamental group of the whole space. The difference between the special case of Van Kampen's Theorem above and the general case is that in the general case the intersection of the two open sets need not be simply connected. How do non-trivial loops in the intersection influence the fundamental group of the whole space? Let's consider another special case of Van Kampen's Theorem where both of the open sets that make up the space are simply connected, but the intersection of the two open sets need not be simply connected. Think of an example of this phenomenon.

Theorem 13.45. *Let $X = U \cup V$ where U, V are open, path connected, and simply connected and $U \cap V$ is nonempty and path connected. Then X is simply connected.*

Let's consider another special case of Van Kampen's Theorem where only one of the two open sets that make up the space is simply connected, but the intersection of the two open sets is not simply connected. This case will help us understand what happens to loops in the intersection when viewed as elements of the fundamental group of the whole space.

Theorem 13.46. *Let $X = U \cup V$ where U, V are open and path connected and $U \cap V$ is path connected, $x \in U \cap V$, and $\pi_1(U, x) \cong 1$. Let $i : U \cap V \to V$ be the inclusion map. Then*

$$\pi_1(X, x) \cong \frac{\pi_1(V, x)}{N}$$

where N is the smallest normal subgroup of $\pi_1(V, x)$ containing the subgroup $i_*(\pi_1(U \cap V, x))$.

The above theorem captures the idea that loops in V that lie in $U \cap V$ become trivial, because loops in the intersection shrink in U. When $U \cap V$ had trivial fundamental group, the fundamental group of X was the free product of the fundamental groups of U and V. But when the fundamental group of $U \cap V$ is not trivial, the loops in the intersection of U and V can be considered to be elements of either the fundamental group of U or the fundamental group of V and so those two views of those loops in the intersection are the same in the fundamental group of the whole space. That insight is what Van Kampen's Theorem captures.

We are now ready to state Van Kampen's Theorem in its full generality.

Theorem 13.47 (Van Kampen's Theorem)**.** *Let $X = U \cup V$ where U, V are open and path connected and $U \cap V$ is path connected and $x \in U \cap V$. Let $i : U \cap V \to U$ and $j : U \cap V \to V$ be the inclusion maps. Then*

$$\pi_1(X, x) \cong \frac{\pi_1(U, x) * \pi_1(V, x)}{N}$$

where N is the smallest normal subgroup containing $\{i_*(\alpha) j_*(\alpha^{-1})\}_{\alpha \in \pi_1(U \cap V, x)}$ (so N contains elements created by taking a finite sequence of products and conjugates starting with elements of the form $i_*(\alpha) j_*(\alpha^{-1})$).

The proof of Van Kampen's Theorem is challenging; however, it is straightforward, meaning that if you follow clearly what needs to be proved, those steps lead to a proof. The strategy is to first notice that there is a natural map from the free product of $\pi_1(U)$ and $\pi_1(V)$ onto $\pi_1(X)$. What needs to be proved is that the kernel of that map is exactly the subgroup N of the free product that is described in the theorem. So two things need to be shown: An element in N is in the kernel and an element in the kernel is in N.

In proving Van Kampen's Theorem you will find yourself analyzing a homotopy between a loop and the constant loop. Figures 14.4 and 14.5 may be helpful as you create a proof.

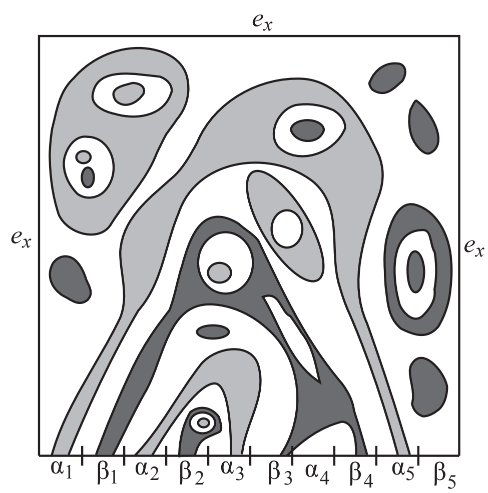


Figure 13.10: A covering of the regions mapped to $X \backslash U$ and those mapped to $X \backslash V$.

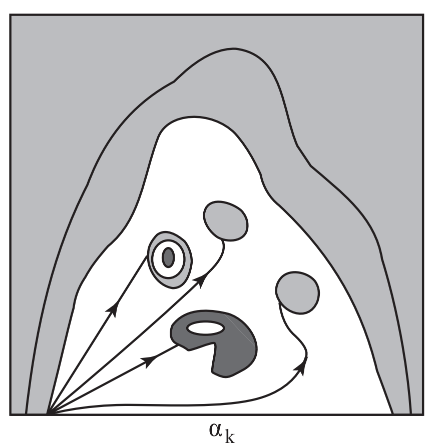


Figure 13.11: A possibly suggestive close up of a portion of a homotopy.

Van Kampen's Theorem can also be stated in the language of group presentations. Group presentations are useful, concrete ways to represent $\pi_1(X)$ (and any group), but they are not without problems as it is often extremely difficult to decide whether two different presentations describe the same group.

Theorem 13.48 (Van Kampen's Theorem; group presentations version)**.** *Let $X = U \cup V$ where U, V are open and path connected and $U \cap V$ is path connected and $x \in U \cap V$. Let $i : U \cap V \to U$ and $j : U \cap V \to V$ be the inclusion maps. Suppose $\pi_1(U, x) = \langle g_1, \ldots, g_n | r_1, \ldots, r_m \rangle$, $\pi_1(V, x) = \langle h_1, \ldots, h_t | s_1, \ldots, s_u \rangle$ and $\pi_1(U \cap V, x) = \langle k_1, \ldots, k_v | t_1, \ldots, t_w \rangle$ then*

$$\begin{aligned} \pi_1(X, x) &= \langle g_1, \ldots, g_n, h_1, \ldots, h_t \mid \quad r_1, \ldots, r_m, s_1, \ldots, s_u, \\ &\qquad\qquad\qquad\qquad\qquad\qquad i_*(k_1) = j_*(k_1), \ldots, i_*(k_v) = j_*(k_v) \rangle \, . \end{aligned}$$

Among many other things, Van Kampen's Theorem allows us to calculate the fundamental group of a polygonal presentation of a compact, connected 2-manifold (as long as we assume the vertices are all identified).

Exercise 13.49. *Let P be a polygonal representation of a compact, connected 2-manifold such that the all the vertices of P are identified in the corresponding quotient. Give a presentation for $\pi_1(P)$.*

In particular, we can now calculate the fundamental groups of all the connected, compact, triangulated 2-manifolds we saw in the previous chapter.

Exercise 13.50. *Give presentations of the fundamental groups for our canonical polygonal presentations of $\mathop{\#}\limits_{i=1}^{n} \mathbb{T}^2$ and $\mathop{\#}\limits_{i=1}^{n} \mathbb{R}\mathrm{P}^2$.*

By showing that the groups described above are all different, we can give a proof of the classification of 2-manifolds that does not use Euler Characteristic or orientability. Note: It may help to remember that if the abelianizations of two groups are different, then the groups must be different.

Theorem 13.51. *Each 2-manifold in the following infinite list is topologically different from all the others on the list: $\mathbb{S}^2$, $\mathop{\#}\limits_{i=1}^{n} \mathbb{R}\mathrm{P}^2$, and $\mathop{\#}\limits_{i=1}^{n} \mathbb{T}^2$.*

Theorem 13.52. *Suppose G is a finitely presented group. Then there exists a 2-complex (K, T) such that $\pi_1(K) \cong G$.*

In particular, every finite group is the fundamental group of a 2-complex. Hence if one could understand complexes sufficiently (in dimension 2 even), one could classify all finite groups, a major result (and instant Fields Medal).

The fundamental group of any finite complex is finitely generated, but there are topological spaces whose fundament groups are more elaborate. The Hawaiian earring is one.

Theorem 13.53. *The fundamental group of the Hawaiian earring is not finitely generated. In fact, it is not countably generated.*

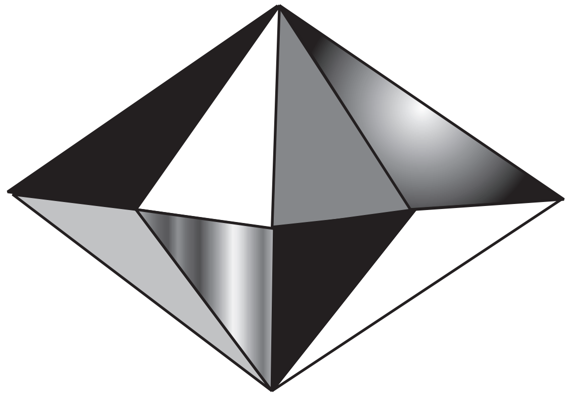


Figure 13.12: A lens space as a quotient of a lens.

13.5 Lens Spaces

As we will have seen, the topology of 2-manifolds is very well-understood. On the other hand, the study of 3-manifolds is still a very active field. In this section we will study a relatively simple (and yet still interesting) class of 3-manifolds called **lens spaces**. These 3-manifolds are well-understood and completely classified.

Let p and q be relatively prime integers. A (p, q)-lens space, denoted $L(p, q)$, can be defined in several different ways. It was first defined as an identification space that started with a 3-ball drawn in the shape of a lens (hence the name). The top and bottom hemispheres of this lens are each divided into p triangle-like wedges. Each triangle from the top hemisphere is identified with a triangle in the bottom hemisphere that is a certain specified number (relatively prime to p) of triangles around the equator. (See Figure 13.12).

We, however, will give a different formulation which will take a few steps to develop. Recall now that we have three ways of viewing the torus. The torus is $\mathbb{S}^1 \times \mathbb{S}^1$, it is the surface of a donut, and, via the process we carried out in Figure 12.18, it is a rectangle in the plane with points identified pairwise. Notice, however, the the 'surface of a donut' could also be described in the following way. Imagine a unit circle in 3-space that lies in the yz-plane and whose center lies at the point $(0, 2, 0)$ (so that a portion of the y-axis forms a diameter of the circle and so that the origin is not contained on the circle or on the flat disk it would bound). We could form the surface of a donut by rotating this circle around another circle that lies in the xy plane, is centered at the origin, and contains the center of our first circle. Notice that if we had imagined our original circle to be 'filled in,' we would instead get a solid torus.

In this way, we can view the torus as a product of our original circle with a larger circle. We will now use this viewpoint to pick out two very important type of curves on the torus. Suppose that we pick a point on our original circle and trace its path as the circle was rotated. We see that we get a circle on the torus. This type of curve is called a **longitude** or **longitudinal** curve. Likewise, a curve that travels around a copy of the original circle cross a point of the larger circle is called a **meridian**. Notice that if our torus were a solid torus, a meridian would form the boundary of a disk in the 'inside' of the torus.

Both a longitude and a meridian go 'around' the torus, but they go around in very different ways (this distinction will be made more precise when we study the fundamental group, particularly that of a solid torus). It is also important to realize that neither curve is a homeomorphic invariant of the torus. In fact, we will see that there is a homeomorphism of the torus that takes one to the other (as well as homeomorphisms that take them to many different curves). Hence we need to always have a particular manifestation of the torus in mind when we discuss one of these curves.

Now, if we consider Figure 12.18, we can get yet another description of these curves. Notice that under the process shown, a straight-line path that goes up the rectangle corresponds to a meridian. A path that goes down the rectangle also corresponds to a meridian, but traverses in the opposite direction. Likewise, a path that goes across the rectangle corresponds to a longitude. Given this rectangle, we'll take the path given by traversing the top (and so bottom) edge to be the canonical choice for a longitude and the path given by traversing the right edge upwards as our choice for a meridian. Notice that the path we get on the surface of the donut depends on the process used to identify it with the rectangle. Hence, we should always have a fixed process in mind (such as the one in Figure 12.18).

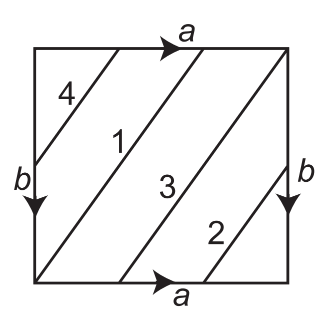


Figure 13.13: This path goes up and to the right and traverses the diagram in the order of the numbers indicated. It travels around 3 times vertically (in the meridian direction on the torus) and 2 times horizontally (in the longitudinal direction on the torus).

Now suppose that $p, q \in \mathbb{N}$ are relatively prime. We can use our second interpretation to define a curve on this torus that 'goes around' the torus p times in the sense of a meridian and q times longitudinally. To accomplish this and to keep things simple, consider the torus as a square with edges identified (rather then the rectangle we were picturing before: apply the same cutting process as in Figure 12.18, but 'stretch' the end rectangle vertically until it is a square). Imagine a path that begins at the lower left-hand vertex of the square. From this point, draw a line segment of slope $\frac{q}{p}$ until it hits an edge of the square. If the line segment hits the top edge of square, move down to the point that corresponds on the bottom edge of the square and draw a line of the same slope until it again hits an edge. Likewise if the line segment hits the right hand side move to the left. Proceed with this process until the line hits the upper-right vertex.

This process will give a curve on the torus which essentially goes around q times in the meridian direction and p times longitudinally. The 'line' we get for $p = 2$ and $q = 3$ is shown in Figure 13.5.

Lemma 13.54. *If $p, q \in \mathbb{N}$ are relatively prime, the line described in the above process will eventually intersect the upper-right vertex of the square. Moreover, the line will not intersect itself until it does.*

We can allow p or q to be negative by again taking our slope to be $\frac{q}{p}$ but moving up along it if p is positive and down along it if q is negative. If we are, for example, moving down and to the right, we will need to begin at the upper-left vertex. Since all the vertices are really identified on the torus, this is not a problem. We can also allow one of p or q to be zero by stipulating that the other must be one (i.e., 0 is not relatively prime to any integer other than ± 1) and taking the appropriately traversed vertical or horizontal line on an edge.

Lemma 13.55. *Let p and q be relatively prime integers and let $\rho_{p,q}$ be the simple closed curve constructed above. Then there is a homeomorphism of the square (with the standard identifications made for the torus) that takes $\rho_{p,q}$ to our canonical meridian.*

We can now give the definition of a lens space.

Definition. Fix relatively prime integers p and q. Let V_1 be a solid torus (manifested as solid donut to be precise) and let V_2 be an additional (disjoint) copy. Let $h : \mathrm{Bd}(V_1) \to \mathrm{Bd}(V_2)$ be the homeomorphism defined in the previous lemma. Then the quotient space $V_1 \cup_h V_2$ is the (p, q)-*lens space*, denoted $L(p, q)$.

In addition to the two we have mentioned, Lens spaces can be defined other ways. For example, they can be viewed as quotient spaces of $\mathbb{S}^3$ under certain group actions. Beware that some authors use $L(p, q)$ to mean $L(q, p)$.

Theorem 13.56. *For $p, q \in \mathbb{Z}$ relatively prime, the lens space $L(p, q)$ is triangulable.*

Exercise 13.57. *Let $p, q \in \mathbb{Z}$ be relatively prime. Calculate the fundamental group of the Lens space $L(p, q)$.*

13.6 Knot Complements

To introduce knot theory, we begin with a special type of homotopy called *isotopy*.

Definition. Suppose that X and Y are topological spaces and $f, g : X \to Y$ are topological embeddings. An **isotopy** from f to g is a function $H : X \times [0,1] \to Y$ such that $H(\cdot, 0) = f$, $H(\cdot, 1) = g$, and $H(\cdot, t)$ is a topological embedding for each t. If such a function exists, we say that f and g are isotopic.

An embedding of $\mathbb{S}^1$ in $\mathbb{S}^3$ is called a knot. If we ignore the point at infinity, such an embedding corresponds to the intuitive idea of a closed knot made of string or rope. We will also assume that our knots are 'tame' in the sense that they can be thickened to an embedding of a solid torus (which we will assume is PL for regularity). The study of these types of knots is an active branch of mathematics. It has major applications to physics and to the study of 3-manifolds.

Definition. Let $i : \mathbb{D}^2 \times \mathbb{S}^1 \to \mathbb{S}^3$ be a PL embedding, that is, an embedding that is simplicial with respect to some triangulation of $\mathbb{D}^2 \times \mathbb{S}^1$ and some triangulation of $\mathbb{S}^3$. Then $K = i|_{\{0\} \times \mathbb{S}^1}$ is a **knot** in $\mathbb{S}^3$. Let $N(K)$ be the image of the interior (in the manifold with boundary sense) of $D^2 \times \mathbb{S}^1$, a neighborhood of $i(\{0\} \times \mathbb{S}^1)$. The **knot complement** or **knot exterior** of K is $M_K = \mathbb{S}^3 - N(K)$. Note that M_K is a 3-manifold with boundary whose boundary is homeomorphic to the 2-torus.

The knot complement is an important characteristic of the knot. Cameron Gordon and John Luecke proved that at most two knots (a knot and its mirror image) can have the same knot complement up to homeomorphism. Furthermore, knot complements give an important method for generating compact connected 3-manifolds: namely gluing together knot complements along their boundary.

In this section we will study the fundamental groups of knot complements. If we look at a knot from above, we see a curve with crossings where it goes over or under itself. For example, Figure 13.14 shows the *trefoil knot*.

If we are given a picture of a projection of a knot K into $\mathbb{R}^2$ where gaps indicate where crossings occur and where all crossings are transverse crossings of two arcs, then we can use the pictures, along with Van Kampen's Theorem to produce a presentation of $\pi_1(M_K)$. This group is often called the fundamental group of the knot, even though it is actually the fundamental group of the complement of the knot.

Roughly speaking, each arc in the picture gives a generator and each crossing represents a relation. For each arc in a knot projection, draw a labeled perpendicular arrow as shown in Figure 13.15):

The arrow a_i, for example, represents the loop in M_K obtained by starting well above the knot

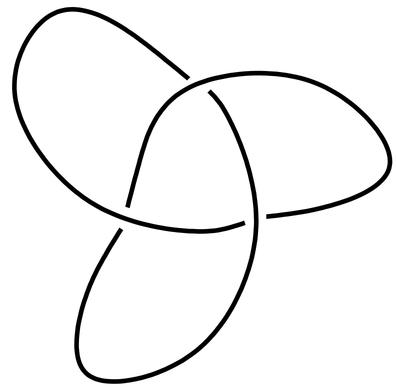


Figure 13.14: The trefoil knot.

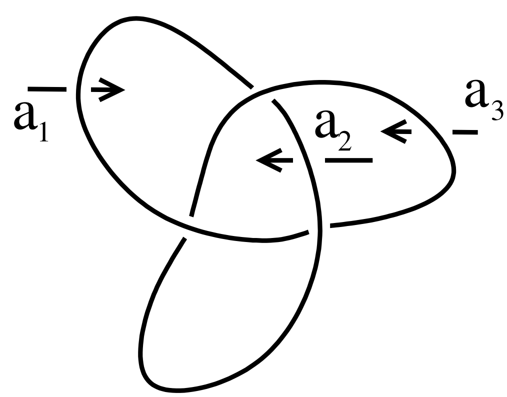


Figure 13.15: The arrows for the arcs of a trefoil knot.

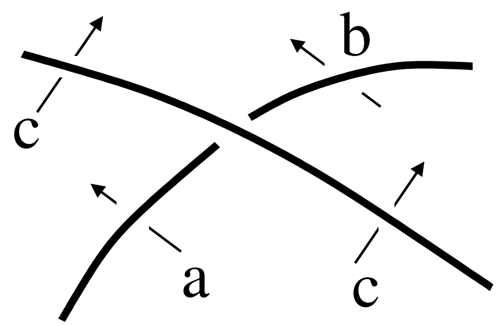


Figure 13.16: The arrows around a crossing.

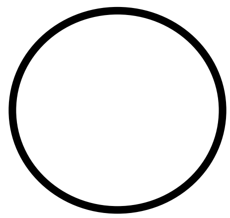


Figure 13.17: The unknot.

(at the base point chosen for $\pi_1(M_K)$), going straight down to the tail of a_i, then going along a_i under the knot, and finally returning to the starting point going straight from the head of a_i.

Lemma 13.58. *Every loop in M_K is homotopic in M_K to a product of a_i's. In other words, the loops $\{a_i\}$ generate $\pi_1(M_K)$.*

Lemma 13.59. *At every crossing, such as that illustrated in Figure 13.16, the following relation holds: $acb^{-1} = c$ or $acb^{-1}c^{-1} = 1$.*

Theorem 13.60. *Let K be a knot in $\mathbb{S}^3$ and let $\{a_i\}$ be the set of loops consisting of one loop for each arc in a knot projection of K as described above. Then $\pi_1(M_K) = \{a_1, a_2, \ldots, a_n | a_i a_j a_k^{-1} a_j^{-1}$ where there is one relation of the form $a_i a_j a_k^{-1} a_j^{-1}$ for each crossing in the knot projection$\}$.*

Exercise 13.61. *Find the fundamental group of the complement of the unknot (See Figure 13.17).*

Exercise 13.62. *Find the fundamental group of the complement of the trefoil knot.*

Exercise 13.63. *Find the fundamental group of the complement of the figure-8 knot, shown in Figure 13.18.*

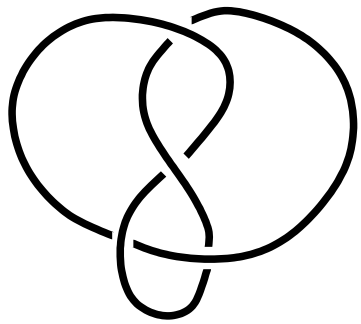


Figure 13.18: The figure-8 knot.

13.7 Higher Homotopy Groups

In this section, we will briefly describe a way to generalize the fundamental group by constructing the *higher homotopy groups*. Recall that a loop in X based at $x_0 \in X$ is a map $[0,1] = \mathbb{D}^1 \to X$ where both 0 and 1 are sent to $x_0 \in X$. In other words, it is a map from $\mathbb{D}^1$ which maps the boundary of $\mathbb{D}^1$ to $x_0 \in X$. This leads to the following natural generalization.

Definition. Let X be a topological space and $x_0 \in X$. Let $f, g : (\mathbb{D}^n, \partial\mathbb{D}^n) \to (X, x_0)$ be continuous (that is, let $f, g : \mathbb{D}^n \to X$ be continuous maps that take $\partial\mathbb{D}^n$ to x_0). If $[f]$ and $[g]$ denote the homotopy classes of these maps relative to $\partial\mathbb{D}^n$, then we define the product $[f] \cdot [g]$ to be the homotopy class of:

$$f \cdot g(x_1, x_2, \ldots, x_n) = \begin{cases} \alpha(2x_1, x_2, \ldots, x_n), & 0 \leq x_1 \leq \frac{1}{2} \\ \beta(2x_1 - 1, x_2, \ldots, x_n), & \frac{1}{2} < x_2 \leq 1 \end{cases}$$

where $(x_1, x_2, \ldots, x_n) \in \mathbb{D}^n$.

Maps of this type can be considered to be the higher dimension analogues of loops. They can also be viewed as maps from $\mathbb{S}^n$ which map a certain fixed point in $\mathbb{S}^n$ to x_0.

Exercise 13.64. *The collection of homotopy classes of continuous maps of the type $f : (\mathbb{D}^n, \partial\mathbb{D}^n) \to (X, x_0)$, with the product defined above, forms a group.*

Definition. The above mentioned group is called the n^{th} **homotopy group of X based at x_0** and is denoted $\pi_n(X, x_0)$. The point x_0 is called the **base point** of the homotopy group.

Theorem 13.65. *Homotopy equivalent spaces have the same homotopy groups.*

Homotopy groups are generally hard to compute (even for $\mathbb{S}^n$). In later chapters we will develop the study of homology groups, which are easier to compute. Homology groups are generally more useful than the higher homotopy groups in distinguishing higher dimensional topological spaces from one another, so we will turn our attention now to covering spaces.

13.8 The Fundamental Group—Not Such a Loopy, Loopy Idea

Holes in spaces are often easy to see intuitively. When we are confronted with an obvious piece of reality, one of the great ways to create ideas is to take the trouble to pin down that intuition and turn an impression into a concrete concept. The fundamental group arose from specifying what we mean by a hole in a space.

In order to pin down the intuition of surrounding a hole with a loop, we were required to come up with the idea of a homotopy in order to clarify what it means for two loops to go around these holes in the 'same' manner. We have seen already and will see much more in the chapters ahead how valuable the concept of homotopy really is.

Surely one of the great satisfactions in mathematics is to follow the thread of mathematical creation from an intuitive impression to a whole edifice of mathematical richness and insight. The fundamental group presents us with a great example of that journey.

Chapter 14

Covering Spaces: Layering It On

Suppose you are given an open set U in the plane whose total area is strictly more than 1. It can even be disconnected, maybe in a million little pieces. Here's a puzzle: show that U can be translated, without rotation, so that two different lattice points of $\mathbb{R}^2$ lie in the translated set. (Lattice points are points (a, b) in $\mathbb{R}^2$ in which both coordinates are integers.)

This puzzle has a clever resolution. Think about rolling the plane in the horizontal direction onto a cylinder, and then taking the tube-like cylinder and wrapping it along its infinite length onto a torus. This produces a map of the plane onto the torus, and it can be done so that when completed, lattice points will all be on top of each other. Since the area of the open set U is greater than 1, its image on the torus must overlap using some pair of points in the plane. Can you see how to use this pair to produce a translation that moves both points of the pair to lattice points?

This technique of finding a 'covering' map from the plane onto the torus that wraps nicely many times (in this case, infinitely many times) highlights a relationship between the plane and the torus that leads to the concept of a covering space. Certain paths in the plane correspond to loops in the torus. So there is a strong connection between a covering space and the fundamental group of a given space. Recall our proof that $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$. We 'covered' $\mathbb{S}^1$ with the real line using various wrapping maps. In this chapter, we will generalize that idea. Just as the wrapping map was useful for computing the fundamental group of the circle, more general covering spaces are useful for understanding the fundamental groups of other spaces.

We have seen a couple of ways in which the fundamental group allows us to use group theoretic results to answer important questions about topology. Covering spaces help us understand topological spaces, but they also can help us to prove results about the structure of certain groups.

14.1 Basic Results and Examples

Effective Thinking Principle. *Think Backwards.* Think about where things come from or could be constructed from.

Definition. Let X and $\widetilde{X}$ be connected, locally path connected spaces and let $p : \widetilde{X} \to X$ be a continuous function. The pair $(\widetilde{X}, p)$ is a **covering space of** X if and only if for each $x \in X$ there exists a neighborhood U of x such that p restricted to each component of $p^{-1}(U)$ is a homeomorphism onto U. When $(\widetilde{X}, p)$ is a covering space of X, the space $\widetilde{X}$ is a **cover** of X and p is a **covering map**. The space X is called the **base space**.

Example. Let $X = \mathbb{S}^1$, $\widetilde{X} = \mathbb{R}^1$, and $p : \mathbb{R}^1 \to \mathbb{S}^1$ be defined by $p(t) = (\cos t, \sin t)$. Then $(\mathbb{R}^1, p)$ is a covering space of $\mathbb{S}^1$.

Example. Let $X = \mathbb{S}^1$, $\widetilde{X} = \mathbb{S}^1$, and $p : \mathbb{S}^1 \to \mathbb{S}^1$ be defined by $p(z) = z^n$, where we view $\mathbb{S}^1$ as a subset of $\mathbb{C}$ in the usual way and use complex multiplication. Then $(\mathbb{S}^1, p)$ is a covering space of $\mathbb{S}^1$.

Example. If X is the figure eight, and $\widetilde{X}$ is any of the spaces shown in Figure 14.1, with the corresponding map $p : \widetilde{X} \to X$, then $(\widetilde{X}, p)$ is a covering space of X. Note that the third choice for $\widetilde{X}$ in the figure continues the pattern infinitely in the obvious way; that is, the picture just shows part of the actual covering space, which has infinitely many segments that map around each of the two circles in the figure eight.

Theorem 14.1. *Let $(\widetilde{X}, p)$ be a covering space of X. If $x, y \in X$, then $|p^{-1}(x)| = |p^{-1}(y)|$.*

Definition. If $(\widetilde{X}, p)$ is a covering space of a space X and $n = |p^{-1}(x)|$ for some $x \in X$, then $(\widetilde{X}, p)$ is called an **n-fold covering** of X and $\widetilde{X}$ is a cover of **degree** n. Note that n is allowed to be an infinite cardinal number.

Example. $p : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $p(z) = z^n$ (as above) gives an n-fold covering of $\mathbb{S}^1$ by itself.

Exercise 14.2.

1. *Describe two non-homeomorphic 2-fold covers of the Klein bottle.*

2. *Describe all non-homeomorphic 2-fold covers of the figure eight.*

3. *Describe all non-homeomorphic 3-fold covers of the figure eight.*

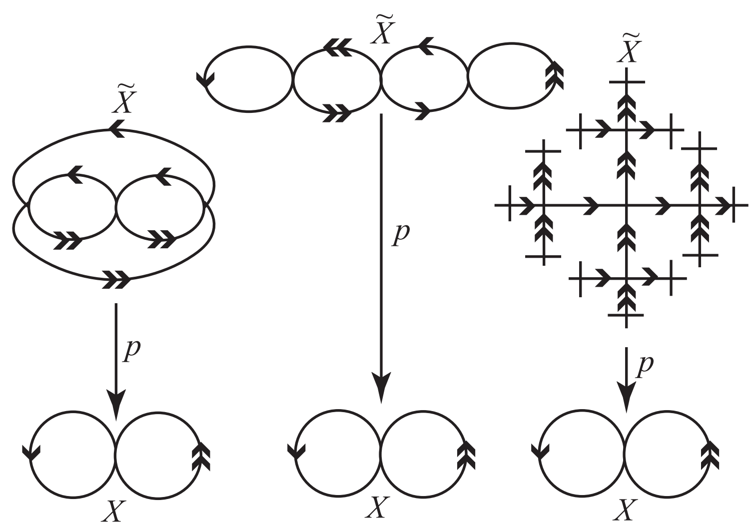


Figure 14.1: Several coverings of the figure eight. The single arrow segments or loops in $\widetilde{X}$ are mapped to the single arrow loop in X. Similarly for the double arrow segments or loops.

14.2 Lifts

Now that we have established the basic properties of covering spaces, we consider their interactions with continuous functions. Specifically, we want to study the correspondence between continuous functions into the base space and continuous functions into the covering space. Certainly, if we have a map into the covering space, we can compose with the covering map to get a map into the base space, but when can we go the other direction?

Definition. Given a covering space $(\widetilde{X}, p)$ of X and a continuous function $f : Y \to X$, then a continuous function $\widetilde{f} : Y \to \widetilde{X}$ is called a **lift of** f if and only if $p \circ \widetilde{f} = f$, that is, if the following diagram commutes.

$$\begin{tikzcd} & \widetilde{X} \arrow[d, "p"] \\ Y \arrow[ur, "\tilde{f}"] \arrow[r, "f"'] & X \end{tikzcd}$$

Theorem 14.3. *Let $(\mathbb{R}^1, \omega)$ be the standard wrapping map covering of $\mathbb{S}^1$. Then any path $f : [0,1] \to \mathbb{S}^1$ has a lift $\widetilde{f} : [0,1] \to \mathbb{R}^1$.*

The next result shows that lifts satisfy a certain type of uniqueness.

Theorem 14.4. *If $(\widetilde{X}, p)$ is a cover of X, Y is connected, and $\widetilde{f}, \widetilde{g} : Y \to \widetilde{X}$ are continuous functions such that $p \circ \widetilde{f} = p \circ \widetilde{g}$, then $\{y \mid \widetilde{f}(y) = \widetilde{g}(y)\}$ is empty or all of Y.*

We will often study lifts of paths.

Theorem 14.5. *Let $(\widetilde{X}, p)$ be a cover of X and let f be a path in X. Then for each $x_0 \in \widetilde{X}$ such that $p(x_0) = f(0)$, there exists a unique lift $\widetilde{f}$ of f satisfying $\widetilde{f}(0) = x_0$.*

> **Effective Thinking Principle.** *Explore Simple Cases.* Simple cases often lead to the most important insights.

Exercise 14.6. *Let p be a k-fold covering of $\mathbb{S}^1$ by itself and α a loop in $\mathbb{S}^1$ which when lifted to $\mathbb{R}^1$ by the standard lift satisfies $\widetilde{\alpha}(0) = 0$ and $\widetilde{\alpha}(1) = n$. What are the conditions on n under which α will lift to a loop?*

The next result will provide the machinery we need to study the relationship between the fundamental group of the base space and the fundamental group of the cover.

Theorem 14.7 (Homotopy Lifting Lemma)**.** *Let $(\widetilde{X}, p)$ be a cover of X and α, β be two paths in X. If $\widetilde{\alpha}$, $\widetilde{\beta}$ are lifts of α, β satisfying $\widetilde{\alpha}(0) = \widetilde{\beta}(0)$, then $\widetilde{\alpha} \sim \widetilde{\beta}$ if and only if $\alpha \sim \beta$.*

Theorem 14.8. *If $(\widetilde{X}, p)$ is a cover of X, then p_* is a monomorphism (i.e. an injective homomorphism) from $\pi_1(\widetilde{X})$ into $\pi_1(X)$.*

The previous theorem implies that the fundamental group of a cover of X is isomorphic to a subgroup of the fundamental group of the space X.

Theorem 14.9. *Let $(\widetilde{X}, p)$ be a cover of X, α a loop in X, and $\widetilde{x}_0 \in \widetilde{X}$ such that $p(\widetilde{x}_0) = \alpha(0)$. Then α lifts to a loop based at $\widetilde{x}_0$ if and only if $[\alpha] \in p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$.*

Exercise 14.10. *Recast a proof of the fact that $\pi_1(\mathbb{S}^1) \cong \mathbb{Z}$ using the language of covering spaces.*

Suppose p is a k-fold cover of $\mathbb{S}^1$ by $\mathbb{S}^1$. We can figure out what maps from $\mathbb{S}^1$ into the base space would lift. Looking at the specific example suggests criteria for deciding when a map will lift.

> **Effective Thinking Principle.** *Iconic Cases Suggest General Truths.* Examples with the clearest salient features of a phenomenon or concept often expose general truths.

Theorem 14.11. *Let $(\widetilde{X}, p)$ be a covering space of X and let $x_0 \in X$. Fix $\widetilde{x}_0 \in p^{-1}(x_0)$. Then a subgroup H of $\pi_1(X, x_0)$ is in $\{p_*(\pi_1(\widetilde{X}, \widetilde{x}))\}_{p(\widetilde{x})=x_0}$ if and only if H is a conjugate of $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$.*

Theorem 14.12. *Let $(\widetilde{X}, p)$ be a covering space of X. Choose $x \in X$, then $|p^{-1}(x)| = [\pi_1(X) : p_*(\pi_1(\widetilde{X}))]$, where the equation has the obvious interpretation if either side is infinite.*

In particular, the index of the subgroup of $\pi_1(X)$ corresponding to a finite covering $(\widetilde{X}, p)$ equals the degree of the covering.

Exercise 14.13. *Give a covering space of $\mathbb{S}^1$ that corresponds to a subgroup of index 3. If p is the covering map, describe p_*.*

Theorem 14.14. *Let $(\widetilde{X}, p)$ be a covering space of X and $\widetilde{x}_0 \in \widetilde{X}$, $x_0 \in X$ with $p(\widetilde{x}_0) = x_0$. Also let $f : Y \to X$ be continuous where Y is connected and locally path connected and $y_0 \in Y$ such that $f(y_0) = x_0$. Then there is a lift $\widetilde{f} : Y \to \widetilde{X}$ such that $p \circ \widetilde{f} = f$ and $f(y_0) = \widetilde{x}_0$ if and only if $f_*(\pi_1(Y, y_0)) \subseteq p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$. Furthermore, $\widetilde{f}$ is unique.*

Exercise 14.15. *Let $X = \mathbb{S}^1$, $\widetilde{X} = \mathbb{R}$, $(\widetilde{X}, \omega)$ be the covering space of X given by the standard wrapping map, and Y as in Figure 14.2. When does a map $f : Y \to X$ not have a lift? Why is this example here?*

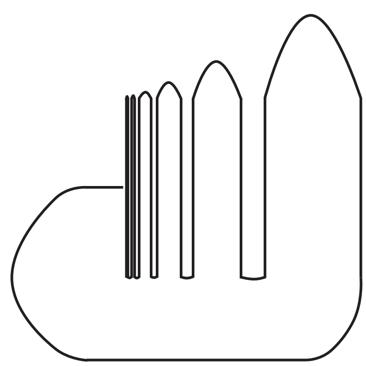


Figure 14.2: Consider this picture.

Exercise 14.16. *Show that $\pi_2(\mathbb{T}^2) = 0$, i.e., every map of a sphere $\mathbb{S}^2$ into $\mathbb{T}^2$ is null homotopic.*

14.3 Regular Covers and Cover Isomorphism

Effective Thinking Principle. *Define Equivalence.* After creating a concept, investigate equivalence with respect to that concept.

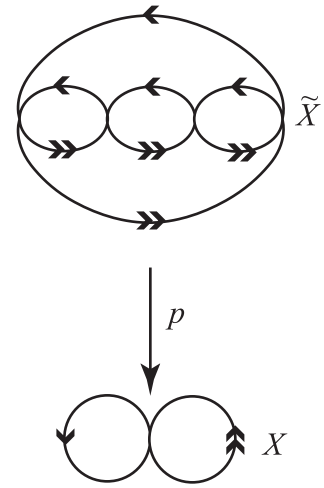


Figure 14.3: A covering of the figure eight.

As with almost any structure in mathematics, we will want a notion of two covering spaces being equivalent. Certainly we will want the covers to be homeomorphic as topological spaces. In addition, this homeomorphism should respect the covering map.

Definition. Let $(\widetilde{X}_1, p_1)$ and $(\widetilde{X}_2, p_2)$ be covering spaces of X. Then a map $f : \widetilde{X}_1 \to \widetilde{X}_2$ such that f is a homeomorphism and $p_2 \circ f = p_1$ is called a *cover isomorphism*.

As we might expect, our notion of equivalence behaves well with respect to the relationship we have established between covering spaces and subgroups of the fundamental group.

Theorem 14.17. *Let $(\widetilde{X}_1, p_1)$ and $(\widetilde{X}_2, p_2)$ be covering spaces of X. Let $\widetilde{x}_1 \in \widetilde{X}_1$ and $\widetilde{x}_2 \in \widetilde{X}_2$ such that $p_1(\widetilde{x}_1) = p_2(\widetilde{x}_2)$. Then there is a cover isomorphism $f : \widetilde{X}_1 \to \widetilde{X}_2$ with $f(\widetilde{x}_1) = \widetilde{x}_2$ if and only if $p_*(\pi_1(\widetilde{X}_1, \widetilde{x}_1)) = p_*(\pi_1(\widetilde{X}_2, \widetilde{x}_2))$.*

Definition. Let $(\widetilde{X}, p)$ be a covering space. Then a cover isomorphism from $\widetilde{X}$ to itself is called a **covering transformation** or a **deck transformation**. The set of covering transformations, denoted $\mathcal{C}(\widetilde{X}, p)$ is a group under composition.

Exercise 14.18. *What is $\mathcal{C}(\widetilde{X}, p)$ for the covering space of the figure eight shown in Figure 14.3?*

Theorem 14.19. *If $(\widetilde{X}, p)$ is a covering space of X and $f \in \mathcal{C}(\widetilde{X}, p)$, then $f = Id_{\widetilde{X}}$ if and only if f has a fixed point.*

As far as group structure is concerned, we know that normal subgroups are an important type of subgroup. We will therefore want to give special attention to the covering spaces whose fundamental groups correspond to normal subgroups of the fundamental group of the base space.

Definition. Let $(\widetilde{X}, p)$ be a covering space of X. If $p_*(\pi_1(\widetilde{X})) \triangleleft \pi_1(X)$, then $(\widetilde{X}, p)$ is a **regular** covering space.

Exercise 14.20. *Consider the second three-fold covering space of the figure eight discussed in Exercise 14.18. Find an element of $p_*(\pi_1(\widetilde{X}))$ which, when conjugated, is not in $p_*(\pi_1(\widetilde{X}))$. Conclude that the covering space is not regular.*

Theorem 14.21. *If $(\widetilde{X}, p)$ is a regular covering space of X and $x_1, x_2 \in \widetilde{X}$ such that $p(x_1) = p(x_2)$, then there exists a unique $h \in \mathcal{C}(\widetilde{X}, p)$ such that $h(x_1) = x_2$.*

The preceding theorem tells us that for a regular covering space, there is a (unique) covering transformation carrying any point in the set $p^{-1}(x)$ to any other point in the same set.

Exercise 14.22. *Do such covering transformations necessarily exist in irregular covering spaces?*

Theorem 14.23. *A covering space is regular if and only if for every loop in the base space either all its lifts are loops or all its lifts are paths that are not loops.*

Exercise 14.24. *Find a covering space $p : \widetilde{X} \to X$ and generators $e_1, \ldots, e_n$ of $\pi_1(X)$ such that each e_i satisfies the criteria of the previous theorem but the cover is not regular.*

Exercise 14.25.

1. *Describe all regular 3-fold covering spaces of the figure eight.*
2. *Describe all irregular 3-fold covering spaces of the figure eight.*
3. *Describe all regular 3-fold covering spaces of the bouquet of 3 circles.*

There is an important correspondence between the covering transformations of regular covers of X and the normal subgroups of $\pi_1(X)$.

Theorem 14.26. *Let $(\widetilde{X}, p)$ be a regular covering space of X. Then $\mathcal{C}(\widetilde{X}, p) \cong \pi_1(X)/p_*(\pi_1(\widetilde{X}))$. In particular, $\mathcal{C}(\widetilde{X}, p) \cong \pi_1(X)$ if $\widetilde{X}$ is simply connected.*

Exercise 14.27. *Observe that the standard wrapping map is a regular covering map of $\mathbb{S}^1$ by $\mathbb{R}^1$. Describe the covering transformations for this covering space. Describe the covering map that maps $\mathbb{R}^2$ to the torus $\mathbb{T}^2$ and describe the covering transformations for this covering space.*

14.4 The Subgroup Correspondence

In this section, we prove a substantial and beautiful result about the relationship between covering spaces and the fundamental group of a space.The following strange-looking, but relatively mild criterion is involved in the theorem statement.

Definition. A space X is called **semi-locally simply connected** if and only if every $x \in X$ is contained in an open set U such that every loop in U based at x is homotopically trivial in X.

Note that the open set U in the previous definition need not be simply connected itself (though being simply connected would certainly suffice). We require every loop in U to be homotopic to a constant, but the homotopy is allowed to go outside of U (and into X).

Theorem 14.28 (Existence of covering spaces)**.** *Let X be connected, locally path connected, and semi-locally simply connected. Then for every $G < \pi_1(X, x_0)$ there is a covering space $(\widetilde{X}, p)$ of X and $\widetilde{x}_0 \in \widetilde{X}$ such that $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0)) = G$. Furthermore, $(\widetilde{X}, p)$ is unique up to isomorphism.*

In summary, we have proved the following.

Corollary 14.29. *Let X be connected, locally path connected, and semi-locally simply connected. Then there is a one-to-one correspondence between the subgroups of $\pi_1(X)$ and the collection of isomorphism classes of covering spaces of X where the covering space $p : \tilde{X} \to X$ corresponds to $p_*(\pi_1(\widetilde{X}))$.*

Definition. A connected, locally path connected cover is called **universal** if and only if its fundamental group is trivial.

Corollary 14.30. *Every connected, locally path connected, semi-locally simply connected space admits a unique universal covering space.*

> **Effective Thinking Principle.** *Apply Insights to Familiar Objects.* Seeing implications of new insights to familiar objects enriches understanding of both the insight and the familiar world.

Exercise 14.31. *Find a universal cover $\widetilde{X}$ for each of the Klein bottle, the torus, and the projective plane. In each case, show explicitly that $\mathcal{C}(\widetilde{X}, p) \cong \pi_1(X)$.*

14.5 Theorems about Free Groups

We have pointed out previously that the topology of a space can be used to prove results about the structure of its fundamental group. Later in this section, we will give an important example of this technique.

Theorem 14.32. *A finite tree is contractible.*

Theorem 14.33. *Let G be a finite graph, and T be a maximal tree in G. Then if $\{e_1, \dots e_n\}$ is the set of edges that are not in T, $\pi_1(G) = F_n$, the free group on n generators; and there is a system of generators that are in one-to-one correspondence with the edges $\{e_1, \dots e_n\}$.*

Lemma 14.34. *Let X be the bouquet of n circles. Every finite cover of X is homeomorphic to a finite graph.*

Theorem 14.35. *Let F_n be the free group on n letters. Then every subgroup of F_n of finite index is isomorphic to a free group on finitely many letters.*

By allowing ourselves to use graphs that may not be finite, we can substantially improve the previous result. The next result shows that much of our work can be assumed to take place in a finite graph.

Lemma 14.36. *Suppose that G is a graph and that $K \subset G$ is compact. Then K is contained in a finite subgraph of G.*

Theorem 14.37. *Every tree is simply connected.*

Theorem 14.38. *Let G be a graph, then $\pi_1(G)$ is free.*

Exercise 14.39. *Show that the free group of rank 2 has finite index subgroups that are isomorphic to free groups of arbitrarily large rank.*

Lemma 14.40. *Let X be the bouquet of n circles. Then every cover of X is homeomorphic to a graph.*

The following improvement is a substantial result in group theory. There are proofs which do not rely on topology, but they are considerably longer and more difficult.

Corollary 14.41 (Nielsen-Schreier Theorem)**.** *Let F_n be the free group on n letters. A subgroup of F_n is always free.*

The Nielsen-Schreier Theorem is not as trivial as it might appear. For example, consider $F_2 = \langle x, y \rangle$ and set

$$a = x^2y^3,\; b = y^{-3}x^2y^{-1}x^{-1}y^{-3}x^{-2}y^{-1}x^{-1},\text{ and } c = xy.$$

Let H be the subgroup generated by a, b, c. Then H is *not* isomorphic to the free group on $\{a, b, c\}$ because we have the relation $abcac = 1$. The theorem says that there *exists* some subset of H that generates H and such that the elements have no nontrivial relations among them.

To avoid (even more) technicalities, we have restricted ourselves to free groups on finitely many letters, but the same techniques can be used to prove the result for a more general free group.

Exercise 14.42. *Describe a regular k-fold cover $\widetilde{X}$ of a bouquet of n-circles. What (in terms of k and n) is the rank of the free group $\pi_1(\widetilde{X})$? What does this insight tell us about the normal subgroups of finite index of the free group on n letters?*

Exercise 14.43.

1. *Let F be a free group on n letters. Let $G < F$ be of finite index k and contain 7 free generators. What can the value of n be?*

2. *Let F be a free group on n letters. Let $G < F$ be of finite index k and contain 4 free generators. What can the value of n be?*

3. *Let F be a free group on n letters. Let $G < F$ be of finite index k and contain 24 free generators. What can the value of n be?*

14.6 Covering Spaces and 2-Manifolds

Since covering spaces are locally homeomorphic to the spaces they cover, the covering spaces of 2-manifolds are 2-manifolds. Investigating the relationships between 2-manifolds and covering spaces is especially satisfying.

Theorem 14.44. *Let F be a 2-manifold and $(\widetilde{F}, p)$ be a covering space of F. Then $\widetilde{F}$ is a 2-manifold.*

The properties of being an n-fold cover allow us to describe the relationship between the Euler characteristics of a covering space and its base space. And orientability of the base space implies orientability of the covering space.

Theorem 14.45. *Let F be a compact connected surface and $p_n : \widetilde{F} \to F$ be an n-fold covering of F (for $n < \infty$). Then $\widetilde{F}$ is a compact surface and $\chi(\widetilde{F}) = n\chi(F)$. Moreover, if F is orientable, then $\widetilde{F}$ is as well.*

Exercise 14.46.

1. *Describe all non-homeomorphic 3-fold covers of the Klein bottle.*

2. *Describe all non-homeomorphic 2-fold covers of* $\mathbb{T}^2 \# \mathbb{T}^2$.

3. *Describe all non-homeomorphic 3-fold covers of* $\mathbb{T}^2 \# \mathbb{T}^2 \# \mathbb{T}^2$.

4. *Describe all non-homeomorphic 3-fold covers of* $\mathbb{R}\mathrm{P}^2$.

Exercise 14.47. *Given a compact, connected 2-manifold and a natural number* n*, describe all non-homeomorphic* n*-fold covers of that surface.*

14.7 Covers are Cool

The idea of a covering space arises from the idea of gracefully layering one space over another one. Rolling a line around and around a circle or cleverly wrapping a torus over a Klein bottle invite us to find relationships between the wrapping space and the wrapped space. Those relationships were rich indeed. Many involved the fundamental group—capturing the intuition that following a wrapping of one space over another somehow corresponds to loops.

Much of the development of the idea of covering spaces proceeded by looking at some simple examples, such as circles covering circles, and seeing how much the insights there could be extended. In the case of covering spaces, the extensions were wide and deep.

Applying the insights of covering spaces to graphs and to 2-manifolds let us see connections among familiar objects that revealed insights not only about those objects themselves, but related mathematical constructs such as free groups and collections of homeomorphisms. Covering spaces are both geometrically and algebraically appealing.

Chapter 15

Manifolds, Simplexes, Complexes, and Triangulability: Building Blocks

Our entire journey through topology has been based on the idea of taking familiar concepts and objects and then generalizing or abstracting them in one way or another. Ancient Greek mathematicians defined axioms of geometry with the idea that those features of geometry were the Platonic ideals on which our actual universe was built. So Euclidean geometry and Euclidean spaces still feel like home. Euclidean spaces also still hold a central place in mathematics, especially in the motivation and development of ideas of topology.

There are many methods for taking a familiar concept and letting it lead to mathematical ideas. In the first part of the book, we followed the strategy of looking for basic set-theoretic essences of our familiar surroundings. But a more direct way to use Euclidean spaces is to think of ways to construct spaces that are built using Euclidean pieces. In this chapter we will explore two methods for creating spaces built from Euclidean pieces—n-manifolds, in which every point lies in an open set homeomorphic to $\mathbb{R}^n$; and simplicial complexes, which are constructed from simple, rectilinear building blocks in $\mathbb{R}^n$.

15.1 Manifolds

> **Effective Thinking Principle.** *Extend Ideas.* Perhaps the richest source of new ideas comes from taking existing ideas and seeking extensions and generalizations.

We have already explored 2-manifolds such as the 2-sphere and torus—that is, spaces that are locally like the plane. Of course, there is no reason to restrict ourselves to 2-dimensionality when we are thinking of locally Euclidean spaces. Intuitively speaking, an *n-manifold* is a space that is locally the same as Euclidean n-space. These spaces are extremely important in many branches of mathematics as well as in many sciences. Manifolds are probably the most frequently studied spaces in all of topology. This section contains definitions, several examples, and a few basic theorems about manifolds.

Let's start by agreeing on the definitions of some basic subsets of Euclidean spaces such as cubes, balls, and spheres.

Definition. The n**-dimensional cube**, denoted $\mathbb{D}^n$, is defined as

$$\begin{aligned}\mathbb{D}^n &= \{(x_1, \ldots, x_n) \in \mathbb{R}^n : 0 \le x_i \le 1 \text{ for } i = 1, \ldots, n\} \\ &= \overbrace{[0,1] \times [0,1] \times \cdots \times [0,1]}^{n \text{ times}} \subset \mathbb{R}^n.\end{aligned}$$

For example, $\mathbb{D}^1 = [0,1]$. $\mathbb{D}^1$ is also called the unit interval. Likewise $\mathbb{D}^2$ is often called the unit square.

Definition. The **standard** n**-ball**, denoted B^n, is:

$$B^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + \ldots + x_n^2 \le 1\}.$$

We will often refer to B^2 as the unit disk or just the disk.

Exercise 15.1. *Show that the standard n-ball and the standard n-cube are homeomorphic spaces and each is compact and connected.*

Definition. The **standard** n**-sphere**, denoted by $\mathbb{S}^n$, is

$$\mathbb{S}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} : x_0^2 + \ldots + x_n^2 = 1\}.$$

$\mathbb{S}^1$ is called the **unit circle**.

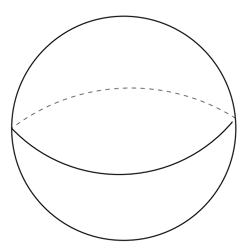


Figure 15.1: The 2-sphere.

Note. $\operatorname{Bd} B^{n+1} = \mathbb{S}^n$. Here we are referring to the topological boundary when B^{n+1} is embedded in $\mathbb{R}^{n+1}$ (as we viewed it above). When we mention the interior of an n-ball, we mean its interior as a subset of $\mathbb{R}^n$.

As usual, the terms n-ball and n-sphere will apply to any space homeomorphic to the standard n-ball and standard n-sphere respectively.

Exercise 15.2. *Show that for $n \geq 1$, the n-sphere is compact and connected.*

Exercise 15.3. *Consider $\mathbb{S}^0$, $\mathbb{S}^1$, and $\mathbb{S}^2$. Is any pair of them homeomorphic? If not, are there properties that allow you to distinguish them?*

While point-set topology provides tools to answer the previous question for these low dimensional spheres, it does not provide apparent means for answering this question in higher dimensions. Why are spaces of different dimensions different topologically? That question is one of the motivations for developing some of the concepts of algebraic topology in the chapters ahead.

Definition. An **n-dimensional manifold** or **n-manifold** is a separable, metric space, M, such that for each $p \in M$, there is a neighborhood U of p that is homeomorphic to an open set V in $\mathbb{R}^n$. Often an n-dimensional manifold is denoted M^n. A **curve** is a synonym for a 1-manifold and **surface** is a synonym for 2-manifold .

Hence a manifold is a topological space that is locally Euclidean.

Effective Thinking Principle. *Find Equivalences of Definitions.* After making a definition, see whether alternative definitions are equivalent.

Theorem 15.4. *For a separable, metric space M^n, the following are equivalent:*

1. *M^n is an n-manifold;*
2. *for each point $p \in M^n$, p has a neighborhood base of open sets each homeomorphic to the interior of an n-ball;*
3. *for every point $p \in M^n$, $p \in U$ where U is an open set homeomorphic to $\mathbb{R}^n$.*

Most people who study manifolds are not interested in topologically exotic spaces such as non-metric spaces. But as a matter of good mathematical practice, it is healthy to explore the consequences of removing technical assumptions. In the definition of an n-manifold, the technical assumptions of being separable and metric are important because otherwise the definition would admit strange spaces that do not behave in the spirit of Euclidean spaces and their subsets.

Exercise 15.5. *If you are comfortable with ordinal numbers, construct a topological space where every point has an open set containing it that is homeomorphic to $\mathbb{R}^1$, and yet the space is not metrizable. You might call your space the **long line**.*

The following exercise shows that in the definition of manifolds, the technical conditions of being separable and metrizable could be replaced by alternative conditions. You might enjoy this exercise if you enjoyed the theorems about metrizability.

Exercise 15.6. *Show that a locally Euclidean space is Hausdorff and second countable if and only if it is separable and metrizable.*

Let's look at a basic example of an n-manifold.

Exercise 15.7. *Show that $\mathbb{S}^n$ is an n-manifold.*

We conclude this section by listing two easy ways to produce new manifolds from others.

Theorem 15.8. *If M is an n-manifold and U is an open subset of M, then U is also an n-manifold.*

Theorem 15.9. *If M is an m-manifold and N is an n-manifold, then $M \times N$ is an $(m+n)$-manifold.*

Effective Thinking Principle. *Explore Variations.* After defining a concept, seek variations and extensions.

Many familiar objects look like manifolds at most points except they have a boundary to them. For example, in the closed disk $\mathbb{D}^2$, every point in the interior has a neighborhood homeomorphic to $\mathbb{R}^2$ while points on the boundary have neighborhoods homeomorphic to $\mathbb{R}^2_+$. So a category of spaces similar to manifolds are manifolds with boundary.

Definition. An n**-dimensional manifold with boundary** or n**-manifold with boundary** is a separable, metric space, M^n, such that for each $p \in M^n$, there is a neighborhood U of p that is homeomorphic either to $\mathbb{R}^n$ or to $\mathbb{R}^n_+$. Let M^n be an n-dimensional manifold with boundary. Then points of M^n that have neighborhoods homeomorphic to $\mathbb{R}^n_+$ are **boundary points** and the union of all boundary points of M^n is the **boundary of** M^n, denoted ∂M^n.

Theorem 15.10. *Let M^n be an n-dimensional manifold with boundary. Then ∂M^n is an $(n-1)$-manifold.*

15.2 Simplicial Complexes

Another strategy for using Euclidean pieces to construct objects is to start with rectilinear, simple building blocks and assemble them like you would assemble a bookshelf from a kit where each piece fits neatly with the others.

Effective Thinking Principle. *Identify Simple Building Blocks.* One of the most potent strategies for dealing with complexity is to identify a collection of elemental building blocks from which complex objects are made.

The first axiom of classical Euclidean geometry basically says that two points can be connected with a straight line segment. Straight lines are in some sense the most basic objects in Euclidean geometry. When we move to Cartesian coordinates, those straight lines can be described using linear combinations of the coordinates of two points to determine other points on the straight line determined by those points. Linear combinations of three non-collinear points in Euclidean space determine a flat plane; and if we restrict ourselves to linear combinations where the scalars add to 1, the linear combinations determine a triangle. This pattern leads us to define basic, rectilinear building blocks in each dimension: points, line segments, triangles, tetrahedra, and so on. These are the simple pieces, appropriately called *simplices*, from which complex objects, appropriated called *complexes* are built.

After defining the building blocks more formally, we will consider those spaces that are created by putting together a finite number of these geometrically appealing simplices. Since such spaces are built from a finite number of neatly assembled pieces, we can use our knowledge of the local geometry of these objects to help us analyze them. Also, when we have only a finite number of building blocks, induction becomes a very useful tool.

Effective Thinking Principle. *Start Concretely; Abstract Later.* Concrete ideas are the foundations for more abstract generalizations. Explore the concrete settings first to gain experience and intuition.

As is frequently the case in mathematics, when we define ideas, we need to make choices about the level of generality and abstraction with which to start. For example, if we were introducing the idea of the natural numbers, we could talk about equivalence classes of finite sets, or we could introduce the idea by talking about cows in a field. Most people understand ideas more meaningfully if they are first introduced in a concrete setting with more abstract versions reserved for later. We will take that concrete approach here by introducing simplices and complexes as subsets of $\mathbb{R}^n$. Later we will discuss how the ideas of simplices and complexes can be viewed more abstractly.

Let's begin by defining these fundamental objects. The basic building block is the **simplex**. The plural of simplex is *simplices*. A 0-simplex is simply a point in $\mathbb{R}^n$. A 1-simplex is a line segment, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron. To define a simplex more generally, recall that a set of points $v_0, \ldots, v_k$ in $\mathbb{R}^n$ is **affinely independent** if $\{v_1 - v_0, \ldots, v_k - v_0\}$ is a linearly independent set; informally speaking, no three points are collinear, no four points lie on a plane, etc. A **convex combination** of $v_0, \ldots, v_k$ is a linear combination of those points whose coefficients sum to 1. You can think of a convex combination of points as a weighted average of those points. The collection of all convex combinations of points fills in the convex hull of those points—producing a generalization of a triangle or tetrahedron.

Definition. A **k-simplex** is the set of all convex combinations of $k+1$ affinely independent points in $\mathbb{R}^n$. For affinely independent points $v_0, \ldots, v_k$ in $\mathbb{R}^n$, $\{v_0 \cdots v_k\}$ denotes the k-simplex

$$\left\{ \lambda_0 v_0 + \lambda_1 v_1 + \ldots + \lambda_k v_k \mid \text{ for each } i = 0, 1, \ldots, k,\ 0 \leq \lambda_i \leq 1 \text{ and } \sum_{i=0}^{k} \lambda_i = 1 \right\}.$$

Each v_i is called a **vertex** of $\{v_0 \cdots v_k\}$. (The plural of vertex is *vertices*.) Any point x in the k-simplex is specified uniquely by the $k+1$ coefficients (λ_i); these coefficients are called the **barycentric coordinates** of x. The **barycentric coordinate of x with respect to vertex** v_i is the coefficient λ_i.

In this notation, order does not matter: the 2-simplex $\{v_0 v_1 v_2\}$ is the same 2-simplex as $\{v_2 v_1 v_0\}$.

Definition. Any simplex τ whose vertices are a nonempty subset of the vertices of a k-simplex σ is a called a **face** of σ. If the number of vertices is $i+1$, then τ has **dimension** i and is called an **i-face** of σ and τ has **codimension** $k-i$, the number of dimensions below the top dimension.

If $\sigma = \{v_0 \cdots v_k\}$, the $(k-1)$-dimensional face of σ obtained by deleting the vertex v_j from the list of vertices of σ is denoted by $\{v_0 \cdots \widehat{v_j} \cdots v_k\}$, where the carat symbol means "omit."

Notice that our definition says that a simplex is a face of itself.

Exercise 15.11. *Show that if σ is a simplex and τ is one of its faces, then $\tau \subset \sigma$.*

Exercise 15.12. *Show that an n-simplex is homeomorphic to a closed n-dimensional ball.*

Although the n-simplex is homeomorphic to the closed n-dimensional ball, we will typically reserve the term n-simplex to refer to a subset of $\mathbb{R}^n$ like $\{v_0 \cdots v_n\}$ as described above.

Definition. A **simplicial complex** K (in $\mathbb{R}^n$) is a collection of simplices in $\mathbb{R}^n$ satisfying the following conditions.

1. If a simplex σ is in K, then each face of σ is also in K.

2. Any two simplices in K are either disjoint or their intersection is a face of each.

The **vertices** of K are the 0-simplices of K. The **dimension** of K is the maximum dimension of all the simplices in K.

In this book, we will restrict attention to finite simplicial complexes. However, topologists do study simplicial complexes containing infinitely many simplices. In that case, a condition of local finiteness of the simplices making up the complex is imposed to avoid pathologies.

Exercise 15.13. *Exhibit a collection of simplices that satisfies condition (1) but not (2) in the definition of a simplicial complex.*

A simplicial complex is a collection of simplices, but sometimes we want to refer to the topological space formed by these simplices.

Definition. The **underlying space** $|K|$ of a simplicial complex K is the set $\bigcup_{\sigma \in K} \sigma$, the union of all simplices in K, with a topology consisting of sets whose intersection with each simplex σ in K is open in σ.

For finite simplicial complexes, this topology is the topology inherited as a subspace of $\mathbb{R}^n$.

Exercise 15.14. *Let K be the simplicial complex in $\mathbb{R}^2$:*

$$K = \{\sigma, e_1, e_2, e_3, e_4, e_5, v_1, v_2, v_3, v_4\}$$

where $\sigma = \{(0,0)(0,1)(1,0)\}$, $e_1 = \{(0,0)(0,-1)\}$, $e_2 = \{(0,-1)(1,0)\}$, $e_3 = \{(0,0)(0,1)\}$, $e_4 = \{(0,1)(1,0)\}$, $e_5 = \{(1,0)(0,0)\}$, $v_1 = \{(0,0)\}$, $v_2 = \{(0,1)\}$, $v_3 = \{(1,0)\}$, and $v_4 = \{(0,-1)\}$. Draw K and its underlying space.

Definition. A topological space X is **triangulable** if it is homeomorphic to the underlying space of a simplicial complex K. In that case, we say K is a **triangulation** of X.

Note that a given topological space may have many triangulations.

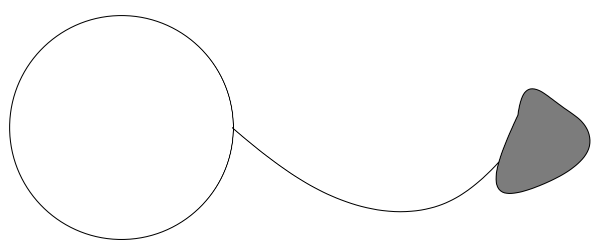


Figure 15.2: A triangulable space.

Exercise 15.15. *Show that the space shown in Figure 15.2 is triangulable by exhibiting a simplicial complex whose underlying space it is homeomorphic to.*

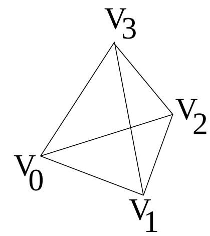


Figure 15.3: Tetrahedral surface.

Example. Suppose v_0, v_1, v_2, and v_3 are affinely independent points. The simplicial complex K (shown in Figure 15.3) is

$$\left\{ \begin{array}{c} \{v_0v_1v_2\}, \{v_0v_1v_3\}, \{v_0v_2v_3\}, \{v_1v_2v_3\}, \\ \{v_0v_1\}, \{v_0v_2\}, \{v_0v_3\}, \{v_1v_2\}, \{v_1v_3\}, \{v_2v_3\}, \\ \{v_0\}, \{v_1\}, \{v_2\}, \{v_3\} \end{array} \right\}.$$

Its underlying space is a triangulated 2-manifold homeomorphic to $\mathbb{S}^2$.

Exercise 15.16. *For each n, $\mathbb{S}^n$ is triangulable.*

15.3 Simplicial Maps and PL Homeomorphisms

Effective Thinking Principle. *After Objects, Transformations.* After defining a category of mathematical entities, a natural step is to investigate transformations between them.

We now define maps between simplicial complexes that respect their triangulations.

Definition. Suppose X and Y are topological spaces. A function $f : X \to Y$ is a **simplicial map** if and only if there exist simplicial complexes K and L such that $|K| = X$ and $|L| = Y$ and f maps each simplex of K linearly onto a (possibly lower-dimensional) simplex in L. We use the notation

$$f : K \to L$$

where K and L are complexes (rather than topological spaces) to signify f is a simplicial map. Every such simplicial map gives rise to an underlying continuous function $f : |K| \to |L|$ on the underlying topological spaces. (We'll use the same name to indicate the simplicial map and the underlying continuous function, and context will make clear which one we mean.)

Definition. A simplicial map f is a **simplicial homeomorphism** if and only if it is a bijection; and in that case, the two complexes are **simplicially homeomorphic**.

The linearity of f on simplices means the following. Suppose x is in the k-simplex σ spanned by the vertices $\{v_i\}_{i=0}^k$. Then x is a convex combination of the vertices $\{v_i\}_{i=0}^k$. Being a convex combination means $x = \lambda_0 v_0 + \lambda_1 v_1 + ... + \lambda_k v_k$ where $0 \leq \lambda_i \leq 1$ for each $i = 0, 1, \ldots, k$, and $\sum_{i=0}^k \lambda_i = 1$. The linearity of f means that $f(x)$ will be the same convex combination of the vertex images $\{f(v_i)\}_{i=0}^k$, that is, $f(x) = \lambda_0 f(v_0) + \lambda_1 f(v_1) + ... + \lambda_k f(v_k)$.

Note that a simplicial map sends vertices to vertices, but not necessarily injectively. If the $\{f(v_i)\}_{i=0}^k$ are not distinct, the simplex spanned by those points will be of lower dimension than the simplex spanned by $\{v_i\}_{i=0}^k$.

Theorem 15.17. *A simplicial map from K to L is determined by the images of the vertices of K.*

Theorem 15.18. *A composition of simplicial maps is a simplicial map.*

Theorem 15.19. *If two complexes are simplicially homeomorphic, then there are 1-1 correspondences between their k-simplices for each $k \geq 0$.*

Theorem 15.20. *A simplicial map $f : K \to L$ is continuous as a map on the underlying spaces. In particular, simplicially homeomorphic complexes have homeomorphic underlying spaces.*

Effective Thinking Principle. *What Things are Equal?.* After defining something, a natural question to ask is when two items are equivalent.

The problem with the concept of simplicial homeomorphism as defined above is that the very same physical object might not be equivalent to itself if it had two different triangulations. So we need an expanded idea of equivalence of simplicial complexes. For that purpose, we introduce the idea of a subdivision of a complex.

Definition. Let K be a simplicial complex. Then a simplicial complex K' is a **subdivision** of K if and only if each simplex of K' is a subset of a simplex of K and each simplex of K is the union of finitely many simplices of K'.

Figure 15.4 illustrates a finite simplicial complex and a subdivision of it.

Definition. If K and L are complexes, a continuous map $f : |K| \to |L|$ is called **piecewise linear** or **PL** if and only it there are subdivisions K' of K and L' of L such that f is a simplicial map from K' to L'. If there exist subdivisions such that f is a simplicial homeomorphism, then f is a **PL homeomorphism** and the spaces are **PL homeomorphic**.

The letters "PL" are used interchangeably with *piecewise linear*.

Theorem 15.21. *A composition of PL maps is PL. A PL homeomorphism is an equivalence relation.*

Theorem 15.22. *PL homeomorphic complexes are homeomorphic as topological spaces.*

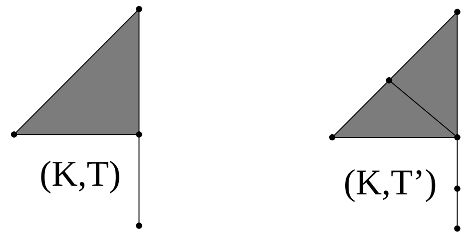


Figure 15.4: A simplicial complex and a subdivision.

15.4 Simplicial Approximation

One of the most useful and important facts about simplicial maps is that every continuous function between simplicial complexes can be approximated arbitrarily closely by a simplicial map. To make this statement true, we cannot necessarily use the originally given triangulations—we will have to take subdivisions first.

Exercise 15.23. *Let K is a complex consisting of the boundary of a triangle (three vertices and three edges) and L be an isomorphic complex. Both $|K|$ are $|L|$ are topologically circles. There is a continuous map that takes the circle $|K|$ and winds it twice around the circle $|L|$; however, show that there is no simplicial map from K to L that winds the circle $|K|$ twice around the circle $|L|$.*

Our goal is to take a continuous function f between the underlying spaces of simplicial complexes and find another continuous function g that has two additional virtues, namely, (1) being a simplicial map; and (2) for every point x in the domain, we want $d(f(x), g(x))$ to be small.

Given the previous exercise, we see that it is impossible to accomplish our task of finding simplicial approximations if we stick with the triangulations of the domain and range that we are given. So obtaining a simplicial approximation of a continuous function may require us to pass to finer and finer triangulations.

A systematic method for finding appropriate subdivisions is to use a technique called *barycentric subdivision*. We were introduced to barycentric subdivisions for triangles in the chapter on Classification of 2-Manifolds. Here we extend that idea to arbitrary dimensions.

Definition. The **barycenter** of a k-simplex $\{v_0 v_1 \cdots v_k\}$ in $\mathbb{R}^n$ is the point $\frac{1}{k+1}(v_0 + \cdots + v_k)$.

The barycenter is at the very center of the simplex. For example, the barycenter of the 2-simplex $\{v_0 v_1 v_2\}$ is $\frac{1}{3}v_0 + \frac{1}{3}v_1 + \frac{1}{3}v_2$. The barycentric coordinates of the barycenter of a 2-simplex are $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The barycenter of a 0-simplex $\{v_0\}$ is just v_0.

The definition below of a barycentric subdivision is hard to parse, but after you see what it means in the case of a 2-simplex as pictured and described after the definition, the definition will make sense.

Definitions. 1. Let σ be an n-simplex. The **first barycentric subdivision of** σ, denoted $\mathsf{sd}\,\sigma$, is the complex of all simplices of the form $\{b_0 \cdots b_k\}$, where b_i is the barycenter of a face σ^i of σ from a chain of faces of σ, $\sigma^0 \subset \sigma^1 \subset \cdots \subset \sigma^k$ of increasing (not necessarily consecutive) dimensions. The maximal simplices, that is, the n-simplices of $\mathsf{sd}\,\sigma$ each arise from a maximal sequence of faces, that is, from faces of consecutive dimensions starting with a vertex of σ. Thus an n-simplex of $\mathsf{sd}\,\sigma$ corresponds exactly to a permutation of the vertices of σ.

2. Let K be a simplicial complex. The **first barycentric subdivision of** K, denoted $\mathsf{sd}\,K$, is the complex consisting of all the simplices in the barycentric subdivision of each simplex of K.

3. The **second barycentric subdivision**, denoted $\mathsf{sd}^2\,K$, is the first barycentric subdivision of $\mathsf{sd}\,K$. Proceeding in this way, the **m-th barycentric subdivision** is denoted $\mathsf{sd}^m\,K$.

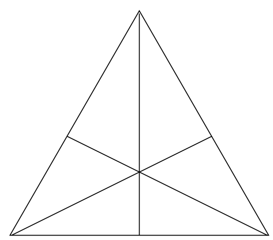


Figure 15.5: Barycentric subdivision of a 2-simplex.

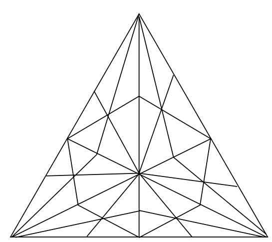


Figure 15.6: Second barycentric subdivision of a 2-simplex.

Thus in the first barycentric subdivision of a 2-simplex there are 6 maximal simplices (triangles), each of which has one corner at the barycenter of a vertex, another at the barycenter of an edge, and the third at the barycenter of the given 2-simplex. See Figure 15.4.

Exercise 15.24. *How many n-simplices are there in the first barycentric subdivision of an n-simplex?*

Exercise 15.25. *Convince yourself that the barycentric subdivision of a complex K is, in fact, a subdivision of K.*

One of the important necessities when making approximations of maps is to be able to deal with pieces of increasingly small sizes. So one good feature of barycentric subdivisions is that the simplices become smaller.

Theorem 15.26. *Let K be a finite simplicial complex and let a_n be the maximum among the diameters of simplices in $\mathsf{sd}^n K$. Then $\lim_{n\to\infty} a_n = 0$.*

Now we are ready to construct a simplicial "approximation" g to a given continuous map $f : |K| \to |L|$. Our first challenge is to decide in what sense the approximation should approximate f. The natural choice would be to make the simplicial approximation g point-wise close to the given f. In order to produce such a function g, subdivisions of both K and L would be required. So we will first find an approximation to f that is a sort of half-way step that will involve subdividing only K.

> **Effective Thinking Principle.** *Half Way Steps.* Break a difficult challenge into smaller steps.

It would be useful to develop a concept of a local neighborhood in a simplicial complex that reflects the simplicial structure of the complex. Simplicial complexes are naturally broken into their simplicial pieces, each of which is determined by its vertices. So one way to think about closeness of two points x and y in a simplicial complex is to think about whether x and y lie in the same or nearby simplices. The set of all the simplices that share a given vertex is a natural neighborhood to think about in a simplicial complex. That perspective leads to the following definitions.

Definitions. 1. Let K be a simplicial complex. The **minimal face** of $x \in |K|$ is the simplex of K of smallest dimension that contains x.

2. The **star of a vertex** v in K, denoted $\mathrm{St}(v)$, is the set of all points whose minimal face contains v.

So the star of a vertex is a natural notion of a neighborhood in a simplicial complex. Notice that the definition of the star of a vertex is basically the interior of the union of all simplices that contain v. Making the definition of star refer to minimal faces allows us to conclude that the star of a vertex is an open set, which fact is the content of the next exercise.

Exercise 15.27. *The star of a vertex v in a complex K is an open set of $|K|$, and the collection of all vertex stars covers $|K|$.*

Exercise 15.28. *If the simplex $\sigma = \{v_0, \dots, v_k\}$ in K is the minimal face of a point $x \in |K|$, then $x \in \mathrm{St}(v_0) \cap \dots \cap \mathrm{St}(v_k)$.*

Simplicial maps take simplices in the domain to simplicies in the range. A partway step in that direction is a map that takes simplices in the domain into the star of a vertex in the range.

Definition. Suppose K and L are simplicial complexes. A continuous function $f : |K| \to |L|$ satisfies the **star condition with respect to** K **and** L if and only if for each vertex $v \in K$, there is a vertex $w \in L$ such that

$$f(\mathrm{St}(v)) \subset \mathrm{St}(w).$$

So now we are in a position to develop a notion of approximation of maps that refers to the simplicial structure of simplicial complexes rather than distances.

Definition. Let X and Y be the underlying spaces of simplicial complexes and let $f : X \to Y$ be a continuous map. A **simplicial approximation** of f is a simplicial map $g : K \to L$ where K is a simplicial complex with $|K| = X$ and L is a simplicial complex with $|L| = Y$ such that for each vertex v of K,

$$f(\mathrm{St}(v)) \subset \mathrm{St}(g(v)).$$

Suppose f is a continuous function satisfying the star condition with respect to some K and L, and we want to construct a simplicial approximation g of f. There really is only one thing to try. Namely, let's define a function g on the vertices of K by setting $g(v) = w$ for any w that satisfies $f(\mathrm{St}(v)) \subset \mathrm{St}(w)$. Can such a map g on the vertices of K always be extended to a simplicial map from K to L so that simplices map to simplices? The next theorem answers that question in the affirmative.

Theorem 15.29. *Suppose K and L are simplicial complexes. Then a continuous function $f : |K| \to |L|$ satisfies the star condition with respect to K and L if and only if f has a simplicial approximation $g : K \to L$.*

Theorem 15.30. *If $g, g' : K \to L$ are both simplicial approximations to a continuous function $f : |K| \to |L|$, then for any point $x \in |K|$, if σ is the minimal face of x in K, the point $f(x)$ and the simplices $g(\sigma)$ and $g'(\sigma)$ all lie in a single simplex of L.*

When two points are in the same simplex, the straight line segment between them is also in that simplex. Such straight lines can help us construct a homotopy between a function and a simplicial approximation of it.

Recall that two maps $f, g : X \to Y$ are **homotopic** if there is a continuous map $F : X \times [0, 1] \to Y$ such that for all $x \in X$, $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$.

The previous theorem will allow you to construct a homotopy between f and g.

Theorem 15.31. *Let K and L be simplicial complexes. If $f : |K| \to |L|$ has a simplicial approximation $g : K \to L$, then f is homotopic to $g : |K| \to |L|$.*

Intuitively, if K has very small simplices, then a continuous function f is likely to satisfy the star condition. Barycentric coordinates will help us break the domain into small enough pieces.

Theorem 15.32. *Suppose K and L are finite simplicial complexes and $f : |K| \to |L|$ is a continuous function between their underlying spaces. Then there exists $m \geq 1$ such that the function $f : |\mathsf{sd}^m K| \to |L|$ satisfies the star condition with respect to $\mathsf{sd}^m K$ and L.*

Putting all these theorems together, we can show that every continuous function between simplicial complexes can be approximated by a homotopic simplicial map.

Theorem 15.33. *Suppose K and L are simplicial complexes and $f : |K| \to |L|$ is a continuous function between their underlying spaces. Then there exists $m \geq 1$ such that f has a simplicial approximation $g : \mathsf{sd}^m K \to L$.*

Note that the simplicial approximation g uses a subdivision of K but not L, and note that g may not be pointwise close to f. However, if we allow subdivisions of L as well as K, we can choose our simplicial approximations to be point-wise as close as we like.

Theorem 15.34 (Simplicial Approximation Theorem)**.** *Let K and L be simplicial complexes, let $f : |K| \to |L|$ be a continuous function between their underlying spaces, and let $\epsilon > 0$. Then there exist $m, n \geq 1$ and a simplicial map $g : \mathsf{sd}^n K \to \mathsf{sd}^m L$ such that f is homotopic to g and for every $x \in |K|$, $d(f(x), g(x)) < \epsilon$.*

This theorem tells us that whenever we are dealing with continuous functions between spaces that are triangulable and when close approximations are good enough, then we can use simplicial maps. Simplicial maps are often very helpful.

15.5 Sperner's Lemma and the Brouwer Fixed Point Theorem

In the chapters ahead we will see the value of simplices and complexes in their relationship to homology. But the ideas of simplices and complexes can be used to prove some of the most interesting theorems of topology directly. In this section, we will prove the famous Brouwer Fixed Point Theorem, which states that every continuous function from a closed n-ball to itself leaves some point fixed. The Brouwer Fixed Point Theorem has important applications in many fields. Some notable applications include the existence of solutions to differential equations and the existence of mixed strategy Nash equilibria in game theory.

The adventure begins with a combinatorial insight called Sperner's Lemma. We will start with dimension one and work our way up.

Theorem 15.35. *Let K be a subdivision of a 1-simplex σ. Label every vertex of K with a 0 or a 1 such that one of the two vertices of σ is labeled with a 0 and the other is labeled with a 1. Then there is a 1-simplex τ in K such that one vertex of τ is labeled 0 and the other vertex of τ is labeled 1.*

The preceding theorem can be thought of as a combinatorial version of the Intermediate Value Theorem.

> **Effective Thinking Principle.** *Seek Extensions.* After seeing an insight, seek ways to extend it.

In a way, the above theorem seems rather simple—and perhaps does not seem too thrilling. But it gives us a chance to think about what the analogous version might be in dimension two. Here is a possible 2-dimensional extension.

Theorem 15.36. *Let K be a subdivision of a 2-simplex σ. Label every vertex of K with 0, 1, or 2 such that the three vertices of σ are labeled with different numbers. Then there is a 2-simplex τ in K such that its vertices are labeled with all different numbers.*

Surely, we cannot resist extending this result to arbitrary dimensions. The following theorem is called Sperner's Lemma.

Theorem 15.37 (Sperner's Lemma)**.** *Let K be a subdivision of a n-simplex σ. Label every vertex of K with one of $\{0, 1, \ldots, n\}$ such that the $(n+1)$ vertices of σ are labeled with different numbers. Then there is an n-simplex τ in K such that its vertices are labeled with all different numbers.*

Sperner's Lemma and knowledge about uniform continuity of continuous maps between compact spaces will allow you to prove the No Retraction Theorem.

Theorem 15.38. *Let σ be an n-simplex with boundary $\partial\sigma$. There does not exist a continuous function $r : \sigma \to \partial\sigma$ such that for every $x \in \partial\sigma$, $r(x) = x$.*

Next you will use the No Retraction Theorem to prove the most famous fixed point theorem—the Brouwer Fixed Point Theorem.

Theorem 15.39 (n-dimensional Brouwer Fixed Point Theorem)**.** *Let σ^n be an n-simplex. For every continuous function $f : \sigma^n \to \sigma^n$ there exists a point $x \in \sigma^n$ such that $f(x) = x$.*

Perhaps the easiest way to prove the Brouwer Fixed Point Theorem is to prove that it is equivalent to the No Retraction Theorem. So instead of proving the Brouwer Fixed Point Theorem directly, prove that it is equivalent to the No Retraction Theorem. To prove the equivalence of those two statements, you need to answer the two questions below. It might be convenient to think of the n-simplex as an n-ball (to which it is homeomorphic), because an n-ball is rounder.

1. Suppose you were given a retraction from an n-simplex to its boundary. Then how could you use that map to construct a fixed point free map from the n-simplex to itself?

2. Suppose you were given a fixed point free map from the n-simplex to itself. Then how could you use that map to produce a retraction from the n-simplex to its boundary?

Convince yourself that answering these two questions would in fact prove the equivalence of the No Retraction Theorem and the Brouwer Fixed Point Theorem. Since you have already proved the No Retraction Theorem, then after proving the equivalence, you will have proved the Brouwer Fixed Point Theorem as well.

15.6 The Jordan Curve Theorem, the Schoenflies Theorem, and the Triangulability of 2-Manifolds

In dimensions one, two, and three, every manifold is triangulable. Hence, to study these low dimensional manifolds, it suffices to study simplicial complexes. The proofs of these triangulability theorems are quite involved for dimensions two and three, but the hardy reader might enjoy the adventure in dimension two. The 3-dimensional versions would take us too far astray, so, alas, we are forced to omit those delights here, but hope you will meet them sometime during your lifetime of learning.

In this section, we guide you through a possible approach to proving several fundamental theorems in plane topology: the Jordan Curve Theorem, the Schoenflies Theorem, the triangulability of 2-manifolds, and the fact that orientability and the Euler Characteristic of compact 2-manifolds are well-defined features. These theorems are difficult because they involve grappling with detailed plane geometry and taking limits of sequences of homeomorphisms. We will start with a sort of heuristic overview of the strategy and then afterward dive into the deep end. Feel free to skip this whole section.

A surface, or 2-manifold, is locally homeomorphic to the plane. Analyzing the global structure of a compact, connected surface in the chapter about the classification of 2-manifolds required us to look at the whole object. As we saw in that chapter, one of the most important steps we took to simplify our task was to accept the fact that every compact surface is triangulable. After accepting that 2-manifolds are triangulable, we knew how the triangles fit together locally and we knew there are only finitely many of them, so we were able to analyze our surfaces by moving from triangle to adjacent triangle or seeing why that implied that compact, connected, 2-manifolds were actually homeomorphic to polygonal disks with edges identified in pairs. The geometric simplicity of triangles was greatly helpful to our analysis, so proving that every compact 2-manifold is homeomorphic to a simplicial complex in $\mathbb{R}^n$ is worth considerable effort.

Outline of the proof of the triangulability of 2-manifolds: We know that a compact 2-manifold has a finite cover of open sets each of which is homeomorphic to an open disk. You can show that the 2-manifold can, in fact, be covered by a finite number of closed disks. Suppose those topological disks in the 2-manifold intersected one another neatly. Then overlaps could be removed to create a cover of the 2-manifold by subsets each homeomorphic to a closed disk where the interiors of the closed disks were disjoint and every two of those closed disks either were disjoint or met at a single point or met along a common arc on the boundary of each. From there, each disk could be broken up into topological triangles where each pair of those triangles met with the same constraints as the disk intersections had. Then it would be easy to create a homeomorphism to a desired subset of $\mathbb{R}^5$ by choosing points in $\mathbb{R}^5$ more or less randomly and taking the vertices of the topological triangles to those points in $\mathbb{R}^5$, creating triangles from every triple of points that corresponded to a topological triangle in the 2-manifold, and creating the homeomorphism from each topological triangle in the 2-manifold to each rectilinear triangle in $\mathbb{R}^5$ appropriately.

To fill in the details of the argument above, the difficult part occurs when we try to come to grips with the sentence: "Suppose those topological disks in the 2-manifold intersected one another neatly." First we will have to decide what 'neatly' means. Then we will need to deal with the challenge of getting topological arcs in the plane to intersect one another in reasonable ways. Also, we will need to prove the Jordan Curve Theorem (which we will also prove in more generality in a later chapter of this book). We will also need to prove the Schoenflies Theorem. The heart of the challenges in the triangulability question lies in the details of the proof of the Schoenflies Theorem.

Fair warning: many of the following theorems are challenging; however, they are not impossibly difficult.

We can start with a polygonal version of the Jordan Curve Theorem. Do not hesitate to use simple facts about $\mathbb{R}^2$ such as the local geometry of a line segment or an angle. There are several approaches to proving the next two theorems, but one way to take advantage of the polygon hypothesis is to induct on the number of segments that make up the polygon.

Theorem 15.40. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding that is a polygon, that is, $h(\mathbb{S}^1)$ consists of a finite* *number of straight line intervals. Then $h(\mathbb{S}^1)$ separates $\mathbb{R}^2$ into two components and each point of $h(\mathbb{S}^1)$ is a limit point of each component.*

Next we can prove a polygonal version of the Schoenflies Theorem.

Theorem 15.41. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding that is a polygon, that is, $h(\mathbb{S}^1)$ consists of a finite number of straight line intervals. Then there is a homeomorphism $H : \mathbb{R}^2 \to \mathbb{R}^2$ such that $H(h(S^1))$ is the unit circle.*

To prove the next theorem, you might think of covering up the square with bricks, like a brick wall, that are so small that no brick touches both A and B. One good thing about bricks is that if you take any collection of bricks, such as those that touch A, the boundary of any such collection of bricks is either a polygonal arc or a polygonal simple closed curve.

Theorem 15.42. *Let A and B be disjoint closed subsets of $[0,1] \times [0,1]$ such that $A \cap ([0,1] \times \{0,1\} \cup \{1\} \times [0,1]) = \varnothing$ and $B \cap ([0,1] \times \{0,1\} \cup \{0\} \times [0,1]) = \varnothing$. Then there exists a path in $[0,1] \times [0,1]$ from $(1/2, 0)$ to $(1/2, 1)$ that misses $A \cup B$.*

The next theorem states that no embedding of a closed interval in the plane separates the plane. In proving that theorem, you might want to first prove the case where you can find a polygonal arc A (that is, one made of a finite number of straight line segments) from p to q that misses the first half of the embedding and another polygonal arc B from p to q that misses the second half of the embedding such that $A \cap B = \{p, q\}$ and where the midpoint of the embedded arc is outside the simple closed curve $A \cup B$.

Theorem 15.43. *Suppose $h : [0,1] \to \mathbb{R}^2$ is an embedding and suppose p and q are points in $\mathbb{R}^2$ not contained in $h([0,1])$. Then there exists a path $f : [0,1] \to \mathbb{R}^2$ such that $f(0) = p$, $f(1) = q$, and $f([0,1]) \cap h([0,1]) = \varnothing$.*

The preceding theorem can be summarized by saying that no arc separates the plane. Knowing that no arc separates the plane, you can prove the Jordan Curve Theorem for simple closed curves that have a flat spot.

Theorem 15.44. *Suppose $g, h : [0,1] \to \mathbb{R}^2$ are embeddings such that $g([0,1])$ is a straight line segment, $g(0) = h(0)$, $g(1) = h(1)$, and $g((0,1)) \cap h((0,1)) = \varnothing$. Then $g([0,1]) \cup h([0,1])$ separates $\mathbb{R}^2$ into two components and each point of $g([0,1]) \cup h([0,1])$ is a limit point of each component.*

That special case of the Jordan Curve Theorem allows you to prove the general case.

Theorem 15.45. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding. Then $h(\mathbb{S}^1)$ separates $\mathbb{R}^2$ into two components and each point of $h(\mathbb{S}^1)$ is a limit point of each component.*

Effective Thinking Principle. *Explore Extensions.* After proving a theorem, investigate whether a more general statement would be true.

The Jordan Curve Theorem tells us that the complement of an embedded circle has two components and is the boundary of each. A natural question is whether it would be possible to construct any continuum $C \subset \mathbb{R}^2$ such that $\mathbb{R}^2 - C$ has more than two components and yet every point of C is a limit point of each of the three or more components of its complement. Surprisingly, creating such a continuum C is possible. We will use the fact that if $C_0 \supset C_1 \supset C_2 \supset \cdots \supset C_i \supset \cdots$ are a nested sequence of continua, then $C = \bigcap_{i=0,1,2,\cdots} C_i$ is a continuum.

Exercise 15.46. *1. Let C_0 be a disk with two holes. Construct a subset C_1 of C_0 such that C_1 is also homeomorphic to a disk with two holes, and for which each point $x \in C_1$ is within distance 1 of points in each of the three components of $\mathbb{R}^2 - C_1$.*

2. *Construct a continuum $C \subset \mathbb{R}^2$ such that $\mathbb{R}^2 - C$ has three components and each point $x \in C$ is a limit point of each component of $\mathbb{R}^2 - C$.*

3. *Construct a continuum $C \subset \mathbb{R}^2$ such that $\mathbb{R}^2 - C$ has infinitely many components and each point $x \in C$ is a limit point of each component of $\mathbb{R}^2 - C$.*

The examples you constructed in the previous exercise are called the Lakes of Wada.

Now that we know the Jordan Curve Theorem, we know what we mean by the inside of a simple closed curve in the plane. We will now proceed to prove the Schoenflies Theorem by proving that any simple closed curve in the plane can be filled up with a sort of expanding bull's eye where the outer rings are getting increasingly closer to the simple closed curve. To construct such rings, let's start by finding some polygonal simple closed curves that are close to the embedded simple closed curve. You might consider using small bricks again. Notice that the polygonal simple closed curve you will construct in the theorem below is close to the topological embedding as a set, but it is not necessarily close to the embedding as a map.

Theorem 15.47. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding. Let p be a point in the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$ and let $\epsilon > 0$. Then there exists an embedding $g : \mathbb{S}^1 \to \mathbb{R}^2$ such that $g(\mathbb{S}^1)$ is a polygonal simple closed curve in the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$, $g(\mathbb{S}^1)$ lies in the ϵ-neighborhood of $h(\mathbb{S}^1)$, and p is in the bounded component of $\mathbb{R}^2 - g(\mathbb{S}^1)$.*

Since $g(\mathbb{S}^1)$ is in the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$, it follows that $h(\mathbb{S}^1)$ is in the unbounded component of $\mathbb{R}^2 - g(\mathbb{S}^1)$.

The problem with our polygonal simple closed curve $g(\mathbb{S}^1)$ is that it is not close as a map to the simple closed curve $h(\mathbb{S}^1)$. We will start to remedy that defect by creating a map g that is somewhat close to the map h on part of $\mathbb{S}^1$. In the following theorem you are given an arc A on $\mathbb{S}^1$ with endpoints a and b. In proving the theorem, you might consider taking a point p_a in the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$ that is very close to $h(a)$ and drawing a straight line from p_a toward $h(a)$ and finding the first point q_a where that straight line hits $h(\mathbb{S}^1)$. By choosing p_a very close to $h(a)$, you can be assured that q_a is as close to $h(a)$ as you wish. The segment from p_a to q_a is a sort of sticker coming off the topological simple closed curve heading into the interior of it. Perhaps you can prove that any polygonal simple closed curve such as you found in the previous theorem would have to cross through that segment. In any case, erecting those stickers might be useful. The statement of the following theorem is perhaps more difficult than its proof.

Theorem 15.48. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding. Let $\epsilon > 0$ and let A be an arc on $\mathbb{S}^1$ with endpoints a and b such that the diameter of $h(A)$ is less than ϵ and let p be a point in the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$. Then there exists an embedding $g : \mathbb{S}^1 \to \mathbb{R}^2$ such that $g(\mathbb{S}^1)$ is a polygonal simple closed curve in the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$, p is in the bounded component of $\mathbb{R}^2 - g(\mathbb{S}^1)$, and for every $x \in A$, $d(g(x), h(x)) < \epsilon$.*

Putting in more stickers at each stage will allow you to create a sequence of polygonal simple closed curves that become increasingly close as maps to h and that create a whole constellation of ever finer polygonal simple closed curves nearer and nearer to $h(\mathbb{S}^1)$. Using the polygonal Schoenflies Theorem in each of those pockets will allow you to prove that the closure of the interior of $h(\mathbb{S}^1)$ is homeomorphic to a disk.

Theorem 15.49. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding, let U be the bounded component of $\mathbb{R}^2 - h(\mathbb{S}^1)$, and let D be the closed unit ball in $\mathbb{R}^2$. Then there is a homeomorphism $H : (U \cup h(\mathbb{S}^1) \to D$.*

Realizing that adding a point at infinity to $\mathbb{R}^2$ gives $\mathbb{S}^2$ is one way to see that dealing with the unbounded component is not a huge challenge. Alternatively, you could repeat the arguments above on the unbounded component. Either way, you have successfully proved the Schoenflies Theorem.

Theorem 15.50. *Let $h : \mathbb{S}^1 \to \mathbb{R}^2$ be an embedding. Then there is a homeomorphism $H : \mathbb{R}^2 \to \mathbb{R}^2$ such that $H(h(\mathbb{S}^1))$ is the unit circle.*

During the course of the proof of the Schoenflies Theorem, you have actually learned that any embedding of a simple closed curve or an arc in the plane is arbitrarily close to a polyhedral embedding.

Theorem 15.51. *Let $h : [0,1] \to \mathbb{R}^2$ be an embedding of $[0,1]$ in the plane and let $\epsilon > 0$. Then there exists an embedding $g : [0,1] \to \mathbb{R}^2$ such that $h(0) = g(0)$, $h(1) = g(1)$, and for every $x \in [0,1]$, $d(h(x), g(x)) < \epsilon$.*

We can use our insights from the proofs to show that for any two pointwise close embeddings of an arc into the plane, we can find a homeomorphism of the plane that takes one embedding to the other. This next theorem says that two close embeddings of an arc into the plane are morally the same.

Theorem 15.52. *Let $f, g : [0,1] \to \mathbb{R}^2$ be two embeddings of $[0,1]$ in the plane such that $f(0) = g(0)$ and $f(1) = g(1)$. Let $\epsilon > 0$. Suppose for every $x \in [0,1]$, $d(f(x), g(x)) < \epsilon$. Then there exists a homeomorphism $h : \mathbb{R}^2 \to \mathbb{R}^2$ such that for every $t \in [0,1]$, $h(f(t)) = g(t)$ and for every $x \in \mathbb{R}^2$, $d(x, h(x)) < \epsilon$.*

The insight that one embedded arc is the same as another via a homeomorphism of the plane will allow you to triangulate 2-manifolds. The fundamental problem that arises when trying to triangulate 2-manifolds is that overlapping patches that are each homeomorphic to a disk do not agree on what should be viewed as 'straight.' So it is not easy to match up something that looks like a triangle in one patch with something that looks like a triangle in an overlapping patch, but from the point of view of the first patch just looks like a topological triangle. You can now match up topological embeddings neatly and successfully prove the triangulability of compact 2-manifolds.

Theorem 15.53. *Every compact 2-manifold is triangulable, that is, it is homeomorphic to a subset C of $\mathbb{R}^n$ consisting of a finite collection $T = \{\sigma_i\}_{i=1}^k$ of (rectilinear) 2-simplices where each pair of 2-simplices are disjoint or they meet in one vertex of each or they share a single edge. Since the space C is homeomorphic to a 2-manifold, each edge of each 2-simplex making up C is shared by exactly two triangles, and around each vertex is a circle of triangles whose union is a disk.*

In fact, since we know that every metric space is paracompact, we can actually prove the more general theorem that every 2-manifold is triangulable. For a non-compact 2-manifold, being triangulable means that it is the union of a locally finite collection of subsets each of which is homeomorphic to a 2-simplex such that those embeddings of 2-simplices fit together like the rectilinear ones did in the case of compact 2-manifolds.

Theorem 15.54. *Every 2-manifold is triangulable.*

The remaining detail to be considered regarding triangulations is to show that for compact 2-manifolds, any two triangulations are equivalent, meaning that given any two rectilinear triangulations of the same 2-manifold, it is possible to subdivide all the triangles in each triangulation to arrive at a common triangulation. That is, the subdivisions of the two given triangulations produce two triangulations that are combinatorically identical, that is, there is a one-to-one correspondence between the vertices of one triangulation to the vertices of the other such that that correspondence induces a one-to-one correspondence between the edges and the 2-simplices as well.

As a first step toward that end, we need to show how to change a topological homeomorphism that goes from one rectilinearly triangulated 2-manifold into another rectilinearly triangulated 2-manifold and show that the homeomorphism can be adjusted so that the images of edges are polygonal arcs in the image 2-manifold. This step amounts to showing that topological arcs can be adjusted to become polygonal via a homeomorphism.

Lemma 15.55. *Let M_1 and M_2 be two rectilinearly triangulated 2-manifolds in $\mathbb{R}^n$. Let $h : M_1 \to M_2$ be a topological homeomorphism. Then there exists a homeomorphism $g : M_1 \to M_2$ such that the image of every edge in the triangulation of M_1 is a polyhedral arc in M_2.*

Taking edges to polygonal arcs is a step toward making our homeomorphism take triangles to triangles. In the next theorem, the originally given triangles that make up M_1 will probably be subdivided into smaller triangles in the process of proving this theorem. In other words, the conclusion of the theorem does not assert that the original triangles of M_1 go to triangles, only that the subtriangles go to triangles.

Theorem 15.56. *Let M_1 and M_2 be two rectilinearly triangulated 2-manifolds in $\mathbb{R}^n$. Let $h : M_1 \to M_2$ be a topological homeomorphism. Then there exists a homeomorphism $g : M_1 \to M_2$ such that the image of every triangle in a triangulation of M_1 is a rectilinear triangle in M_2.*

Now we are in a position to prove that different triangulations of the same compact 2-manifold are equivalent in the sense that subdivisions of each are combinatorially identical.

Theorem 15.57. *Any two triangulations of a compact 2-manifold are equivalent.*

After you prove that a subdivision of a triangulation will yield the same answer when you compute the Euler Characteristic, you can prove that the Euler characteristic of a compact 2-manifold is well-defined.

Theorem 15.58. *The Euler characteristic is well-defined for compact 2-manifolds.*

The same insights about subdivisions and equivalence of triangulations also shows that orientability is a well-defined concept.

Theorem 15.59. *Orientability is well-defined for compact 2-manifolds.*

The long sequence of technicalities in this section finally have filled in the details in the proofs of the classification of 2-manifolds. The geometric intricacies entailed in the proofs of the above theorems show us some of the rich structure that exists in something as mundane as the plane. A few people on the planet find these details to provide a certain delight. The authors happen to be among them, but we are well aware that they are not everyone's cup of tea.

15.7 Simple Simplices; Complex Complexes; Manifold Manifolds

In this chapter we defined topological objects whose names appropriately suggest their meaning. Simplices are the simple building blocks that arise from constructing the simplest objects we can think of that are made of straight line combinations of a few points in $\mathbb{R}^n$. Complexes can become complex because we can use any number of simplices to create a complex. Any physical object we can envision is likely to be constructible as a complex. And our classification of 2-manifolds and our existence in what appears to be a 3-manifold suggests that manifolds are many and varied.

The creation of these ideas and our exploration of them followed the strategies we have come to expect. We started by thinking of simple objects and explored how they could be put together to create more complex objects. The whole of this study was motivated by looking at simple spaces with an eye toward dealing later with more complicated spaces. Complexes and manifolds are both examples of objects that we imagined because they are built from spaces that we view as natural—namely, Euclidean spaces. After defining them, we followed the usual methods of seeing their relationships and seeing how transformations behaved that respected their structure.

Complexes can be thought of as things we can build with simple pieces. We can now explore them further by taking advantage of their combinatorial virtues as we develop ideas of homology to come.

Chapter 16

Simplicial $\mathbb{Z}_2$-Homology: Physical Algebra

The fundamental group, as we have seen, is a valuable tool for understanding a topological space. The central idea of the fundamental group is to understand a space by the nature of its non-trivial loops, i.e., loops that cannot be contracted to a point. Such loops reflect the presence of a certain kind of 'hole', such as the tunnel-like hole inside a torus. Loops on the torus that wind around the tunnel are non-trivial. By contrast, the sphere $\mathbb{S}^2$ has no non-trivial loops, and therefore has a trivial fundamental group. And yet it too has a 'hole', though not the kind the doughnut has. The *higher homotopy groups*, which appear to be the natural generalizations of the fundamental group, do detect spherical holes. Unfortunately, these groups have major drawbacks: they are difficult to compute even in simple situations, and they can be bizarre—giving results that do not reflect the apparent geometry we are trying to capture. For example, unbelievably there is a homotopically non-trivial map of $\mathbb{S}^3$ into $\mathbb{S}^2$, which would seem to suggest $\mathbb{S}^2$ has a 3-dimensional 'hole' even though it is patently 2-dimensional.

In this chapter, we introduce the concept of *homology*, which associates to a space a sequence of abelian groups (one for every dimension) that detect features of spaces that can be regarded as variations of 'holes'. Homology groups lack the bizarre behavior of higher homotopy groups and are typically much easier to compute. For these reasons, homology gives us an attractive, intuitively meaningful strategy to capture geometrical differences among topological spaces.

In the preceding chapter we saw a combinatorial proof of the No Retraction Theorem, which states that there is no continuous map of a ball to its boundary that fixes the boundary. The intuition behind the No Retraction Theorem is that the boundary of a ball has a hole, while the filled in ball does not. Homology theory is especially designed to measure such distinctions, so you will see that using homology to provide an alternative proof of the No Retraction Theorem is an iconic example of an application of homology.

The applications of homology are vast. In the next chapter you will see that the ideas of homology that you develop in this chapter will enable you to prove many theorems that can rightly be viewed as among the highlights of topology.

16.1 Motivation for Homology

> **Effective Thinking Principle.** *Start with Simple Cases.* Understanding simple cases deeply is a great step toward understanding more abstract versions later.

There are many homology theories. We will begin by describing the most concrete version, which holds for simplicial complexes and in which each element of the homology groups has a clear geometric manifestation as a representative that is a subcomplex.

The motivating insight for homology theory for a space X is this: a way to detect the presence of a 'hole' in X is to find an object that surrounds it. For instance, suppose you triangulate the region between a bigger cube and a smaller cube to create a finite simplicial complex K where $|K| = [-3,3] \times [-3,3] - (-1,1) \times (-1,1)$. The 'hole' in K can be detected by a sphere, for example, the boundary of $[-2,2] \times [-2,2]$, which is not filled in, that is, it is not the boundary of any 3-manifold-with-boundary in $|K|$. On the other hand, every loop in $|K|$ is homotopically trivial in $|K|$, so the fundamental group does not detect a hole. So in this case, a 2-dimensional manifold detects the hole in $|K|$.

Although not exactly accurate (as you will soon see), a good way to start to understand homology for a space X is to view an n-manifold in X that is not the boundary of an $(n+1)$-manifold-with-boundary as capturing some geometry of X while an n-manifold that is the boundary of an $(n+1)$-dimensional manifold-with-boundary is not detecting any hole or structure. In the example above, the boundary of $[-2,2] \times [-2,2]$ is not the boundary of a ball in $|K|$, so it was identifying something important in the space, in this case, the hollowness of $|K|$.

So n-manifolds that are not the boundaries of $(n+1)$-manifolds can be viewed as detecting geometrical features analogous to 'holes' in the space. The next step is to realize that a smaller sphere around the hole in K (such as the boundary of $[-1,1] \times [-1,1]$) and a larger sphere around the hole in K (such as the boundary of $[-2,2] \times [-2,2]$) are both detecting the same hole, so they should be considered the same. In this case, the smaller sphere and the larger sphere together create the boundary of the material between the two spheres. So this example suggests the idea that two n-manifolds that together form the boundary of an $(n+1)$-manifold in a space X should be viewed as equivalent n-manifolds in X from the point of view of detecting 'holes.'

Effective Thinking Principle. *Turn Intuition into Precision.* After gaining an intuitive idea for a concept, make the ideas precise.

Our example gives an intuitive idea that actually contains all the main features of homology; however, the example is neither precise nor complete. We must now try to turn this intuition into precise definitions.

Effective Thinking Principle. *Turn Examples into General Statements.* Examples constructed to illustrate salient issues provide great lessons for creating general definitions or theorems.

Our intuition can be made clearer by examining another example. We will again see these emerging concepts physically and combinatorially, but this time we will look at the specific triangulation involved. Let's consider the concrete simplicial complex shown in Figure 16.1. It comprises two triangles (that is, two 2-simplices) $\{\sigma_1, \sigma_2\}$, seven edges (that is, seven 1-simplices) $\{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$, and five vertices (that is, five 0-simplices) $\{v_1, v_2, v_3, v_4, v_5\}$.

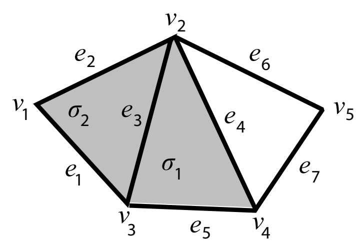


Figure 16.1: A small example of a simplicial complex.

This example is simple enough that we can look at all the 1-dimensional loops that may or may not encircle the hole. We write each of these loops as a sum rather than a union because we will soon be viewing each simplex as an element of a group, but you should think of each sum simply as a physical collection of edges. There are several loops in this example; for instance, $e_1 + e_2 + e_4 + e_5$ is one loop, $e_1 + e_2 + e_6 + e_7 + e_5$ is another loop, and $e_6 + e_7 + e_4$ is another loop.

Which ones surround 'holes' in our space? Well, $e_1 + e_2 + e_4 + e_5$ bounds a solid piece of our space, namely $\sigma_1 + \sigma_2$. Hence it doesn't encircle a hole, and it will be considered trivial, that is, equivalent to 0 in the homology group we want to define. The two loops $e_1 + e_2 + e_6 + e_7 + e_5$ and $e_6 + e_7 + e_4$ both surround the same hole, so they should be considered "equal" in our group.

In fact, these two sets of edges differ (where we take the difference mod 2, meaning we just look at the set of all the edges that are in one set, but not it in the other) by $e_1 + e_2 + e_4 + e_5$, which we already agreed should be equal to 0 since it bounds $\sigma_1 + \sigma_2$, a 2-dimensional object. Thus we should declare two cycles to be equivalent if their difference is a loop that is the boundary of a set of 2-simplexes in the space. The arithmetic of these combinatorial objects and the desire to simplify them by considering equivalence classes capture the core ideas of a homology group.

Let's look at one more loop in this example, namely, $e_1+e_2+e_3+e_6+e_7+e_4$. That set of edges is not a 1-manifold, yet it is a loop in the sense that it has no boundary. So this figure-eight shaped loop suggests that the natural objects to consider are not necessarily 1-manifolds, but instead are collections of edges that have no boundary whether they are 1-manifolds or not.

Since we are expanding our notion of what objects to view as the elements of our emerging idea of a homology group, we will use a different name for boundary-less objects—we will call them *cycles* and define them precisely below.

16.2 Chains, Cycles, Boundaries, and the Homology Groups

We can now give the exact definition of simplicial $\mathbb{Z}_2$-homology for a simplicial complex K. Actually, we define for every non-negative integer n the n^{th} $\mathbb{Z}_2$-homology group of K. The next paragraph is a one-paragraph summary. The remainder of this section pins down the details.

A collection Z of n-simplices in K such that Z has no boundary is called an n-cycle. Two n-cycles Z and W are called equivalent if $Z + W$ is the boundary of a collection of $(n+1)$-simplices of K. Each element of the n^{th} $\mathbb{Z}_2$-homology group of K is an equivalence class of n-cycles.

To pin this summary down, let's start with a simplicial complex K with a fixed triangulation. For reasons that are lost in the mists of history, we will give the name *n-chain* to any collection of n-simplices in K and we will denote such an n-chain using plus signs rather than commas or union symbols so we can think of all such n-chains as a forming a group.

Definition. An n-**chain** of K is a finite formal sum $\sum_{i=1}^{k} \sigma_i$ of distinct n-simplices in K. Note that the dimensions of the simplices must be the same. So **chain** will mean n-chain whenever the dimension is either unimportant or understood.

Recall the group $\mathbb{Z}_2 = \{0, 1\}$, the group of two elements where 0 is an additive identity and $1 + 1 = 0$. We can view an n-chain as a formal linear combination of n-simplices in K with coefficients in $\mathbb{Z}_2$ (i.e., a simplex has a coefficient of 1 if it appears in the sum and 0 if it does not). So in $\mathbb{Z}_2$-homology, we can think of an n-chain as merely a collection of n-simplices. We use the sum notation to suggest an operation on the collection of n-chains, thus making the set of n-chains into a group.

Definition. The n**-chain group of** K (with coefficients in $\mathbb{Z}_2$), denoted $\mathsf{C}_n(K)$, is the collection of n-chains in K under formal addition modulo 2. If there are no n-simplices in K, the n-chain group of K is defined to be trivial (containing the 'empty chain').

The n-chain group of K with coefficients in $\mathbb{Z}_2$ is more formally denoted $C_n(K;\mathbb{Z}_2)$, but since we focus on $\mathbb{Z}_2$-homology exclusively in this chapter, to simplify notation we omit writing the $\mathbb{Z}_2$ and use a sans serif letter to denote the chain group: $\mathsf{C}_n(K)$. Similarly, when we (soon) define the n-th homology group of K with coefficients in $\mathbb{Z}_2$, we will use a sans serif letter to denote it: $\mathsf{H}_n(K)$ rather than the more formal notation $H_n(K;\mathbb{Z}_2)$.

Exercise 16.1. *Check that* $\mathsf{C}_n(K)$ *is an abelian group.*

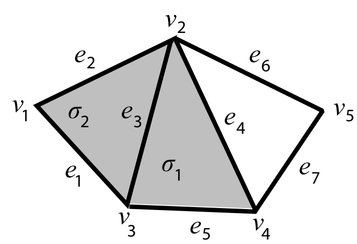


Figure 16.2: The simplicial complex from Figure 16.1, repeated.

Example. Consider the simplicial complex in Figure 16.2. Then $\mathsf{C}_2(K)$ has two generators: the simplices σ_1 and σ_2. Hence $\mathsf{C}_2(K)$ consists of four chains: the empty chain, σ_1, σ_2, and $\sigma_1+\sigma_2$. As a group, $\mathsf{C}_2(K)$ is isomorphic to $(\mathbb{Z}_2)^2$, the direct sum of two copies of $\mathbb{Z}_2$. The chain group $\mathsf{C}_1(K)$ is generated by $e_1,e_2,e_3,e_4,e_5,e_6,e_7$ and is isomorphic to $(\mathbb{Z}_2)^7$, the direct sum of seven copies of $\mathbb{Z}_2$. As an example of $\mathbb{Z}_2$ addition in this group, note that

$$(e_1+e_2+e_3)+(e_3+e_4+e_5)=e_1+e_2+e_4+e_5.$$

The chain group $\mathsf{C}_0(K)$ is generated by v_1,v_2,v_3,v_4,v_5 and is isomorphic to $(\mathbb{Z}_2)^5$.

Now let's develop a notion of the *boundary* of a chain. Since we want the boundary of an n-chain to be an $(n-1)$-chain, we will define the **boundary operator** ∂ in each dimension to be a homomorphism between chain groups:

$$\partial_n : \mathsf{C}_n(K) \to \mathsf{C}_{n-1}(K)$$

That is, the boundary operator ∂_n takes an n-chain (which is just a sum of n-simplices) and associates with it an $(n-1)$-chain (which is just a sum of $(n-1)$-simplices) that we will call the boundary of the n-chain. If you guessed at the definition of ∂_n, you would get it right. One technicality: we set $\mathsf{C}_{-1}(K) = \{0\}$ for convenience so we can define ∂_0. It will be our general practice to omit the subscript n on the '∂' symbol when it is clear from context.

Definition. The $\mathbb{Z}_2$**-boundary of an n-simplex** $\sigma = \{v_0 \cdots v_n\}$ is defined by

$$\partial\sigma = \sum_{i=0}^{n} \{v_0 \cdots \widehat{v_i} \cdots v_n\}.$$

Thus the $\mathbb{Z}_2$-boundary of the n-simplex σ is just a formal sum of the $(n-1)$-faces of σ. For a 0-simplex, the $\mathbb{Z}_2$-boundary is defined to be 0 in $\mathsf{C}_{-1}(K)$. The $\mathbb{Z}_2$**-boundary of an n-chain** is the sum of the boundaries of the simplices in the chain, thereby making ∂ linear:

$$\partial\left(\sum_{i=1}^{k} \sigma_i\right) = \sum_{i=1}^{k} \partial(\sigma_i)$$

where the sum is taken modulo 2.

Exercise 16.2. *Verify that ∂ is a homomorphism, and use the definition to compute the $\mathbb{Z}_2$-boundary of $\sigma_1 + \sigma_2$ in Figure 16.1.*

The above exercise shows that the definition of ∂ does the right thing: namely, the common edge of σ_1 and σ_2 is not counted as a piece of the Z_2-boundary of $\sigma_1 + \sigma_2$, because that common edge appears *twice* in the sum of $\partial\sigma_1$ and $\partial\sigma_2$ and in $\mathbb{Z}_2$ this coefficient becomes 0. Thus working over $\mathbb{Z}_2$ coefficients is critical. Later on, when we define homology with coefficients in $\mathbb{Z}$ we will need to alter our definition of boundary so that it still works out as it should. But since we will only work with $\mathbb{Z}_2$-boundary in this chapter, we will feel free for the remainder of this chapter to just use the word **boundary** in place of the more cumbersome $\mathbb{Z}_2$-boundary.

Definition. An n**-cycle** is an n-chain of K whose boundary is zero. The set of all n-cycles of K is denoted $\mathsf{Z}_n(K)$. An n**-boundary** is an n-chain that is the boundary of an $(n+1)$-chain of K. The set of all n-boundaries is denoted $\mathsf{B}_n(K)$.

Exercise 16.3. *Explore:*

1. *Which 2-chains of Figure 16.1 are cycles?*
2. *Which* 1*-chains of Figure 16.1 are cycles?*
3. *Which* 1*-chains of Figure 16.1 are boundaries?*
4. *Which* 0*-chains of Figure 16.1 are cycles?*

5. *Which* 0*-chains of Figure 16.1 are boundaries?*

The next theorem contains one of the most basic of all observations at the heart of homology, namely, that the boundary of any chain is a cycle.

Theorem 16.4. *Both* $\mathsf{Z}_n(K)$ *and* $\mathsf{B}_n(K)$ *are subgroups of* $\mathsf{C}_n(K)$. *Moreover,*

$$\partial \circ \partial = 0,$$

in other words, $\partial_n \circ \partial_{n+1} = 0$ *for each index* $n \geq 0$. *Hence,* $\mathsf{B}_n(K) \subset \mathsf{Z}_n(K)$.

Remember that our motivation for developing homology theory is to detect holes in a space by finding cycles that are not boundaries. Our motivation suggests that we should try to get rid of cycles that are boundaries, and count as equivalent two cycles whose difference is a boundary.

Definition. Two n-cycles α and β in K are **equivalent** or **homologous** if and only if $\alpha - \beta = \partial\gamma$ for some $(n+1)$-chain γ. In other words, α and β are homologous if they differ by an element of the subgroup $\mathsf{B}_n(K)$, and is denoted:

$$\alpha \sim_{\mathbb{Z}_2} \beta.$$

The equivalence class of α is denoted by enclosing it in brackets thusly: $[\alpha]$.

For $\mathbb{Z}_2$ n-chains, observe that $\alpha - \beta = \alpha + \beta$. (We use the minus sign in $\alpha - \beta$ in anticipation of $\mathbb{Z}$-homology to come.) So we see that two n-cycles are equivalent if together they bound an $(n+1)$-chain.

Exercise 16.5. *List all the equivalence classes of* 0*-cycles,* 1*-cycles and* 2*-cycles in the complex in Figure 16.1.*

Exercise 16.6. *List all the equivalence classes of* 0*-cycles,* 1*-cycles and* 2*-cycles in a triangulated 2-sphere with its standard triangulation as the faces of a 3-simplex.*

Definition. The n^{th}**-homology group** (with coefficients in $\mathbb{Z}_2$) of a finite simplicial complex K, denoted $\mathsf{H}_n(K)$, is the additive group whose elements are equivalence classes of cycles under the $\mathbb{Z}_2$-equivalence defined above, where $[\alpha] + [\beta] := [\alpha + \beta]$. That is,

$$\mathsf{H}_n(K) = \mathsf{Z}_n(K)/\mathsf{B}_n(K).$$

This construction is quite natural. Indeed, $\mathsf{H}_n(K)$ is the group formed when we take the collection of n-cycles of K and set to zero those that are the boundary of an $(n+1)$-chain. An n-cycle is something that potentially could detect a hole, but saying it is the boundary of an $(n+1)$-chain is tantamount to saying that the hole is 'filled' (so there is no hole). Thus the homology group is designed to consider only those n-cycles that actually detect some structure of the space.

Since K is finite, $\mathsf{H}_n(K)$ is a finite abelian group, and it is easy to see that every element (aside from the identity) has order 2. Thus by the fundamental theorem of finite abelian groups, each homology group $\mathsf{H}_n(K)$ will be isomorphic to a (finite) direct sum of copies of $\mathbb{Z}_2$.

The following exercises provide good warm-ups for understanding homology groups. If the reasoning seems tedious at first, do not worry—soon we will develop tools to compute homology groups more efficiently.

Theorem 16.7. *If K is a one-point space, $\mathsf{H}_n(K) \cong 0$ for $n \geq 0$ and $\mathsf{H}_0(K) \cong \mathbb{Z}$.*

Definition. Any space with the homology groups of a point is called **acyclic**.

Theorem 16.8. *If K is connected, then $\mathsf{H}_0(K)$ is isomorphic to $\mathbb{Z}_2$. If K has r connected components, then $\mathsf{H}_0(K)$ is isomorphic to $\mathbb{Z}_2^r$.*

Exercise 16.9. *Let K be a triangulation of a 3-dimensional ball that consists of a 3-simplex together with its faces. Compute $\mathsf{H}_n(K)$ for each n.*

Exercise 16.10. *Let K be a triangulation of a 2-sphere that consists of the proper faces of a 3-simplex. Compute $\mathsf{H}_n(K)$ for each n.*

Definitions.

1. Let K be a simplicial complex with $|K| \subset \mathbb{R}^n$. A point $x \notin K$ can **see** K if any ray from x intersects $|K|$ at most once.
2. Let K be a finite complex and x a point that sees K. If $\sigma = \{v_0 \cdots v_k\}$ is a simplex of K, define the **cone** of x over σ to be the simplex:
$$\mathsf{Cone}_x(\sigma) = \{xv_0 \cdots v_k\}.$$
3. Define $x * K$, the **cone over** K to be the simplicial complex comprising all simplices $\mathsf{Cone}_x(\sigma)$ for $\sigma \in K$, and all faces of such simplices.
4. Define the **simplicial cone operator** $\mathsf{Cone}_x : \mathsf{C}_n(K) \to \mathsf{C}_{n+1}(w * K)$ by extending the definition of $\mathsf{Cone}_x(\sigma)$ linearly to chains.

Theorem 16.11. *For x seeing K, and σ a simplex of K,*

$$\partial\, \mathsf{Cone}_x(\sigma) + \mathsf{Cone}_x(\partial\sigma) = \sigma.$$

Corollary 16.12. *For any complex K and x seeing K, the complex $x * K$ is acyclic.*

Theorem 16.13. *The complex K consisting of an n-simplex together with all its faces is acyclic.*

16.3 Induced Homomorphisms and Invariance

> **Effective Thinking Principle.** *Consider Transformations.* After defining a concept, explore how it interacts with transformations.

Now that we have seen some concrete examples of homology groups, we will show that they are indeed topological invariants for simplicial complexes, that is, if the underlying spaces of two simplicial complexes are homeomorphic, their homology groups are isomorphic. Just as we did for fundamental groups, we will accomplish this goal through the use of induced homomorphisms. That is, given a continuous map between simplicial complexes, we will create maps between their homology groups. Unfortunately, since the construction of homology groups depends heavily on triangulations and since continuous maps do not in general respect simplicial structure, the construction of the induced homomorphisms for homology is somewhat more involved than the corresponding construction for fundamental groups.

> **Effective Thinking Principle.** *Start With Simple Cases.* When creating a concept, start with simple cases and add complexity incrementally.

Fix two simplicial complexes K and L. Our first step is to consider the case where our map does happen to respect simplicial structure. In this case, the definition of the induced map on homology is straightforward.

Exercise 16.14. *Let $f : K \to L$ be a simplicial map. Carefully write out the definition of the natural induced map from n-chains of K to n-chains of L: $f_{\#n} : \mathsf{C}_n(K) \to \mathsf{C}_n(L)$ and show that it is a homomorphism.*

The map $f_{\#n}$ is called the **induced chain map**. The next exercise contains an important technicality about the induced chain map in the case where the image of an n-simplex is an $(n-1)$-simplex.

Exercise 16.15. *If the simplicial map $f : K \to L$ maps an n-simplex σ to an $(n-1)$-simplex τ, what is $f_{\#n}(\sigma)$?*

As we did with the boundary operator, we typically drop the subscript n from our notation and simply write $f_{\#}$.

The following theorem is perhaps the most important theorem in homology. It is the theorem that will be used in the proofs of most of the applications of homology theory, including the No

Retraction Theorem, the Jordan-Brouwer Separation Theorem, the Borsuk-Ulam Theorem, and many more. This theorem is summarized by the phrase: ”The boundary of the image is the image of the boundary.”

Theorem 16.16. *Let $f : K \to L$ be a simplicial map, and let $f_\#$ be the induced map $f_\# : \mathsf{C}_n(K) \to \mathsf{C}_n(L)$. Then for any chain $c \in \mathsf{C}_n(K)$,*

$$\partial(f_\#(c)) = f_\#(\partial(c)).$$

In other words, the diagram:

$$\begin{array}{ccc} \mathsf{C}_n(K) & \xrightarrow{f_\#} & \mathsf{C}_n(L) \\ \partial\big\downarrow & & \big\downarrow\partial \\ \mathsf{C}_{n-1}(K) & \xrightarrow{f_\#} & \mathsf{C}_{n-1}(L) \end{array}$$

commutes.

As you will see, certain theorems in algebraic topology can be succinctly expressed by so-called **commutative diagrams** like this one, which means that the result of following arrows is the same no matter which path is taken.

Definition. Let $f : K \to L$ be a simplicial map. The **induced homomorphism** $f_* : \mathsf{H}_n(K) \to \mathsf{H}_n(L)$ is defined by $f_*([z]) = [f_\#(z)]$.

Note that the induced homomorphism f_* is technically a sequence of maps, one for each homology group. We could have indicated this fact with a subscript, e.g., f_{*n} for the map on the n-th homology group. However, unless needed for emphasis, we omit the subscript and write f_* when the context is clear.

Of course, we have to check that f_* is well-defined, which involves basic applications of the fact that the boundary of the image of a chain is the image of the boundary of that chain.

Theorem 16.17. *Let $f : K \to L$ be a simplicial map. Then the induced homomorphism $f_* : \mathsf{H}_n(K) \to \mathsf{H}_n(L)$ is a well-defined homomorphism.*

Effective Thinking Principle. *Look at Examples to Understand Theorems.* Specific examples can help us understand the implications of theorems more deeply.

Let’s look at an example of a simplicial map with a non-trivial homology group and observe how the induced homomorphism behaves.

Exercise 16.18. *Let K be a complex comprising the proper faces of a hexagon: six edges and six vertices $v_0, \ldots, v_5$. Let L be the complex comprising the proper faces of a triangle: three edges and three vertices w_0, w_1, w_2. Let f be a simplicial map that sends v_i to $w_{(i \mod 3)}$. Compute the homology groups of K and L and describe the simplicial map f and the induced homomorphism f_*.*

Now that we can induce a homomorphism on homology from a simplicial map, we want to show that arbitrary continuous functions induce a homomorphism on homology. In the last chapter we saw how to approximate arbitrary continuous functions between the underlying sets of simplicial complexes by simplicial maps that are homotopic to our original continuous function. So our next step is to see how the steps involved in creating simplicial approximations relate to induced homology maps.

In constructing a simplicial approximation, one of the basic steps involved taking barycentric subdivisions. So we should check that if we compute homology groups using a triangulation, then we should get the same homology groups if we use the barycentric subdivision of that triangulation to compute the homology groups of our complex. Taking a barycentric subdivision gives us more n-simplices and more n-cycles, but our (correct) intuition tells us that there are no more equivalence classes of n-cycles—that is, breaking simplices into pieces does not create more holes in the space.

> **Effective Thinking Principle.** *Pin Down Intuition.* Take the trouble to pin down details that justify intuition. Either you will better understand why your intuition is correct or you will realize you are mistaken—both good outcomes.

Recall that if K is a triangulated complex, then $\mathsf{sd}\, K$ is the barycentric subdivision of K. Our next goal is to show that $\mathsf{H}_n(\mathsf{sd}\, K)$ is isomorphic to $\mathsf{H}_n(K)$. A reasonable strategy is to construct a specific simplicial map $\lambda : \mathsf{sd}\, K \to K$ and prove that $\lambda_* : \mathsf{H}_n(\mathsf{sd}\, K) \to \mathsf{H}_n(K)$ is an isomorphism. We will define a candidate map here and ask you to verify that it does what we want. The simplicial map $\lambda : \mathsf{sd}\, K \to K$ is determined by its values on its vertices and is defined as follows. Any vertex v in $\mathsf{sd}\, K$ is the barycenter of a simplex σ in K. Choose any vertex of σ and define $\lambda(v)$ to be that vertex. Notice that if v is a vertex in K, then $\lambda(v) = v$, since a vertex of K is the barycenter of itself.

Since λ is a simplicial map, λ_* is a well-defined homomorphism from $\mathsf{H}_n(\mathsf{sd}\, K)$ to $\mathsf{H}_n(K)$.

We shall prove λ_* is injective and surjective by exhibiting an inverse homomorphism from $\mathsf{H}_n(K)$ to $\mathsf{H}_n(\mathsf{sd}\, K)$. We could find such a homomorphism by finding a map that takes chains in $\mathsf{C}_n(K)$ to chains in $\mathsf{C}_n(\mathsf{sd}\, K)$ that commutes with the boundary operator. Before reading on, think about the following exercise.

Exercise 16.19. *Suggest a homomorphism from* $\mathsf{C}_n(K)$ *to* $\mathsf{C}_n(\mathsf{sd}\,K)$ *that commutes with* ∂*. Could its induced homomorphism on homology be an inverse for* λ_**?*

In doing the preceding exercise, you probably associated each n-simplex in K to a set of n-simplices that comprise it in the barycentric subdivision $\mathsf{sd}\,K$. For the record, let's describe the map from $\mathsf{C}_n(K) \to \mathsf{C}_n(\mathsf{sd}\,K)$ that does that and give it a name. Don't be intimidated by the technical notation involved in the following definition; it just does what we claim.

Definition. Define the **subdivision operator** $\mathsf{SD} : \mathsf{C}_n(K) \to \mathsf{C}_n(\mathsf{sd}\,K)$ by first defining SD on a simplex:

$$\mathsf{SD}(\{v_0 \cdots v_n\}) = \sum_{\pi \in S_{n+1}} \{b_0^\pi \cdots b_n^\pi\}$$

where b_k^π is the barycenter of the face $\{v_{\pi(0)} \cdots v_{\pi(k)}\}$, and $\pi \in S_{n+1}$ is a permutation of $\{0, 1, \ldots, n\}$. Then extend SD linearly to define it on n-chains. Thus SD is a homomorphism that sends an n-simplex to the formal sum of the n-simplices in its barycentric subdivision.

The subdivision operator is not a simplicial map since it takes a single simplex in K to a chain containing many simplices in $\mathsf{sd}\,K$. Nevertheless, as is the case with simplicial maps, the image of the boundary is the boundary of the image.

Theorem 16.20. *The subdivision operator commutes with the boundary operator, that is, if* c *is a chain in* K*, then* $\mathsf{SD}(\partial c) = \partial\,\mathsf{SD}(c)$*.*

So we see how an n-cycle in K that bounds an $(n+1)$-chain in K corresponds to an n-cycle in $\mathsf{sd}\,K$ that bounds an $(n+1)$-chain in $\mathsf{sd}\,K$. This fact will come in handy.

As a result of this theorem, there is a natural induced homomorphism

$$\mathsf{SD}_* : \mathsf{H}_n(K) \to \mathsf{H}_n(\mathsf{sd}\,K).$$

We claim that SD_* and λ_* are inverses and the next two exercises will help verify this fact.

Exercise 16.21. *Show that* $\lambda_\# \circ \mathsf{SD} = \mathsf{id}$*, the identity map on* $C_n(K)$*, and therefore* $\lambda_* \circ \mathsf{SD}_* = \mathsf{id}_*$*, the identity map on* $H_n(K)$*.*

To do the preceding exercise, proceed simplex by simplex.

Exercise 16.22. *Show that* $\mathsf{SD} \circ \lambda_\#$ *and* id*, the identity map on* $C_n(\mathsf{sd}\,K)$*, induce the same homomorphism on homology.*

A moment's reflection reveals why $\mathsf{SD} \circ \lambda_\#$ and id should, intuitively speaking, induce the same homomorphism on homology: if you think of the underlying maps for SD and id as the identity on $|\operatorname{sd} K| = |K|$, then the underlying maps $\operatorname{id}_{|K|} \circ \lambda$ and $\operatorname{id}_{|\operatorname{sd} K|}$ are homotopic. (In fact, via a straight line homotopy.) So if c is a cycle in $\operatorname{sd} K$, the homotopy between image cycles $\mathsf{SD} \circ \lambda_{\#}(c)$ and $\operatorname{id}(c)$ should, intuitively speaking, sweep out a simplicial chain between the two image cycles. Unfortunately, a homotopy, while continuous, is not necessarily a simplicial map.

So we adopt another strategy. We will attempt to construct a chain $D(z)$ whose boundary consists of the image cycles $\mathsf{SD} \circ \lambda_{\#}(z)$ and $\operatorname{id}(z)$. Rather than globally trying to find $D(z)$ for every cycle z, we will instead try to find such chains $D(c)$ for every chain c. Let's start by defining $D(\sigma)$ for every simplex σ, and then extending linearly to produce the definition of D acting on a chain. We cannot expect the boundary of $D(c)$ to be just $\mathsf{SD} \circ f_{\#}(c)$ and $\operatorname{id}(c)$, however. The boundary of $D(c)$ in general may involve ∂c. But that will vanish if c is a cycle.

Exercise 16.23. *If $\sigma \in \operatorname{sd} K$ is contained in $\tau \in K$, then $\mathsf{SD} \circ \lambda_{\#}(\sigma)$ and $\operatorname{id}(\sigma)$ both lie inside τ.*

Because of the previous exercise, the possibility thus opens for us to define $D(\sigma)$ locally inside τ. Start by defining $D(v)$ for v a vertex. Then inductively define D dimension by dimension. If you're really stuck, you may gather some hints by seeing the discussion in Section 19.3.

Theorem 16.24. *Let K be a simplicial complex. Then $\mathsf{H}_n(K)$ is isomorphic to $\mathsf{H}_n(\operatorname{sd} K)$. In fact, if the simplicial map $\lambda : \operatorname{sd} K \to K$ is defined by taking each vertex in $\operatorname{sd} K$ to any vertex of the simplex in K of which it is the barycenter, then*

$$\lambda_* : \mathsf{H}_n(K) \to \mathsf{H}_n(\operatorname{sd} K)$$

is an isomorphism. Also, the induced homomorphism of the subdivision operator

$$\mathsf{SD}_* : \mathsf{H}_n(K) \to \mathsf{H}_n(\operatorname{sd} K)$$

is an isomorphism and is the inverse of λ_*.

Now we can define the induced homomorphism for an arbitrary continuous function.

Definition. Let K and L be simplicial complexes, and let $f : |K| \to |L|$ be a continuous function. Let $g : \operatorname{sd}^m K \to L$ be a simplicial approximation to f, and let SD be the subdivision operator. Then the **induced homomorphism**

$$f_* : \mathsf{H}_n(K) \to \mathsf{H}_n(L)$$

is defined by $f_* = g_* \circ (\mathsf{SD}_*)^m$.

We need to check that f_* is well-defined: that is, it doesn't depend on the choice of simplicial approximation g.

Theorem 16.25. *Let $\mathsf{sd}^\ell K$ and $\mathsf{sd}^m K$ be barycentric subdivisions of K. Suppose $g : \mathsf{sd}^\ell K \to L$ and $h : \mathsf{sd}^m K \to L$ are simplicial approximations to a continuous function $f : |K| \to |L|$. Then $g_* \circ \mathsf{SD}_*^\ell : \mathsf{H}_n(K) \to \mathsf{H}_n(L)$ is the same homomorphism as $h_* \circ \mathsf{SD}_*^m$.*

First check when $\ell = m$ that g and h induce the same homomorphism. You will use similar ideas as you used in Exercise 16.22 to construct a chain "between" $g_\#$ and $h_\#$. Then, if ℓ and m are different, note that $g_* \circ \mathsf{SD}_*^\ell = g_* \circ (\lambda_*^{-1})^\ell$.

We have thus achieved our goal of taking a continuous map and inducing a map on homology.

Now we can use our induced homomorphism to prove important key results about homology. They show that our induced maps on the homology groups have the same functorial properties as did our induced maps on the fundamental group.

Lemma 16.26. *If K, L, and M are simplicial complexes and $f : |K| \to |L|$ and $g : |L| \to |M|$ are continuous maps, then $(g \circ f)_* = g_* \circ f_*$.*

Lemma 16.27. *If $i : |K| \to |K|$ is the identity map, then i_* is the identity homomorphism on each homology group.*

Homeomorphic spaces had better have isomorphic homology, and indeed they do.

Theorem 16.28. *Let K and L be simplicial complexes. If $f : |K| \to |L|$ is a homeomorphism, then f induces an isomorphism between the $\mathbb{Z}_2$-homology groups of K and L.*

We have thus successfully shown that homology is a topological invariant. Notice that, in particular, the isomorphism classes of a complex's homology groups only depend on the underlying space, not on the particular triangulation. Hence, we are justified in using the notation $H_n(K)$ without reference to that particular triangulation involved. Of course, if we want to actually exhibit elements of a homology group, we need to choose a triangulation.

Not only do homeomorphic spaces have the same homology groups, but any homotopy equivalent spaces have isomorphic homology groups.

Theorem 16.29. *Let K and L be simplicial complexes. If $f : |K| \to |L|$ is a homotopy equivalence, then f induces an isomorphism between the $\mathbb{Z}_2$-homology groups of K and L.*

Corollary 16.30. *If K is a strong deformation retract of L. Then K and L have isomorphic $\mathbb{Z}_2$-homologies.*

16.4 The Mayer-Vietoris Theorem

Effective Thinking Principle. *Pieces Make Wholes.* If you understand basic pieces and you understand what happens when pieces are combined, you can understand complicated wholes.

In this section we will prove the Mayer-Vietoris Theorem, which will allow us to break a complex into smaller subcomplexes to compute its homology. In this sense, the Mayer-Vietoris Theorem is the homological analogue to Van Kampen's Theorem for fundamental groups.

Definition. If K is a simplicial complex, a **subcomplex** is a simplicial complex L such that $L \subset K$.

In other words, to get a subcomplex of K we choose some of the simplices in K in such a way that the resulting space is still a complex (that is, if we choose a simplex, we also need to choose its faces).

Exercise 16.31. *If K is a finite simplicial complex, verify that the intersection of two subcomplexes of K is a subcomplex.*

Now imagine that we break up a simplicial complex K into two subcomplexes A and B. The question before us is how the homology groups of K are related to the homology groups of the pieces. We should also expect to include information about how A and B overlap. Thus we should look for relationships among the cycles of K, A, B, and $A \cap B$.

Now is a good time to draw pictures!

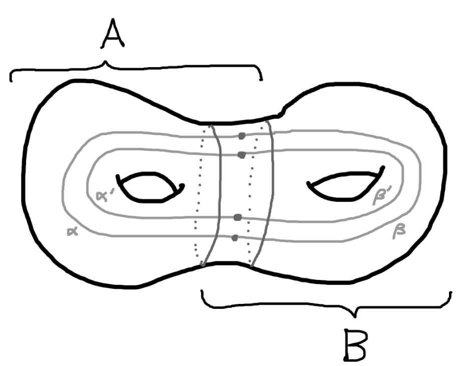


Figure 16.3: Overlapping subcomplexes A and B of a complex K, and cycles broken into chains in each side.

> **Effective Thinking Principle.** *Draw a Picture.* If we haven't said it enough, drawing pictures is an excellent way to build intuition. Choose examples that illustrate various possibilities.

For instance, you might draw examples where the intersection of A and B produces cycles that are homologous to zero in one side but not the other. You might also look at examples in various dimensions.

A first observation starts with cycles in $A \cap B$ and considers what can happen when you view them in the other pieces. The following exercise asks a series of questions that highlights an important feature of discovery.

> **Effective Thinking Principle.** *Invent your own Questions.* As you get used to thinking like an explorer, you will come up with your own questions, whose answers may be interesting and delightful.

Exercise 16.32. *Note that a cycle in $A \cap B$ is still a cycle in A, B, and K. Then answer:*

1. *Can a trivial cycle in $A \cap B$ be non-trivial in A?*

2. *Can a non-trivial cycle in $A \cap B$ be trivial in A?*

3. *Can a non-trivial cycle in $A \cap B$ that's also non-trivial in A and in B be trivial in K?*

A second observation starts with a cycle in A, and asks if, in K, that cycle in A is homologous to a cycle in B, and, if so, is there a cycle in their intersection that is homologous to both?

Theorem 16.33. *Let K be a finite simplicial complex and A and B be subcomplexes such that $K = A \cup B$. If α, β are k-cycles in A and B respectively, and if $\alpha \sim_{\mathbb{Z}_2} \beta$ in K, then there is a k-cycle c in $A \cap B$ such that $\alpha \sim_{\mathbb{Z}_2} c$ in A and $\beta \sim_{\mathbb{Z}_2} c$ in B.*

A third observation starts with cycles in K and considers what can be said when you look at their parts in A and in B.

Theorem 16.34. *Let K be a finite simplicial complex and A and B be subcomplexes such that $K = A \cup B$. Let z be a k-cycle in K. Then there exist k-chains α and β in A and B respectively such that:*

1. *$z = \alpha + \beta$ and*

2. *$\partial\alpha = \partial\beta$ is a $(n-1)$-cycle c in $A \cap B$.*

Furthermore, if $z = \alpha' + \beta'$, a sum of n-chains in A and B respectively, and $c' = \partial\alpha' = \partial\beta'$ is a $(n-1)$-cycle, then c' is homologous to c in $A \cap B$.

There are some natural homomorphisms between various homology groups.

Exercise 16.35. *Let K be a simplicial complex and A and B be subcomplexes such that $K = A \cup B$. Construct natural homomorphisms ϕ, ψ, δ among the groups below and show that $\psi \circ \phi = 0$ and $\delta \circ \psi = 0$.*

1. $\phi : \mathsf{H}_n(A \cap B) \to \mathsf{H}_n(A) \oplus \mathsf{H}_n(B)$.

2. $\psi : \mathsf{H}_n(A) \oplus \mathsf{H}_n(B) \to \mathsf{H}_n(K)$.

3. $\delta : \mathsf{H}_n(K) \to \mathsf{H}_{n-1}(A \cap B)$.

Van Kampen's Theorem was phrased in terms of quotient groups of free products. The analogous theorem in homology is written in terms of *exact sequences*.

Definition. Given a (finite or infinite) sequence of groups and homomorphisms:

$$\cdots \to G_{i-1} \xrightarrow{\phi_{i-1}} G_i \xrightarrow{\phi_i} G_{i+1} \to \ldots$$

the sequence is **exact at** G_i if and only if $\operatorname{Im} \phi_{i-1} = \operatorname{Ker} \phi_i$. The sequence is called an **exact sequence** if and only if it is exact at each group (except at the first and last groups if they exist).

Theorem 16.36 ($\mathbb{Z}_2$ Mayer-Vietoris)**.** *Let K be a finite simplicial complex and A and B be subcomplexes such that $K = A \cup B$. The sequence*

$$\cdots \to \mathsf{H}_n(A \cap B) \to \mathsf{H}_n(A) \oplus \mathsf{H}_n(B) \to \mathsf{H}_n(K) \to \mathsf{H}_{n-1}(A \cap B) \to \ldots$$

using the homomorphisms ϕ, ψ, δ above, is exact.

In using the Mayer-Vietoris sequence it is helpful to record some facts about exact sequences that come in handy.

Exercise 16.37. *Let C, D, E be groups, and arrows represent homomorphisms below.*

1. $0 \to C \xrightarrow{\phi} D$ *is exact at C if and only if ϕ is one-to-one.*

2. $D \xrightarrow{\psi} E \to 0$ *is exact at E if and only if ψ is onto.*

3. $0 \to C \xrightarrow{\phi} D \to 0$ *is exact if and only if ϕ is an isomorphism.*

When applying Mayer-Vietoris to a decomposition of a complex K into parts A and B, do not forget that you can also use results like Theorem 16.30 to simplify homology calculations for A and B.

Exercise 16.38. *Compute the $\mathbb{Z}_2$-homology groups for each complex K below.*

1. *The bouquet of k circles (the union of k circles identified at a point).*
2. *A wedge of a 2-sphere and a circle (the two spaces are glued at one point).*
3. *A 2-sphere union its equatorial disk.*
4. *A double solid torus.*

Exercise 16.39. *Compute the $\mathbb{Z}_2$-homology groups of a torus using Mayer-Vietoris in two different ways (with two different decompositions).*

Exercise 16.40. *Use the Mayer-Vietoris Theorem to compute $\mathsf{H}_n(M)$ for every compact, triangulated 2-manifold M. What compact, triangulated 2-manifolds are not distinguished from one another by $\mathbb{Z}_2$-homology? What does $\mathsf{H}_2(M)$ tell you?*

Exercise 16.41. *Let $p, q \in \mathbb{Z}$ be relatively prime. Calculate $\mathsf{H}_n(L(p,q))$, the homology of the lens space $L(p,q)$.*

Exercise 16.42. *Use the Mayer-Vietoris Theorem to compute $\mathsf{H}_n(K)$ for the complexes K pictured in Figure 16.4.*

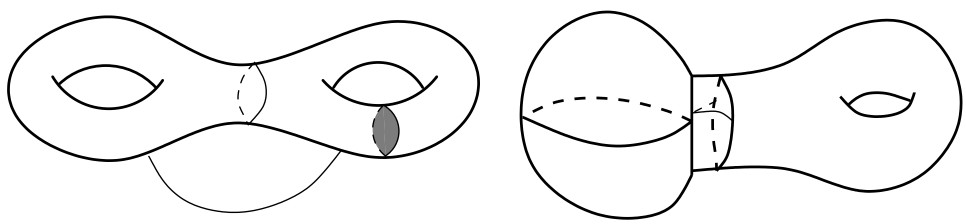


Figure 16.4: Two interesting spaces.

Exercise 16.43. *Use the Mayer-Vietoris Theorem to find the Z_2-homology groups for each of the following spaces.*

1. $\mathbb{S}^n$.
2. *A cone over a finite simplicial complex K.*
3. *A suspension over a finite simplicial complex K (that is, the finite simplicial complex created by gluing two cones over K along K).*
4. $\mathbb{RP}^n$ *(which is $\mathbb{S}^n$ with antipodal points identified).*

16.5 Introduction to Cellular Homology

We have now defined the $\mathbb{Z}_2$-homology groups, and in small examples we have seen how to compute them. But you can anticipate how unwieldy a computation would become if the triangulation of our space were large. Is there an easier way?

We can define another kind of homology theory called **cellular homology**, that can be easier to compute since the associated chain groups will have far fewer generators. Instead of decomposing a triangulated space into simplices, we will try to decompose it into fewer and larger *cells*, which are basically unions of simplices. The boundary maps will still be of interest, so we must decompose in a way that ensures that boundaries of cells remain unions of other cells. Fewer cells means fewer homology computations. Of course, we will need to show that cellular homology is isomorphic to simplicial homology.

Cellular homology begins with an existing triangulation, so we will need the notion of the interior of a simplex.

Definition. Given an n-simplex σ, the interior of σ, denoted $\mathrm{Int}(\sigma)$, is the set of all points of σ that do not lie on a proper face of σ.

Of course, the interior of an n-simplex σ coincides with its topological interior when σ is embedded in $\mathbb{R}^n$.

Theorem 16.44. *Let K be a simplicial complex where $T = \{\sigma_i\}_{i=1}^k$. Then*

$$K = \bigsqcup_{i=1}^{k} \mathrm{Int}(\sigma_i),$$

where $\sqcup$ denotes disjoint union.

Definition. Let K be a simplicial complex. An **open n-cell** of K is a set σ that is the (disjoint) union of interiors of simplices of K (not necessarily all of dimension n) such that σ is homeomorphic to an open n-ball. By convention, an 'open 0-ball' is a point, so a 0-cell is the same as a 0-simplex. We sometimes say **open cell** when the dimension of the cell is understood. Every open n-cell σ in K has an **associated $\mathbb{Z}_2$ n-chain** in $\mathsf{C}_n(K)$: namely the sum of all the n-simplices of K whose interiors are included in the n-cell σ.

We are now ready to specify what we mean by an open cell decomposition.

Definition. Let K be a simplicial complex. A $\mathbb{Z}_2$ **open cell decomposition** of K is a collection K^c of subsets of $|K|$ which satisfies the following conditions.

1. Every set in K^c is an open cell of K.

2. $|K|$ is the disjoint union of the sets in K^c.

3. If $c \in K^c$ is an open n-cell and c' is the associated $\mathbb{Z}_2$ n-chain, then the boundary $\partial c' = \sum_i b_i'$ where each b_i' is the $(n-1)$-chain associated with some $(n-1)$-cell b_i.

Then K^c will be called a $\mathbb{Z}_2$ **cellular complex**.

You can think of a cellular complex as being built up, successively by dimension, by attaching cells of dimension n to a skeleton of cells of dimension up to $(n-1)$. Condition (3) says that the boundary of a n-cell cannot meet the $(n-1)$-cells along partial pieces of cells.

Example. Let K be the proper faces (vertices and edges) of a square. Thus $|K|$ is a simple closed curve. Then one vertex and one open 1-cell would form a open cell decomposition of K.

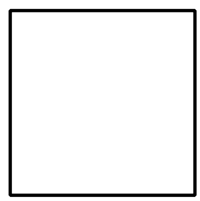


Figure 16.5: A 1-cycle with four edges.

Exercise 16.45. *Let K be a 3-simplex with triangulation shown (a tetrahedron). Find an open cell decomposition of K with one vertex, one open 2-cell, and one open 3-cell. This example shows that it is not necessary to have every dimension less then the dimension of K represented.*

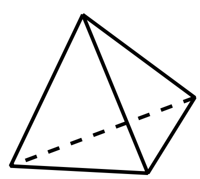


Figure 16.6: A 2-cycle with four faces.

Example. Let K be the 2-complex created by starting with a triangle and attaching a disk whose boundary goes around it twice. We could look at this complex as a quotient space of a 6-sided triangulated disk with opposite edges identified with arrows all going clockwise around the hexagon. It could exist in $\mathbb{R}^4$. Then an open cell decomposition of K could consist of one open 2-cell, one open 1-cell, and one vertex as suggested in Figure 16.7.

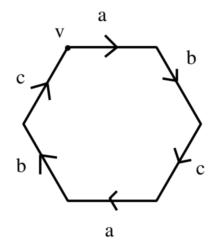


Figure 16.7: A "six sided" two cycle.

We can now define homology for a cellular complex. Our construction will essentially mirror that for simplicial complexes. We will fix a simplicial complex K and a corresponding open cell decomposition K^c.

Definition. A $\mathbb{Z}_2$ **cellular** n**-chain** in K^c is a formal sum, with coefficients in $\mathbb{Z}_2$, of open n-cells in K^c. The collection of such chains forms a group, denoted by $\mathsf{C}_n^c(K^c)$.

As before, we need a notion of boundary. You should try out a few examples to see that the next definition is natural.

Definition. Let $c \in K^c$ be an open n-cell. Let c' be the associated $\mathbb{Z}_2$ n-chain. By assumption, the $\mathbb{Z}_2$-boundary of c' is $\partial c' = \sum b_i'$, where b_i' is an $(n-1)$-chain associated to some open $(n-1)$-cell b_i in S. So we define the $\mathbb{Z}_2$**-boundary of the open cell** c to be $\partial c = \sum b_i$. The $\mathbb{Z}_2$**-boundary of a cellular chain** is defined by extending linearly: the boundary of a sum of open cells is the sum of the boundary of those open cells.

Definition. A $\mathbb{Z}_2$ **cellular boundary** is a $\mathbb{Z}_2$ cellular n-chain that bounds some cellular $(n+1)$-chain. The collection of $\mathbb{Z}_2$ cellular boundaries is denoted $\mathsf{B}_n^c(K^c)$. A $\mathbb{Z}_2$ **cellular** n**-cycle** is a cellular n-chain with zero boundary. The collection of $\mathbb{Z}_2$ cellular n-cycles is denoted $\mathsf{Z}_n^c(K^c)$.

Theorem 16.46. *Both $\mathsf{Z}_n^c(K^c)$ and $\mathsf{B}_n^c(K^c)$ are subgroups of $\mathsf{C}_n^c(K^c)$. Moreover, $\mathsf{B}_n^c(K^c) \subset \mathsf{Z}_n^c(K^c)$.*

We can now define the homology of a cellular complex.

Definition. The $\mathbb{Z}_2$ **cellular** n^{th}**-homology group** of K^c, denoted $\mathsf{H}_n^c(K^c)$ is defined by

$$\mathsf{H}_n^c(K^c) = \mathsf{Z}_n^c(K^c)/\mathsf{B}_n^c(K^c).$$

Finally, we confirm that cellular homology is the same as simplicial homology. The argument may feel familiar if you wrestled with the subdivision operator from a prior section.

Theorem 16.47. *Let K^c be an open cell decomposition of the finite simplicial complex K. Then for each n, the obvious homomorphism $\mathsf{H}_n^c(K^c) \to \mathsf{H}_n(K)$ is an isomorphism.*

In view of this isomorphism, we hereafter drop the superscript on the homology notation and make no distinction between cellular and simplicial homology.

Exercise 16.48. *For each space below, describe a triangulation K and an open cell decomposition K^c. Then use cellular homology to compute $\mathsf{H}_n(K)$ for each n:*

1. *The sphere.*
2. *The torus.*
3. *The projective plane.*
4. *The Klein bottle.*
5. *The double torus.*
6. *Any compact, connected, triangulated 2-manifold.*
7. *The Möbius band.*
8. *The annulus.*
9. *Two (hollow) triangles joined at a vertex.*

Use cellular homology to answer the following questions.

Exercise 16.49. *What is $\mathsf{H}_n(\mathbb{S}^k)$ for $n = 0, 1, 2, \ldots$ and $k = 0, 1, 2, \ldots$?*

Exercise 16.50. *What is $\mathsf{H}_n(\mathbb{T})$ for $n = 0, 1, 2, \ldots$ for a solid torus $\mathbb{T}$?*

16.6 Homology is Easier Than It Seems

The problem with homology is that it is easy to get lost in the woods even though the forest is beautiful and easy to appreciate. We had a great, intuitive idea that something that encircles a hole should be noted and if two such things encircle the same hole, they should be equal. What a simple idea. What happened to that simplicity during the last thirty pages or so? Those pages were not only numerous, they were also full of annoying details. What can we do? How can we relate to this stuff?

The answer is to get to the point where whole chunks become clear and simple. Yes, there are details, but if you can realize that all those details arise from following a rather simple path that all works out, then you can keep the whole outline clearly in mind and you can work your way through the details whenever you need to.

The great thing is that $\mathbb{Z}_2$-homology really is clear and simple. Holes are detected by n-cycles that are physical objects that surround those holes in some sense. Induced maps turn out to be maps that are the first and only thing you can think of as a way to correspond a cycle in the domain with a cycle in the range. It all works out smoothly.

The true payoff is contained in the next chapter. There you will see that $\mathbb{Z}_2$-homology is precisely what you need to deduce some of the classical results in topology. Enjoy them.

Chapter 17

Applications of $\mathbb{Z}_2$-Homology: A Topological Superhero

Many fundamental and important theorems in the field of algebraic topology can be proved using $\mathbb{Z}_2$-homology. There is something extremely delightful about the fact that difficult-sounding and historically challenging mathematical insights can be established using the clear geometric insights that $\mathbb{Z}_2$-homology captures.

In this chapter you will have the unalloyed joy in proving theorems that are among the highlights of topology as we celebrate the consequences that flow from applying $\mathbb{Z}_2$-homology.

In fact, the workhorse theorem about $\mathbb{Z}_2$-homology that is used most is the basic fact summarized by saying that for induced maps on $\mathbb{Z}_2$-homology, the boundary of the image is the image of the boundary (Theorem 16.16). Perhaps even more specifically, often what is used is the special case of that theorem that for induced maps, if you start with a cycle that bounds, then the image of that cycle must also bound.

17.1 The No Retraction Theorem

This first theorem uses the fact that the boundary of an n-manifold is an $(n-1)$-cycle that is the $\mathbb{Z}_2$ boundary of the n-manifold itself.

Theorem 17.1 (No Retraction Theorem). *Let M^n be a connected triangulated n-manifold with $\partial M^n \neq \varnothing$. Then there is no retraction $r : M^n \to \partial M^n$, i.e., no continuous function $r : M^n \to \partial M^n$ such that for each $x \in \partial M^n$, $r(x) = x$.*

Part of the intuition of this theorem is that such a retraction cannot exist because you somehow have to puncture the inside of M^n in order for it to be mapped to its boundary, and this puncturing process would not be continuous. Anytime you are thinking of punching a hole in something you should suspect that homological reasoning may be illuminating.

In this case, if there were such a retraction r, what could you infer about the induced homomorphism r_* going from the $(n-1)$-dimensional $\mathbb{Z}_2$-homology group of M^n to the $(n-1)$-dimensional $\mathbb{Z}_2$-homology group of ∂M^n?

17.2 The Brouwer Fixed Point Theorem

Theorem 17.2 (n-dimensional Brouwer Fixed Point Theorem)**.** *Let $\mathcal{B}^n$ be the n-dimensional ball. For every continuous function $f : \mathcal{B}^n \to \mathcal{B}^n$ there exists a point $x \in \mathcal{B}^n$ such that $f(x) = x$.*

In Chapter 15, we saw that one way to prove the Brouwer Fixed Point Theorem is to prove that it is equivalent to the No Retraction Theorem for a ball. That is, instead of proving the Brouwer Fixed Point Theorem directly, we prove that it is equivalent to the No Retraction Theorem for an n-ball. Recall that to prove the equivalence of those two statements, you need to answer two questions:

1. Suppose you were given a retraction from a ball to its boundary. Then how could you use that map to construct a fixed point free map from the ball to itself?

2. Suppose you were given a fixed point free map from the ball to itself. Then how could you use that map to produce a retraction from the ball to its boundary?

Since $\mathbb{Z}_2$-homology provided a great proof of the No Retraction Theorem, then the equivalence above proves the Brouwer Fixed Point Theorem as well.

17.3 The Borsuk-Ulam Theorem

The Borsuk-Ulam Theorem has the flavor of a fixed point theorem. A physical version of it in dimension 2 would state that if you smash a beach ball on the pavement, some pair of antipodal points (spherically opposite points) on the beach ball must get smashed right on top of one another. We will state the Borsuk-Ulam Theorem formally later. Now we will prove it through a sequence of preliminary results.

The first lemma boils down to recognizing that a connected n-manifold has exactly one $\mathbb{Z}_2$ n-cycle, namely, itself.

Lemma 17.3. *Let M^n be a triangulated, connected n-manifold. Let $f : M^n \to M^n$ be a simplicial map. Then $f_* : \mathsf{H}_n(M^n) \to \mathsf{H}_n(M^n)$ is surjective if and only if $f_\#(M^n) = M^n$.*

The next theorem basically asserts that antipode preserving maps of $\mathbb{S}^1$ must be onto from a $\mathbb{Z}_2$ perspective. Note that for a point x on a sphere $\mathbb{S}^n$, the antipodal point is denoted $-x$.

Theorem 17.4. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an antipode preserving continuous map (that is, for every $x \in \mathbb{S}^1$, $f(-x) = -f(x)$. Then $f_* : \mathsf{H}_1(\mathbb{S}^1) \to \mathsf{H}_1(\mathbb{S}^1)$ is surjective.*

Your next challenge is to use the preceding theorem to prove the Borsuk-Ulam Theorem in dimension 2 (stated next). Try to show that if there existed a counterexample to the Borsuk-Ulam Theorem in dimension 2, then there would be a continuous, antipode preserving function from the whole 2-sphere to $\mathbb{S}^1$. In particular, the equator would go to $\mathbb{S}^1$ in an antipode preserving way. Recall that the equator bounds a disk in the 2-sphere and then see why that fact creates a problem with the previous theorem when you look at the induced homomorphism on the first homology group $\mathsf{H}_1(\mathbb{S}^2)$.

Theorem 17.5 (Borsuk-Ulam Theorem for $\mathbb{S}^2$)**.** *Let $f : \mathbb{S}^2 \to \mathbb{R}^2$ be a continuous map. Then there exists an $x \in \mathbb{S}^2$ such that $f(-x) = f(x)$.*

The above strategy for proving the Borsuk-Ulam Theorem in dimension 2 can be extended to work in all dimensions. Start by seeing whether you can use the truth of the antipode-preserving map theorem in one dimension to prove it in the next dimension. In other words, see whether you can prove the following theorem by induction on n.

Theorem 17.6. *Let $f : \mathbb{S}^n \to \mathbb{S}^n$ be an antipode preserving map (that is, for every $x \in \mathbb{S}^n$, $f(-x) = -f(x)$). Then $f_* : \mathsf{H}_n(\mathbb{S}^n) \to \mathsf{H}_n(\mathbb{S}^n)$ is surjective.*

The n-dimensional Borsuk-Ulam Theorem follows.

Theorem 17.7 (Borsuk-Ulam)**.** *Let $f : \mathbb{S}^n \to \mathbb{R}^n$ be a continuous function. Then there is an $x \in \mathbb{S}^n$ such that $f(-x) = f(x)$.*

17.4 The Ham Sandwich Theorem

The Ham Sandwich Theorem has a name that suggests nourishment for the body, while its proof certainly provides nourishment for the mind. The fanciful name arises from a manifestation of the theorem when cutting a sandwich. Namely, suppose you have a sandwich consisting of two pieces of awkwardly shaped bread and one piece of ham. Is it possible to cut the sandwich into two pieces with a single flat cut of a knife such that all three objects are cut exactly into two pieces of equal volume? In other words, can you find a single, flat plane simultaneously bisecting all three masses?

The statement of the Ham Sandwich Theorem contains the words 'measurable set,' but do not panic or worry about a definition, just think 'sets that have a volume.'

The Ham Sandwich Theorem appears in this chapter about applications of $\mathbb{Z}_2$-homology, although the proof you will supply might not directly mention $\mathbb{Z}_2$-homology. The totality of your proof might just show that the Ham Sandwich Theorem is equivalent to the Borsuk-Ulam Theorem, which you just proved—using $\mathbb{Z}_2$-homology.

As you seek to prove the equivalence of the Borsuk-Ulam Theorem and the Ham Sandwich Theorem, you might keep several ideas in mind. First, it might be easier to show the equivalence between the n-dimensional Ham Sandwich Theorem and the $(n-1)$-dimensional Borsuk-Ulam Theorem (although you may find a different equivalence). Second, if you are given a hyperplane H in $\mathbb{R}^n$, how many hyperplanes parallel to H will cut A_1 in half? If there is more than one such parallel plane, can you think of a natural choice of a parallel plane in that family? And, third, for each point on the unit $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$, can you think of a way to associate a family of parallel hyperplanes with that point?

Theorem 17.8 (Ham Sandwich Theorem)**.** *Let $A_1, A_2, ..., A_n$ be measurable sets of finite measure in $\mathbb{R}^n$. Then there exists an $(n-1)$-dimensional hyperplane H in $\mathbb{R}^n$ that simultaneously cuts each A_i in half.*

17.5 Invariance of Domain

Some mathematical facts seem obvious. The problem is that some of those that seem 'obvious' turn out to be untrue. Here is a pictorial example. Consider the two pictures in Figure 17.1. Each represents an embedding of a double torus into $\mathbb{R}^3$. It seems obvious that you could not distort one figure by just making elastic moves, like a movie, and take the object in the left hand picture and make it look like the object in the right hand picture. However, it is actually possible.

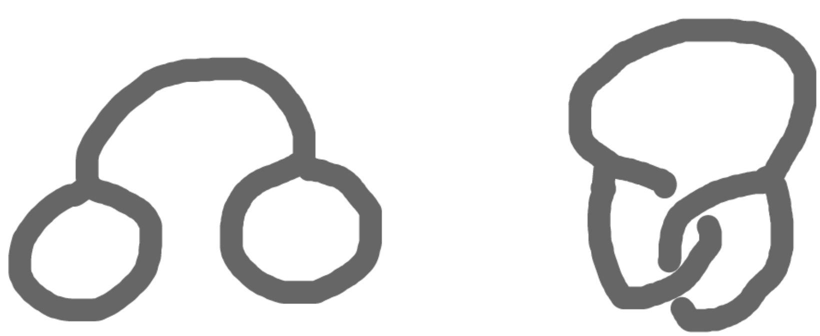


Figure 17.1: These are double tori, and you can get from one to the other with no cutting or gluing.

Exercise 17.9. *Draw a sequence of pictures to demonstrate a sequence of elastic moves with no tricks, no cutting, and no gluing that takes the left hand picture of Figure 17.1 and turns it into the right hand picture.*

This example and many others keep us alive to the reality that the world, particularly the world of topology, is not always as it seems.

This introduction to this section is intended to suggest that it is important to prove some apparently obvious facts. One of those facts, which actually is true, is the theorem that goes by the name of Invariance of Domain. You will find that $\mathbb{Z}_2$-homology provides just the right tools for confirming and proving this result.

Theorem 17.10. *If $m \neq n$ then $\mathbb{R}^m$ is not homeomorphic to $\mathbb{R}^n$.*

This theorem can be stated in a bit more generality.

Theorem 17.11 (Invariance of Domain or Invariance of Dimension)**.** *A space cannot be both an n-manifold and an m-manifold if $n \neq m$.*

17.6 An arc does not separate the plane

A topologically embedded arc in the plane (that is, an embedding of $[0, 1]$ in $\mathbb{R}^2$) can have infinitely many bumps and spirals. The next theorem and exercises ask you to create some examples of exotic embeddings of an interval into the plane so open your mind to the potential variety in embeddings of an arc. Recall that an embedding of X into Y is a continuous map $X \to Y$ that is homeomorphic onto its image. So X has the same topology as that inherited as a subspace of Y.

In this section, we will eventually prove that no embedded arc in the plane separates the plane. You may recall that we actually suggested a proof of this theorem in the penultimate section of the chapter on the classification of 2-manifolds. In this section you will see how naturally this result flows from $\mathbb{Z}_2$-homology.

> **Effective Thinking Principle.** *Gain Experience—Explore Examples.* Give yourself experience with ideas by constructing and investigating examples with unusual features.

We begin our exploration of embedded arcs in the plane by looking at some examples. One source of such embeddings arises from looking at graphs of continuous functions.

Definition. Let $f : X \to Y$ be a function. Then the **graph** of f, denoted G_f, is $\{(x, f(x))\}_{x \in X}$. Notice that $G_f \subset X \times Y$.

Theorem 17.12. *Let $f : [0,1] \to \mathbb{R}^2$ be a continuous function. Then the graph of f is an embedding of $[0,1]$ into the plane.*

The continuous functions that might first come to mind are functions such as polynomials or trigonometric or exponential functions. However, those examples are actually anomalous—they represent the bias of experience over reality. In fact, most continuous functions are not smooth.

Exercise 17.13. *Describe a continuous function from $[0,1]$ to $[0,1]$ that is nowhere differentiable.*

At first it may be difficult to think of any continuous function that is nowhere differentiable, but, in fact, most continuous functions are not differentiable. If you would like a challenge, try to prove that most continuous functions are nowhere differentiable. To define what we mean by 'most', let's consider the topological space of all continuous functions from $[0,1]$ to $[0,1]$.

Definition. Let $\mathcal{C} = \{f \mid f : [0,1] \to [0,1] \text{ is continuous}\}$. The topology on $\mathcal{C}$ is induced by the metric $d(f,g) = max\{|f(x) - g(x)|\}_{x\in[0,1]}$.

By 'most' we will mean a countable intersection of dense open sets in $\mathcal{C}$. In particular, the countable intersection of dense open sets in $\mathcal{C}$ is dense (why?), because $\mathcal{C}$ is a complete metric space (why?).

Theorem 17.14. *Let B be the set of all nowhere differentiable continuous functions from $[0,1]$ to $[0,1]$. Then B is the intersection of countably many dense open sets in $\mathcal{C}$.*

Often things that at first seem strange and anomalous later turn out to be the norm rather than the rare exception. Nowhere differentiable continuous functions provide us with a great example of that phenomenon.

Let's think of some other apparently anomalous features that an embedded arc might have. Embeddings of $[0,1]$ can have infinite length.

Exercise 17.15. *Describe an embedding of $[0,1]$ into the plane that has infinite length. In fact, you might choose the graph of a differentiable function.*

Embedded arcs can have other strange properties.

Exercise 17.16. *Describe an embedding of $[0,1]$ into the unit square and two points x and y in the unit square not on the embedded arc such that to connect x to y by a polygonal path missing the embedded arc requires a polygonal path of length at least a mile.*

All these examples were explored to suggest that the world of embedded arcs may be more varied and complicated than you might at first have supposed. Nevertheless, it is true that no arc separates the plane.

Proving that fact is surprisingly difficult. One method of proving it is to prove that if an embedded arc did separate the plane, then half of it must do so. This technique involves relating a characteristic of a whole space to related characteristics of spaces whose intersection or union is the space we are interested in. This impulse suggests that we consider using the Mayer-Vietoris Theorem. Recall that the Mayer-Vietoris Theorem relates the homology of pieces of a space to the homology of the whole space.

Lemma 17.17. *Let $h : [0,1] \to \mathbb{R}^2$ be an embedding and let p and q be points in $\mathbb{R}^2 - h([0,1])$. If p and q are connected in $\mathbb{R}^2 - h([0,\frac{1}{2}])$ and p and q are connected in $\mathbb{R}^2 - h([\frac{1}{2},1])$, then p and q are connected in $\mathbb{R}^2 - h([0,1])$.*

You might want to observe that saying that p and q are connected in a space is the same as saying that the 0-cycle $\{p,q\}$ bounds a 1-chain in that space. The catechism for using the Mayer-Vietoris Theorem suggests saying, ”Suppose $\{p,q\}$ bounds a 1-chain here; and $\{p,q\}$ bounds a 1-chain there. Then suppose the 1-cycle created by those two 1-chains bounds a 2-cycle in some union, then $\{p,q\}$ must bound in some intersection.” Think about asking the question of connecting $\{p,q\}$ in $\mathbb{S}^2$ rather than $\mathbb{R}^2$, and figure out how to apply the Mayer-Vietoris Theorem to conclude that the lemma is true.

After you have accomplished that insight, you will be able to prove that no embedded arc separates the plane.

Theorem 17.18. *Let $h : [0,1] \to \mathbb{R}^2$ be an embedding. Then $h([0,1])$ does not separate $\mathbb{R}^2$.*

17.7 A ball does not separate $\mathbb{R}^n$

Effective Thinking Principle. *Extend Results.* ”The time to work on a result is after you’ve solved it.” —R.H. Bing
After getting a result, see if it can be extended.

R.H. Bing was a twentieth century topologist. One strategy he advocated for doing successful research was to make maximal use of insights you obtain. That is, when you have discovered or understood a technique, so whether that same insight can be applied to extend your result.

In the previous section, you proved that an arc does not separate the plane. It is natural to ask what extensions of that theorem might be true. For example, was it important that the embedding of the arc be in $\mathbb{R}_2$? Would the theorem still be true if the embedding were in $\mathbb{R}^n$ for larger n’s? Would the theorem still be true if instead of embedding an arc, we embedded a higher dimensional disk? Those questions are natural extensions of the arc not separating the plane result.

The other extension to explore is an extension of the proof. Would the same proof or the same style of proof work in the higher dimensional cases?

It turns out that extensions of both the result and the proof technique work for higher dimensions. But the results are not obvious, because in higher dimensions, embeddings can become quite strange.

In the plane we saw that for any embedding of an arc or a simple closed curve, there is a homeomorphism of the plane that takes that embedding to a nice embedding, for example a smooth or a polygonal embedding. The analogous statements in higher dimensions are not true. That is, there exists an embedding of an arc in $\mathbb{R}^3$ such that there is no homeomorphism of $\mathbb{R}^3$ to itself that takes that embedded arc to a straight line. Such embeddings are appropriately called *wild* embeddings. However, even though such wild embeddings exist, from the point of view of homology, all embeddings behave like standard embeddings behave.

Let's begin by showing that arcs cannot separate $\mathbb{R}^n$ by following the same strategy as we used to show that an arc cannot separate $\mathbb{R}^2$. As before, it is easier to work in $\mathbb{S}^n$ rather than $\mathbb{R}^n$.

Lemma 17.19. *For any natural number n, let $h : [0,1] \to \mathbb{R}^n$ be an embedding and let p and q be points in $\mathbb{R}^n - h([0,1])$. If p and q are connected in $\mathbb{R}^n - h([0,\frac{1}{2}])$ and p and q are connected in $\mathbb{R}^n - h([\frac{1}{2},1])$, then p and q are connected in $\mathbb{R}^n - h([0,1])$.*

As before, you will now be able to prove that arcs cannot separate points in $\mathbb{R}^n$ or $\mathbb{S}^n$.

Theorem 17.20. *For any natural number n, let $h : [0,1] \to \mathbb{R}^n$ be an embedding. Then $h([0,1])$ does not separate $\mathbb{R}^n$.*

This result could be phrased in terms of a 0-cycle bounding a 1-chain. Let's see whether we can extend this result by showing that a 1-cycle in the complement of an embedded arc bounds a 2-chain in the complement of that embedded arc. The best way to get a new idea is to use an old idea, so let's use the same strategy as before, namely, dividing the embedded arc into two parts.

Lemma 17.21. *For any natural number n, let $h : [0,1] \to \mathbb{R}^n$ be an embedding and let Z be a $\mathbb{Z}_2$ 1-cycle in $\mathbb{R}^n - h([0,1])$. If Z bounds a 2-chain in $\mathbb{R}^n - h([0,\frac{1}{2}])$ and Z bounds a 2-chain in $\mathbb{R}^n - h([\frac{1}{2},1])$, then Z bounds a 2-chain in $\mathbb{R}^n - h([0,1])$.*

You are now on a roll. You can now prove that an embedded arc does not get in the way of a 1-cycle bounding.

Theorem 17.22. *For any natural number n, let $h : [0,1] \to \mathbb{R}^n$ be an embedding and let Z be a $\mathbb{Z}_2$ 1-cycle in $\mathbb{R}^n - h([0,1])$. Then Z bounds a 2-chain in $\mathbb{R}^n - h([0,1])$.*

Now we can work our way up in dimension by considering whether an embedded arc could obstruct higher dimensional cycles from bounding. Following that path, you will be able to prove the following.

Theorem 17.23. *For any natural numbers n and k with $k < n$, let $h : [0,1] \to \mathbb{S}^n$ be an embedding and let Z be a $\mathbb{Z}_2$ k-cycle in $\mathbb{S}^n - h([0,1])$. Then Z bounds a $(k+1)$-chain in $\mathbb{S}^n - h([0,1])$.*

Next we can work our way up in the dimension of the embedded object from a 1-dimensional arc to a 2-dimensional disk, then a 3-dimensional disk and so on. Finally, you will be able to prove that no embedded ball of any dimension can get in the way of letting a cycle bound.

Theorem 17.24. *Let B be a topologically embedded m-ball in $\mathbb{S}^n$ and let Z be a $\mathbb{Z}_2$ k-cycle in $\mathbb{S}^n - B$. Then there is an $(k+1)$-chain C in $\mathbb{S}^n - B$ whose boundary is Z.*

In particular, no embedded ball separates $\mathbb{S}^n$.

17.8 The Jordan-Brouwer Separation Theorem

If you draw a circle in the plane, that circle separates the plane into two pieces and is the boundary of each. Even if you distort the circle, that is, you consider any embedding of the circle in the plane, it is still true that the embedded circle separates the plane into two pieces and is the boundary of each, as you may have seen in Chapter 15. Proving that fact is surprisingly difficult and proving its analogs in higher dimensions is also difficult. However, the concepts of homology theory will come to the rescue and allow us to prove it.

Since the proof involves quite a number of technicalities, we will begin with an outline of the proof to help us keep our bearings. Here is the goal statement that we will prove.

Theorem 17.25 (Jordan-Brouwer Separation Theorem)**.** *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding. Then $h(\mathbb{S}^{n-1})$ separates $\mathbb{S}^n$ into precisely two components and is the boundary of each.*

Here is a big-picture outline of the main steps of the proof. Throughout, we will think of $\mathbb{S}^n$ and $\mathbb{S}^{n-1}$ as simplicial complexes where 'straight lines' are great circle segments.

The main idea of the proof is that the embedding $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ can easily be approximated arbitrarily closely by simplicial maps $f : \mathbb{S}^{n-1} \to \mathbb{S}^n$. An approximating simplicial map f is not an embedding, but using homology, we can think of $f(\mathbb{S}^{n-1})$ as dividing $\mathbb{S}^n$ into two pieces. Specifically, the following theorem is true—and not too difficult to prove.

Theorem 17.26 (Two Chains Theorem)**.** *Let $f : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a simplicial map. Then there exist $\mathbb{Z}_2$ n-chains A^n and C^n such that $\partial(A^n) = \partial(C^n) = f_\#(\mathbb{S}^{n-1})$ and $A^n \cup C^n = \mathbb{S}^n$.*

Outline of the proof of the Jordan-Brouwer Separation Theorem

We approximate the embedding h with a sequence of increasingly finer simplicial approximations $f_i : \mathbb{S}^{n-1} \to \mathbb{S}^n$. For each i, we produce an A_i and C_i using the Two Chains Theorem above. We decide on a criterion for which side is the A_i side and which side is the C_i side. Then we shave off a bit from each A_i and C_i to produce connected subsets A_i' and C_i' of A_i and C_i respectively that are disjoint and just miss $h(\mathbb{S}^{n-1})$. Finally, we conclude that $\cup A_i'$ and $\cup C_i'$ are the two components of $\mathbb{S}^n - h(\mathbb{S}^{n-1}$ that we seek.

The Proof

So let's begin by proving the Two Chains Theorem above. As a suggestion for getting started, assume that $f(\mathbb{S}^{n-1})$ misses the north and south poles of $\mathbb{S}^n$. Then draw great circle lines between the north pole and each point of $f(\mathbb{S}^{n-1})$ to create a simplicial map F from the cone over $\mathbb{S}^{n-1}$ (basically B^n or the northern hemisphere of $\mathbb{S}^n$) into $\mathbb{S}^n$ that extends f. Now remember that the boundary of the image is the image of the boundary. After you have produced the n-chain A, you really have no choice about what C must be to satisfy the conclusions of the Two Chains Theorem.

It is easy to create close simplicial approximations of a homeomorphism $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$— just take a fine triangulation of $\mathbb{S}^{n-1}$, have those vertices under f go to the same places that h takes them, and extend linearly. But we are left with the question of how close an approximation would be useful. Part of the answer lies with our realizing that for any embedding of $\mathbb{S}^{n-1}$, some neighborhood of it retracts to it. That insight is the content of the next theorem, which is easier to state using the following definition.

Definition. A topological space Y is an **absolute neighborhood retract** if and only if for every normal space X and embedding $h : Y \to X$ such that $h(Y)$ is a closed subset of X, there is a neighborhood U of $h(Y)$ that retracts to $h(Y)$, that is, there is a neighborhood U of $h(Y)$ and a continuous function $r : U \to h(Y)$ such that $r(U) = h(Y)$ and for every point $x \in h(Y)$, $r(x) = x$.

Theorem 17.27 (Absolute Neighborhood Retract Theorem)**.** *For every k, $\mathbb{S}^k$ is an absolute neighborhood retract.*

We will be applying this fact about spheres being absolute neighborhood retracts to $\mathbb{S}^{n-1}$ in order to conclude that there exists some neighborhood of $h(\mathbb{S}^{n-1})$ that retracts to $h(\mathbb{S}^{n-1})$. First let's prove that spheres are absolute neighborhood retracts.

We will need to remember the Tietze Extension Theorem for $\mathscr{B}^n$, namely, if X is a normal space, $A \subset X$ is closed, and $g : A \to \mathscr{B}^n$ is a continuous function (where $\mathscr{B}^n$ is the n-ball), then there exists a continuous function $G : X \to \mathbb{D}^n$ that extends g.

We have several of the hypotheses of the Tietze Extension Theorem, namely, $h(\mathbb{S}^{n-1})$ is a closed subset of the normal space $\mathbb{S}^n$. There is a natural continuous function from $h(\mathbb{S}^{n-1})$ to $\mathcal{B}^n$, namely, h^{-1}. So the Tietze Extension Theorem says that $h^{-1} : h(\mathbb{S}^{n-1}) \to \mathcal{B}^n$ can be extended to a continuous function $F : \mathbb{S}^n \to \mathcal{B}^n$. A neighborhood of $\partial\mathcal{B}^n = \mathbb{S}^{n-1} = h^{-1}(h(\mathbb{S}^{n-1}))$ certainly retracts to $\partial\mathcal{B}^n$. How can that retraction be used to create a retraction from a neighborhood U of $h(\mathbb{S}^{n-1})$ to $h(\mathbb{S}^{n-1})$?

No neighborhood that retracts to $h(\mathbb{S}^{n-1})$ can equal the whole sphere. Once again homology saves the day.

Theorem 17.28. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding and let U be a neighborhood of $h : \mathbb{S}^{n-1}$ that retracts to it. Then $U \neq \mathbb{S}^n$.*

A corollary to the existence of a neighborhood that retracts to $h(\mathbb{S}^{n-1})$ is that we can find an even smaller neighborhood of $h(\mathbb{S}^{n-1})$ where the retraction can be accomplished by a straight line homotopy.

Corollary 17.29. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding and let U be a neighborhood of $h : \mathbb{S}^{n-1}$ with retraction $r : U \to h(\mathbb{S}^{n-1})$. Then there exists an open set $V \subset U$ such that for every point $v \in V$, the straight line between v and $r(v)$ is contained in U.*

The retracting neighborhood U of $h(\mathbb{S}^{n-1})$ will be useful in proving the following lemma. Also useful is the fact that any map from $\mathbb{S}^{n-1}$ to $\mathbb{S}^{n-1}$ that is homotopic to the identity map of $\mathbb{S}^{n-1}$ induces a non-trivial homomorphism on the $(n-1)$-homology group.

Lemma 17.30. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding. Then there exists an $\epsilon > 0$ such that if $f : \mathbb{S}^{n-1} \to \mathbb{S}^n$ is a simplicial map such that $d(f(x), h(x)) < \epsilon$ for all $x \in \mathbb{S}^{n-1}$, then $f_\#(\mathbb{S}^{n-1})$ does not bound an n-chain in the ϵ-neighborhood of $h(\mathbb{S}^{n-1})$.*

A corollary to the proof of this lemma is the observation that close simplicial approximations of h co-bound n-chains near $\mathbb{S}^{n-1}$.

Lemma 17.31. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding and let $\epsilon > 0$. Then there exists a $\delta > 0$ such that if $f, g : \mathbb{S}^{n-1} \to \mathbb{S}^n$ are simplicial maps such that $d(f(x), h(x)) < \delta$ and $d(g(x), h(x)) < \delta$ for all $x \in \mathbb{S}^{n-1}$, then $f_\#(\mathbb{S}^{n-1})$ and $g_\#(\mathbb{S}^{n-1})$ bound an n-chain in the ϵ-neighborhood of $h(\mathbb{S}^{n-1})$.*

Another corollary of the fact that simplicial approximations of h do not bound in U is that the A and C in the Two Chains Theorem must have distinct anchor points outside of the retracting neighborhood U for any approximation.

Lemma 17.32. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding, U be a neighborhood that retracts to $h(\mathbb{S}^{n-1})$, let $f : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a simplicial map such that for each point $x \in \mathbb{S}^{n-1}$, the straight line segment between $h(x)$ and $f(x)$ lies in U. Let A^n and C^n be $\mathbb{Z}_2$ n-chains such that $\partial(A^n) = \partial(C^n) = f_\#(\mathbb{S}^{n-1})$ and $A^n \cup C^n = \mathbb{S}^n$. Then there exists a point $a \in (A - U)$ and a point $c \in (C - U)$.*

We are inching our way toward constructing objects that are going to become the two components of the complement of $h(\mathbb{S}^{n-1})$. The next lemma says that the A and C anchor points remain on different sides for nearby approximations of h.

Lemma 17.33. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding, U be a neighborhood that retracts to $h(\mathbb{S}^{n-1})$, let $f, g : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be simplicial maps such that for each point $x \in \mathbb{S}^{n-1}$, the straight line segments between $h(x)$ and $f(x)$ and $h(x)$ and $g(x)$ lie in U. Let A^n_f and C^n_f be $\mathbb{Z}_2$ n-chains (from the Two Chains Theorem) such that $\partial(A^n_f) = \partial(C^n_f) = f_\#(\mathbb{S}^{n-1})$ and $A^n_f \cup C^n_f = \mathbb{S}^n$ and let $a \in (A - U)$ and $c \in (C - U)$. Let A^n_g and C^n_g be $\mathbb{Z}_2$ n-chains such that $\partial(A^n_g) = \partial(C^n_g) = g_\#(\mathbb{S}^{n-1})$ and $A^n_g \cup C^n_g = \mathbb{S}^n$ where $a \in A_g$. Then $c \notin A_g$.*

Our goal is to create connected components of $\mathbb{S}^n - h(\mathbb{S}^{n-1})$, so we need to bring connected pieces into the discussion.

Lemma 17.34. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding, U be a neighborhood such that there is a retract $r : U \to h(\mathbb{S}^{n-1}$, and V be an open set in U such that for each $x \in V$, the straight line segment from x to $r(x)$ is in U. Let $a \notin U$ and let T be a standard triangulation of $\mathbb{S}^n$ with simplexes so small that for any simplex $\sigma \in T$, if $\sigma \cap (\mathbb{S}^n - V) \neq \varnothing$, then $\sigma \cap h(\mathbb{S}^{n-1} = \varnothing$. Let τ_0 be an n-simplex in T such that $a \in \tau_0$. Let A be the union of all n-simplexes τ_k in T such that there are n-simplexes $\{\tau_i\}_{i=0,\cdots,k}$ such that (1) each τ_i contains a point in $\mathbb{S}^n - V$, and (2) for each i, τ_i and τ_{i+1} share an $(n-1)$-face. Then $\partial A \subset V$ and $h^{-1}(r(\partial A))$ is the non-trivial element of $H_{n-1}(\mathbb{S}^{n-1})$. Also, $r(\partial A) = h(\mathbb{S}^{n-1})$.*

The final consequence of the above lemma is what will allow us to conclude that each point of $h(\mathbb{S}^{n-1})$ will be a limit point of the boundary of each of its two components.

Putting all these insights together, we can create the sequence of A_i's and C_i's who unions are the components we seek.

Lemma 17.35. *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding, let $\{U_i\}_{i\in\mathbb{N}}$ be open sets each containing $h(\mathbb{S}^{n-1})$ and each contained in the $\frac{1}{n}$-neighborhood of $h(\mathbb{S}^{n-1})$ with retract $r : U_1 \to h(\mathbb{S}^{n-1})$ and such that for every $i \in \mathbb{N}$ and every point $x \in U_{i+1}$, the straight line homotopy between x and $r(x)$ lies in U_i. Let $\{T_i\}_{i\in\mathbb{N}}$ be a sequence of triangulations of $\mathbb{S}^n$ where each triangulation T_{i+1} is a subdivision of T_i with simplices so small that any simplex of T_i that intersects $h(\mathbb{S}^{n-1})$ lies entirely in U_i. Let $a \in (\mathbb{S}^n - U_1)$. Let A_i be the component containing a of the union of all n-simplices of T_i that miss $h(\mathbb{S}^{n-1})$. Then $\partial(A_{i+1})$*

and $\partial(A_{i+2})$ co-bound an n-chain in U_i, $\cup_{i\in\mathbb{N}}A_i \cap h(\mathbb{S}^{n-1}) =$, each point $x \in h(\mathbb{S}^{n-1})$ is a limit point of $\cup_{i\in\mathbb{N}}A_i$, there exists a point c in $(\mathbb{S}_n - \cup_{i\in\mathbb{N}}A_i - U_1)$, and if we do the same process that we did for a for c creating C_i's, then $\mathbb{S}_n = (\cup_{i\in\mathbb{N}}A_i) \cup h(\mathbb{S}^{n-1}) \cup (\cup_{i\in\mathbb{N}}C_i)$.

Finally, we have proved the Jordan-Brouwer Separation Theorem.

Theorem 17.36 (Jordan-Brouwer Separation Theorem)**.** *Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^n$ be a topological embedding. Then $h(\mathbb{S}^{n-1})$ separates $\mathbb{S}^n$ into precisely two components and is the boundary of each.*

Another commonly used version of the Jordan-Brouwer Separation Theorem is the following corollary. It is one the most fundamental results about the global topology of $\mathbb{R}^n$.

Corollary 17.37 (Jordan-Brouwer Separation Theorem)**.** *Let $h : \mathbb{S}^{n-1} \to \mathbb{R}^n$ be a topological embedding. Then $h(\mathbb{S}^{n-1})$ separates $\mathbb{R}^n$ into precisely two components and is the boundary of each. The two components are distinguished topologically by the fact that one has a compact closure and the other does not.*

One of the consequences of understanding the proof of the Jordan-Brouwer Separation Theorem is that every topologically embedded $(n-1)$-manifold must separate $\mathbb{R}^n$ just as an $(n-1)$-sphere does.

Theorem 17.38. *Every connected, compact topologically embedded $(n-1)$-manifold in $\mathbb{R}^n$ separates $\mathbb{R}^n$ into two components and is the topological boundary of each.*

This insight allows us to conclude that a Klein bottle cannot be embedded in $\mathbb{R}^3$.

Theorem 17.39. *The Klein Bottle cannot be embedded in $\mathbb{R}^3$.*

17.9 $\mathbb{Z}_2$-Homology—A Topological Superhero

The results in this section are among the triumphs of topology. They are definitely a rich reward for developing the concepts of homology that allow us to prove them.

One way to evaluate the perceived significance of mathematical results is to notice which ones have names. In this chapter you proved many named theorems, some named after people and some named after fast food cuisine: the No Retraction Theorem, the Brouwer Fixed Point Theorem, the Borsuk-Ulam Theorem, the Ham Sandwich Theorem, the Invariance of Domain Theorem, and the Jordan-Brouwer Separation Theorem.

These theorems are among the highlights of topology. Many of them followed from applying strategies of extending ideas step by step from simple to more complex. That methodical system may appear mundane, but the pay off was enormous here and gives you a strategy of learning and producing ideas that can be used in mathematics and beyond.

Chapter 18

Simplicial $\mathbb{Z}$-Homology: Getting Oriented

In the previous chapters we began our study of homology by defining $\mathbb{Z}_2$ simplicial homology. The $\mathbb{Z}_2$ world was a convenient place to start, because having just two coefficients 1 and 0 corresponded to the idea that a simplex was either present or absent in a chain. Moreover, the boundary of a chain could be easily defined in a way that corresponds with our intuition that an $(n-1)$-simplex appearing where two n-simplices meet should not be regarded as part of the boundary of their union. Having $\mathbb{Z}_2$ coefficients meant that if the boundary of a sum is defined to be the sum of boundaries, then such a shared $(n-1)$-simplex is counted twice, which in $\mathbb{Z}_2$ means they are not counted at all.

> **Effective Thinking Principle.** *Extend Good Ideas.* The best source of new good ideas is old good ideas—generalize and extend them.

In this chapter, we will develop a homology theory that generalizes $\mathbb{Z}_2$ simplicial theory in two important respects. First, we consider how to define simplicial homology with $\mathbb{Z}$ coefficients in much the same way as we did with $\mathbb{Z}_2$ coefficients, and we flesh out several more properties of homology, including a relative version. Then in the next chapter we show how to generalize simplicial homology to a more general homology theory called *singular homology* which holds for any topological space, not just simplicial complexes.

We hope you will also see this chapter as a opportunity to take your learning to a new level. We will provide less scaffolding in these final chapters than in prior chapters. However, you'll be able to draw on the effective learning principles as well as the topological intuition you have developed. And whenever there is a new idea, we'll try to suggest a path forward. As you have already seen, many of the ideas of homology theory are easy to describe, but pinning down the details requires good notation, careful bookkeeping, and attention to just a few important ideas.

> **Effective Thinking Principle.** *Pin Down Intuition.* Let your intuition be your guide—then pin it down.

18.1 Orientation and $\mathbb{Z}$-Homology

In Chapter 16, we developed simplicial homology with $\mathbb{Z}_2$ coefficients, but it turns out that we can develop homology with coefficients in any group G. In order to keep the notation as simple as possible in Chapter 16, the $\mathbb{Z}_2$-homology groups of a complex K were denoted $\mathsf{H}_n(K)$, but customarily they are denoted $H_n(K;\mathbb{Z}_2)$, and $H_n(K;G)$ is the notation used for homology with coefficients in G. When $G=\mathbb{Z}$, it is customary to drop the group from the notation and write $H_n(K)$, since $\mathbb{Z}$ is the most common coefficient group to use. We will study $\mathbb{Z}$-homology throughout this chapter.

To define simplicial $\mathbb{Z}$-homology, we need to alter our definition of the boundary map so that cancellation occurs for a shared $(n-1)$-simplex between two n-simplices in a chain. Using $\mathbb{Z}$ coefficients will allow us to draw finer distinctions among spaces. For example, $\mathbb{Z}_2$-homology was not able to distinguish the torus $\mathbb{T}^2$ and the Klein bottle $\mathbb{K}^2$. But we will find that $\mathbb{T}^2$ and $\mathbb{K}^2$ have different $\mathbb{Z}$-homology groups in dimension 2, and in fact, we will see that $\mathbb{Z}$-homology groups will successfully distinguish all compact connected surfaces from one another.

The first idea to grapple with is what it means to have chains with coefficients in $\mathbb{Z}$. Now we can talk about 5 times a simplex or (-3) times a simplex. The possibility of having a "negative" simplex means that we can salvage our definition of the boundary map at an $(n-1)$-simplex where two n-simplices meet by getting their boundaries to "cancel" along that common face.

We have already seen something similar in Section 12.8. There we noted that every edge and every triangle had two possible *orientations*. Orientation allows us to define the negative of a simplex with a given orientation to be the same simplex but with the other orientation. We revisit this idea here in order to generalize the idea to higher dimensional complexes.

For an edge $\{vw\}$, the two orientation classes correspond to two orderings of the vertices v and w, and are denoted $[vw]$ and $[wv]$. It is customary to think of the oriented edge $[vw]$ as an edge with an arrow pointing from v to w. We set $[vw]=-[wv]$.

For a triangle $\{uvw\}$ with vertices u, v, and w, the two orientation classes correspond geometrically to clockwise or counterclockwise orderings of the vertices when viewed from a particular vantage point. Algebraically, we can think of these as grouping permutations of 3 vertices into two equivalence classes, and as above, we place brackets around the permutations to denote these classes. Thus $[uvw] = [vwu] = [wuv]$ and $[uwv] = [wvu] = [vuw]$. Readers familiar with algebra may recognize that these are the classes of permutations that differ by an even number of transpositions. In a chain group that we will define soon, we set these classes to be additive inverses of one another; e.g., $[uvw] = -[vuw]$.

A natural way to define the boundary of $[uvw]$ is $[uv] + [vw] + [wu]$, a sum of oriented edges. Thus the orientation of the triangle $\{uvw\}$ will induce an orientation in each of the edges of the boundary. This method of inducing orientation is good, because if we were to orient each triangle in Figure 18.1 counterclockwise, then the boundary of the neighboring triangles that meet along edge $\{vw\}$ would have that edge oriented in opposite directions. An alternative way to write the boundary of $[uvw]$ is:

$$\partial[uvw] = [vw] - [uw] + [uv],$$

an alternating sum of oriented edges obtained by removing one vertex from the oriented simplex $[uvw]$.

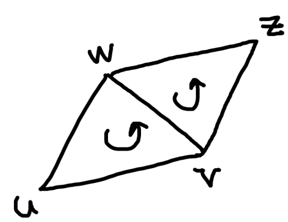


Figure 18.1: Neighboring oriented simplices. Note how the simplices induce opposite orientations on the edge vw.

We can now generalize our concept to apply to n-simplices.

Definition. Let $\{v_0 \cdots v_n\}$ be an n-simplex. Orderings of the vertices $v_0, ..., v_n$ are placed into two **orientation classes** such that the orderings in each class differ by an even number of transpositions. An n-simplex $\{v_0 \cdots v_n\}$ with a chosen ordering is called an **oriented simplex** and denoted $[v_0 \cdots v_n]$.

We shall think of the same simplex with opposite orientations as additive inverses of one another in the chain group we now define. Just as in the case of $\mathbb{Z}_2$-homology, a chain will be a sum of simplices, but this time it will be a sum of oriented simplices.

Definition. Fix a simplicial complex K. The n**-chain group of** K (with coefficients in $\mathbb{Z}$) is the free abelian group on the collection of oriented n-simplices of K, modulo the relation that any oriented simplex and its oppositely oriented version are inverses of one another. This group is denoted $C_n(K)$ or $C_n(K;\mathbb{Z})$ when one wants to mention the coefficient group $\mathbb{Z}$ explicitly. Thus $C_n(K)$ is a free abelian group generated by a set containing exactly one orientation of each n-simplex. When there are no n-simplices in K, $C_n(K) = \{0\}$, the trivial group. An **n-chain** is then an element of $C_n(K)$, a $\mathbb{Z}$-linear combination of a finite number of oriented simplices.

Example. Recall the example from Figure **??**, a simplicial complex

$$K = \{\sigma, e_1, e_2, e_3, e_4, e_5, v_1, v_2, v_3, v_4\},$$

which is a filled in triangle and a hollow triangle, redrawn in Figure 18.2.

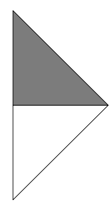


Figure 18.2: A simplicial complex.

Then $C_2(K)$ is generated by the oriented 2-simplex σ. (It is also generated by $-\sigma$.) Hence $C_2(K)$ is isomorphic to $\mathbb{Z}$.

We see also that $C_1(K)$ is a free abelian group on 5 generators: the edges e_1, e_2, e_3, e_4, e_5 (though other generators would also work, for instance: $e_1, -e_2, -e_3, e_4, -e_5$). A sample element c in $C_1(K)$ is $c = 3e_1 - 7e_3 + e_4$.

Similarly, $C_0(K)$ is a free abelian group on 4 generators, namely, the vertices v_1, v_2, v_3, v_4.

Besides viewing a chain as a formal sum of simplices, it may sometimes be convenient to think of a chain as a tuple (one for each generator) or as a function from generators to $\mathbb{Z}$. For instance the element $c = 3e_1 - 7e_3 + e_4$ in $C_1(K)$ can be thought of as a 5-tuple $(3, 0, -7, 1, 0)$ or as an integer-valued function on generators, e.g., $c(e_3) = -7$.

We can define the boundary map:

Definition. For $n \geq 1$, the **boundary of an oriented n-simplex** $\sigma = [v_0 \ldots v_n]$ is defined by

$$\partial(\sigma) = \sum_{i=0}^{n} (-1)^i [v_0 \ldots \widehat{v_i} \ldots v_n].$$

Recall that $[v_0 \ldots \widehat{v_k} \ldots v_n]$ represents the simplex with v_k removed. The boundary of σ can be viewed as the sum of the $(n-1)$-dimensional faces of σ, each of which has the **induced orientation** $(-1)^i[v_0 \ldots \widehat{v_i} \ldots v_n]$. The boundary of a 0-simplex is to defined to be zero.

Exercise 18.1. *Check that this boundary map is well-defined: it does not depend on the oriented representative chosen for the definition.*

Exercise 18.2. *Find the boundary of the oriented 2-simplex $\tau = [v_0 v_1 v_2]$ and the boundary of the oriented 3-simplex $\sigma = [w_0 w_1 w_2 w_3]$. Repeat the procedure for $-\tau$ and $-\sigma$. What is the relationship between the boundary of τ and the boundary of $-\tau$? What is the relationship between the boundary of σ and the boundary of $-\sigma$?*

Theorem 18.3. *For any n-simplex σ*

$$\partial(-\sigma) = -\partial(\sigma).$$

So we may define the boundary of an n-chain by extending the definition on simplices.

Definition. The **boundary of an n-chain** $\sum_{i=1}^{k} c_i \sigma_i$ is an $(n-1)$-chain:

$$\partial\left(\sum_{i=1}^{k} c_i \sigma_i\right) = \sum_{i=1}^{k} c_i \partial(\sigma_i).$$

Thus the boundary operator is a homomorphism

$$\partial_n : C_n(K) \to C_{n-1}(K)$$

for each $n \geq 0$ (note that $C_{-1}(K) = \{0\}$).

Theorem 18.4. *For all $n \geq 0$,*

$$\partial_n \circ \partial_{n+1} = 0.$$

We suppress writing the subscript n on ∂_n when there is no possibility of confusion. The sequence of chain groups and boundary maps

$$\cdots \xrightarrow{\partial} C_{n+1}(K) \xrightarrow{\partial} C_n(K) \xrightarrow{\partial} C_{n-1}(K) \xrightarrow{\partial} \cdots$$

is called a **chain complex**.

Definition. An n-**cycle** of a simplicial complex K is an n-chain whose boundary is zero. The collection of n-cycles, denoted $Z_n(K)$, is the kernel of the homomorphism ∂_n. An n**-boundary** of K is an n-chain that is the boundary of an $(n+1)$-chain. The collection of n-boundaries, denoted $B_n(K)$, is the image of the homomorphism ∂_{n+1}.

Theorem 18.5. *For any simplicial complex K, both $Z_n(K)$ and $B_n(K)$ are subgroups of $C_n(K)$, and $B_n(K) \subset Z_n(K)$.*

Definition. Two n-cycles α and β in the simplicial complex K are **equivalent** or **homologous** if and only if $\alpha - \beta = \partial\gamma$, where γ is some $(n+1)$-chain. In other words, α and β are homologous if they differ by an element of the subgroup $B_n(K)$. Being homologous is denoted:

$$\alpha \sim \beta.$$

The equivalence class of α is denoted by $[\alpha]$.

In the expression $\alpha - \beta = \partial\gamma$, we can see that two n-cycles are equivalent if together they bound an $(n+1)$-chain γ when the orientation of β is reversed.

If we turn our attention to dimension one, we get a sense of a relationship between homology and the fundamental group. Suppose two oriented 1-cycles α and β both involve loops that begin and end at a common point. If together they bound a 2-chain γ, then they are homologous. This scenario is reminiscent of the definition of homotopic loops when we were defining the fundamental group. Indeed, there is a specific relationship between the first homology group (with $\mathbb{Z}$ coefficients) and the fundamental group of a simplicial complex, as we shall see later in this chapter.

Exercise 18.6. *Create a triangulation of a Möbius band such that the central circle forms a* 1*-cycle* γ*. Show that the Möbius band's boundary* 1*-cycle* α *is equivalent to either* 2γ *or* -2γ *(depending on the orientation you give the two cycles).*

Definition. The n^{th}**-homology group** (with coefficients in $\mathbb{Z}$) of a finite simplicial complex K is the additive group of n-cycles under the equivalence defined above. The group is denoted $H_n(K)$ or $H_n(K;\mathbb{Z})$ when one wants to mention the coefficient group $\mathbb{Z}$ explicitly. Thus:

$$H_n(K) = Z_n(K)/B_n(K).$$

Theorem 18.7. *For a finite simplicial complex* K*,* $H_n(K)$ *is a finitely generated abelian group.*

One of the conveniences of homology is that for a finite complex K, the group $H_n(K)$ is always a finitely generated abelian group, a type of structure which is well-understood. In particular, it is usually easy to see whether two such groups are isomorphic or not. After we show that homeomorphic spaces have isomorphic homology groups, then we will know that two spaces with different homology groups are not homeomorphic. So computing homology groups is an effective tool for distinguishing spaces.

Theorem 18.8. *If* K *is a connected simplicial complex, then* $H_0(K)$ *is isomorphic to* $\mathbb{Z}$*. If* K *has* r *connected components, then* $H_0(K)$ *is a free abelian group of rank* r*.*

Theorem 18.9. *If K is a one-point space, $\mathsf{H}_n(K) \cong 0$ for $n \geq 0$ and $\mathsf{H}_0(K) \cong \mathbb{Z}$.*

Let K be a finite simplicial complex. A point $x \notin K$ can "see" K if any ray from x intersects $|K|$ at most once. The cone of x over a simplex $\{v_0 \cdots v_k\}$ is the simplex $\{xv_0 \cdots v_k\}$ and $x * K$ denotes the simplicial complex that is the cone over K. Now we define the cone over an oriented simplex.

Definition. Let K be a finite simplicial complex and let x be a point that sees K. If $\sigma = [v_0 \cdots v_k]$ is an oriented simplex of K, define the **cone** of x over σ to be the oriented simplex:

$$\mathsf{Cone}_x(\sigma) = [xv_0 \cdots v_k].$$

Then there is a **simplicial cone operator** $\mathsf{Cone}_x : \mathsf{C}_n(K) \to \mathsf{C}_{n+1}(w * K)$ that extends the definition of $\mathsf{Cone}_x(\sigma)$ linearly to chains.

Theorem 18.10. *Let x see a complex K, and let $c \in C_n(K)$ be a chain. Then*

$$\partial\, \mathsf{Cone}_x(c) + \mathsf{Cone}_x(\partial c) = c.$$

Recall that a space with the homology groups of a point are called *acyclic*.

Corollary 18.11. *For any complex K and x seeing K, the complex $x * K$ is acyclic.*

Theorem 18.12. *The complex K consisting of an n-simplex together with all its faces is acyclic.*

> **Effective Thinking Principle.** *Generalize Methodically.* When generalizing a concept or theory, follow through and methodically consider generalizations of all features of the source.

For any simplicial map, there's an associated chain map, and an induced homomorphism in homology.

Definition. Let $f : K \to L$ be a simplicial map between complexes K and L. The induced **chain map** $f_{\#n} : C_n(K) \to C_n(L)$ is defined on oriented simplices as follows:

$$f_{\#n}([v_0, ..., v_n]) = \begin{cases} [f(v_0), ..., f(v_n)] & \text{if all the } f(v_i) \text{ are distinct} \\ 0 & \text{otherwise.} \end{cases}$$

As we did with the boundary operator, we will typically drop the subscript n from our notation and simply write $f_{\#}$.

Theorem 18.13. *Let $f : K \to L$ be a simplicial map, and let $f_\#$ be the induced map $f_\# : C_n(K) \to C_n(L)$. Then for any chain $c \in C_n(K)$, $\partial(f_\#(c)) = f_\#(\partial(c))$. In other words, the diagram:*

$$\begin{array}{ccc} C_n(K) & \xrightarrow{f_\#} & C_n(L) \\ \partial\Big\downarrow & & \Big\downarrow\partial \\ C_{n-1}(K) & \xrightarrow{f_\#} & C_{n-1}(L) \end{array}$$

commutes.

Because of this fact, the induced homomorphism is well-defined.

Definition. Let $f : K \to L$ be a simplicial map. The **induced homomorphism** $f_* : H_n(K) \to H_n(L)$ is defined by $f_*([z]) = [f_\#(z)]$.

You have seen how to show this map exists and is well-defined for $\mathbb{Z}_2$-homology. Let's do it again here, but take the opportunity to illustrate the argument as a **diagram chase**, a frequent method of reasoning in homology theory. Consider the following commutative diagram.

$$\begin{array}{ccc} C_{n+1}(K) & \xrightarrow{f_\#} & C_{n+1}(L) \\ \partial\Big\downarrow & & \Big\downarrow\partial \\ C_n(K) & \xrightarrow{f_\#} & C_n(L) \\ \partial\Big\downarrow & & \Big\downarrow\partial \\ C_{n-1}(K) & \xrightarrow{f_\#} & C_{n-1}(L) \end{array}$$

The idea of a diagram chase is to allow the diagram to suggest where the argument should go next. Let's first show that $f_\#$ takes cycles to cycles. Consider a cycle z in $C_n(K)$. What do we know about z in the diagram? Well, because it is a cycle, it maps to 0, going downward in the diagram to $C_{n-1}(K)$. And that maps rightward to 0 in $C_{n-1}(L)$. Whenever we have gone around a square one way, we think: "I could have gone around the square another way." So that means z maps rightward to some c in $C_n(L)$ which must then map downward to 0, by commutativity of the square. So, we conclude that c is a cycle and that means the definition $f_*([z]) = [f_\#(z)]$ makes sense.

Do you see how the diagram often suggests what to do next?

Now, see if you can use a diagram chase to show that f_* is well-defined. Start with two cycles z and z' in $C_n(K)$ that differ by a boundary. That means, looking upward, there is a chain $w \in C_{n+1}(K)$ such that $z - z' = \partial w$. What should we do with w? You take it from here, to finish the proof of the next theorem.

Theorem 18.14. *Let $f : K \to L$ be a simplicial map. Then the induced homomorphism $f_* : H_n(K) \to H_n(L)$ is a well-defined homomorphism.*

18.2 Relative Simplicial Homology

Homology groups are easy to define but they will not be easy to compute unless we develop tools. One strategy that we have seen for $\mathbb{Z}_2$-homology was to break a space into pieces and use the Mayer-Vietoris sequence. Soon, we shall develop a similar sequence for $\mathbb{Z}$-homology.

Another strategy is to find a relationship between the homology groups of a complex and the homology groups of a subcomplex. This approach leads to a more general notion of homology called *relative homology*. This strategy could also have been done for $\mathbb{Z}_2$-homology, though we did not do so earlier.

Definition. Let K' be a subcomplex of a simplicial complex K. Then the chain group $C_n(K')$ can be viewed a subgroup of the chain group $C_n(K)$ consisting of all chains that are zero on any simplex outside $\mathbb{K}'$. Then we can define the **group of relative chains of K modulo K'** as the quotient group:

$$C_n(K, K') = C_n(K)/C_n(K')$$

.

Exercise 18.15. *Check that $C_n(K, K')$ is a free abelian group.*

Note that the boundary map $\partial_n : C_n(K) \to C_{n-1}(K)$ restricts to $\partial_n : C_n(K') \to C_{n-1}(K')$ so that taking boundaries of chains in the subcomplex stays in the subcomplex. This observation produces a boundary map on relative chains.

Theorem 18.16. *There is a boundary map*

$$\partial_n : C_n(K, K') \to C_{n-1}(K, K')$$

such that $\partial_n \circ \partial_{n+1} = 0$ for all $n \geq 0$.

By analogy with the usual homology, we define the **subgroup of relative n-cycles** $Z_n(K, K')$ which are the elements of $\operatorname{Ker} \partial_n$. A relative n-cycle can be represented by a n-chain in K whose boundary lives in the subcomplex K'. And the **subgroup of relative n-boundaries** $B_n(K, K')$ which are the elements in $\operatorname{Im} \partial_{n+1}$. A relative n-boundary can be represented by an n-chain in K that, together with a n-chain in K', forms the boundary of an $(n+1)$-chain in K. By Theorem 19.14, $B_n(K, K')$ is a subgroup of $Z_n(K, K')$. Then the **relative homology group** $H_n(K, K')$ is defined as a quotient—the "relative cycles mod relative boundaries":

$$H_n(K, K') = Z_n(K, K')/B_n(K, K').$$

Exercise 18.17. *Check that if $K' = \varnothing$, the empty set, then $H_n(K, K') = H_n(K)$ for all n, the usual homology groups.*

For a vertex $v \in K$, the relative homology groups $H_n(K, \{v\})$ are sometimes called the **reduced homology groups** of K, written $\tilde{H}_n(K)$. They do not depend on the choice of v.

Exercise 18.18. *Show that $\tilde{H}_n(K) \cong H_n(K)$ for $n > 0$ and $H_0(K) \cong \tilde{H}_0(K) \oplus \mathbb{Z}$.*

Exercise 18.19. *Let K be the complex consisting of a triangle and all its faces. Determine $H_n(K, K')$ for all $n \geq 0$.*

Exercise 18.20. *Let K be a triangulation of an annulus, and let K' be the subcomplex consisting of the inner and outer edges of the annulus. Find a relative 1-cycle in $C_1(K, K')$ that is not a relative 1-boundary.*

Exercise 18.21. *Let K be a triangulation of a Möbius band, and let K' be its boundary. Determine $H_n(K, K')$ for $n \geq 0$.*

The next theorem shows that you can 'excise' a part of the subcomplex without affecting the relative homology.

Theorem 18.22 (Excision)**.** *Suppose K' is a subcomplex of K. Remove an open set U from K' such that what remains is a subcomplex L' of K', and remove U from K so that what remains is a subcomplex L of K. Then*

$$H_n(L, L') \cong H_n(K, K').$$

Now that we've defined the relative homology $H_n(K, K')$, we can consider simplicial maps on pairs that induce homomorphisms on relative homology. Suppose K' is a subcomplex of K and L' is a subcomplex of L. We write

$$f : (K, K') \to (L, L')$$

to denote a simplicial map $f : K \to L$ for which $f(K') \subset L'$.

Theorem 18.23. *Given a simplicial map $f : (K, K') \to (L, L')$ there is an associated chain map $f_\# : C_n(K, K') \to C_n(L, L')$ and induced homomorphism $f_* : H_n(K, K') \to H_n(L, L')$.*

If we can discern the relationship of $H_n(K, K')$ to $H_n(K)$ and $H_n(K')$, we may be able to use this relationship to compute $H_n(K)$ using the other two. For instance, are there natural homomorphisms between them? Before thinking about cycles, let's first think about any homomorphisms that exist between chain groups. Note that the relationship $C_n(K, K') = C_n(K)/C_n(K')$ can be expressed as a statement about homomorphisms. Before reading further, think about these questions.

Exercise 18.24. *There are natural maps between chain groups:*

$$C_n(K') \xrightarrow{\;i\;} C_n(K) \xrightarrow{\;\pi\;} C_n(K, K')$$

What are the maps i and π, and what do you notice about them and their relationship with each other?

The map i is an *inclusion* map, and π is a *projection* map onto cosets. If you did the exercise, you saw that i and π are related in a special way. First you might have noticed that the composition $\pi \circ i = 0$, the zero map. Another way to express this equality is that the image of i is contained in the kernel of π. A second thing you might have noticed is that there are no other elements of $C_n(K)$ contained in the kernel of π. In other words, the image of i exactly equals the kernel of π. This special situation has a name as you may recall from the section on the Mayer-Vietoris Theorem in the $\mathbb{Z}_2$-homology chapter. Namely, the sequence $C_n(K') \to C_n(K) \to C_n(K, K')$ is *exact* at $C_n(K)$. We recall the definition here.

Definition. Given a sequence (finite or infinite) of groups and homomorphisms:

$$\cdots \longrightarrow G_{i-1} \xrightarrow{\phi_{i-1}} G_i \xrightarrow{\phi_i} G_{i+1} \longrightarrow \cdots$$

the sequence is **exact at** G_i if and only if $\operatorname{Im}\phi_{i-1} = \operatorname{Ker}\phi_i$. The sequence is called an **exact sequence** if and only if it is exact at each group (except at the first and last groups if they exist).

Exact sequences show up everywhere in the subject of algebraic topology. They turn out to be a surprisingly effective way to summarize relationships between the groups that appear in the subject. For instance, in the exercise above, you might have also noticed that the map i is injective and the map π is surjective. Both these facts are included in the statement that the sequence

$$0 \longrightarrow C_n(K') \longrightarrow C_n(K) \longrightarrow C_n(K, K') \longrightarrow 0$$

is exact. A sequence of five groups such as this sequence—with first and last groups trivial and exactness at each of the three middle groups—is called a **short exact sequence**.

In fact, because the maps i and π exist in each dimension, they are chain maps, and there is therefore a short exact sequence of chain complexes:

$$\begin{tikzcd}
 & {} \arrow[d, "\partial"'] & {} \arrow[d, "\partial"'] & {} \arrow[d, "\partial"'] & \\
0 \arrow[r] & C_{n+1}(K') \arrow[r, "i"] \arrow[d, "\partial"'] & C_{n+1}(K) \arrow[r, "\pi"] \arrow[d, "\partial"'] & C_{n+1}(K, K') \arrow[r] \arrow[d, "\partial"'] & 0 \\
0 \arrow[r] & C_n(K') \arrow[r, "i"] \arrow[d, "\partial"'] & C_n(K) \arrow[r, "\pi"] \arrow[d, "\partial"'] & C_n(K, K') \arrow[r] \arrow[d, "\partial"'] & 0 \\
0 \arrow[r] & C_{n-1}(K') \arrow[r, "i"] \arrow[d, "\partial"'] & C_{n-1}(K) \arrow[r, "\pi"] \arrow[d, "\partial"'] & C_{n-1}(K, K') \arrow[r] \arrow[d, "\partial"'] & 0 \\
 & {} & {} & {} &
\end{tikzcd}$$

with short exact sequences at every level.

If you think about the meaning of relative homology, an element in $H_n(K, K')$ is the class of a relative cycle in $C_n(K, K')$; that relative cycle is represented by a chain in $C_n(K)$ whose boundary is completely inside K'. What may not be apparent at first is that the boundary map on $C_n(K)$ induces a well-defined map from $H_n(K, K')$ to $H_{n-1}(K')$.

Theorem 18.25. *The boundary map $\partial : C_n(K) \to C_{n-1}(K')$ induces a well-defined map*

$$\partial_* : H_n(K, K') \to H_{n-1}(K').$$

A proof of this fact uses the diagram and a diagram chase . As we saw earlier, a diagram chase starts at some node in the diagram, and at each step, uses the information in the diagram to infer the existence of an object at another node in the diagram. Usually it is clear which node to focus on ("chase") next.

We illustrate how to chase a diagram in a proof of the theorem above. Start with a relative cycle z_n in $C_n(K, K')$. Since z_n is a relative cycle, we know (looking downward in the diagram) that $\partial z_n = 0$ in $C_{n-1}(K, K')$. Since π is surjective, z_n is the image of (looking leftward from z_n) some chain c_n in $C_n(K)$. But then (looking downward from there) ∂c_n is a chain in $C_{n-1}(K)$, and the commutativity of the diagram implies that $\pi(\partial c_n) = \partial \pi(c_n) = \partial z_n = 0$. Then exactness of rows at $C_{n-1}(K)$ suggests that ∂c_n is the image of some c_{n-1} in $C_{n-1}(K')$. You'll need to check that c_{n-1} is a cycle (more diagram chasing) and once you do, you can define:

$$\partial_*[z_n] = [c_{n-1}]$$

where brackets denote the homology class. You will also need to show that ∂_* is well-defined: it doesn't depend on the choice of z_n in $C_n(K, K')$ to represent the homology class. This verification involves another diagram chase. You should try to master such arguments.

The map ∂_* and the induced maps i_* and π_* have a special relationship; they form a long exact sequence in homology.

Theorem 18.26 (Long Exact Sequence of a Pair)**.** *If K' is a subcomplex of a simplicial complex K, then there is a long exact sequence:*

$$\cdots \xrightarrow{\partial_*} H_n(K') \xrightarrow{i_*} H_n(K) \xrightarrow{\pi_*} H_n(K, K') \xrightarrow{\partial_*} H_{n-1}(K') \xrightarrow{i_*} \cdots$$

The ideas underlying these proofs can be generalized, since the core of the arguments are algebraic and don't require the underlying meaning of the groups involved. We do this algebra in the next section.

18.3 Some Homological Algebra

In this section, we prove a purely algebraic result known as the Zig-Zag Lemma, as well as some other algebraic results. You've done a version of the Zig-Zag Lemma in showing the existence of a long exact sequence from a short exact sequence, which involved zig-zagging your way through a diagram. The value in this algebraic abstraction is that we will be able to use it in many other applications; for instance, the Mayer-Vietoris sequence can be viewed as a consequence of the Zig-Zag Lemma. The study of such algebraic arguments motivated by homology is called **homological algebra**.

Before proving the Zig-Zag Lemma, we restate in a purely algebraic fashion some concepts we have already encountered.

Definition. A **chain complex** $\mathcal{C}$ is a family $\{C_n, \partial_n\}$ of abelian groups C_n and homomorphisms $\partial_n : C_n \to C_{n-1}$ such that $\partial_n \circ \partial_{n+1} = 0$ for all n. The n**-th homology group** $H_n(\mathcal{C})$ is defined by

$$H_n(\mathcal{C}) = \operatorname{Ker} \partial_n / \operatorname{Im} \partial_{n+1}.$$

Definition. Given two chain complexes $\mathcal{C} = \{C_n, \partial_n\}$ and $\mathcal{C}' = \{C'_n, \partial'_n\}$, a **chain map** $\phi : \mathcal{C} \to \mathcal{C}'$ is a family of homomorphisms $\phi_n : C_n \to C'_n$ such that the ϕ_n commute with the boundary maps:

$$\partial'_n \circ \phi_n = \phi_{n-1} \circ \partial_n.$$

Theorem 18.27 (Zig-Zag Lemma)**.** *Suppose $\mathcal{C} = \{C_n, \partial_n^C\}, \mathcal{D} = \{D_n, \partial_n^D\}, \mathcal{E} = \{E_n, \partial_n^E\}$ are chain complexes, and $\phi : \mathcal{C} \to \mathcal{D}$ and $\psi : \mathcal{D} \to \mathcal{E}$ are chain maps such that*

$$0 \longrightarrow \mathcal{C} \xrightarrow{\ \phi\ } \mathcal{D} \xrightarrow{\ \psi\ } \mathcal{E} \longrightarrow 0$$

is a short exact sequence of chain complexes. Then there is a long exact sequence:

$$\cdots \xrightarrow{\partial_*} H_n(\mathcal{C}) \xrightarrow{\phi_*} H_n(\mathcal{D}) \xrightarrow{\psi_*} H_n(\mathcal{E}) \xrightarrow{\partial_*} H_{n-1}(\mathcal{C}) \xrightarrow{i_*} \cdots$$

where ∂_* is induced by ∂^D.

The short exact sequence of chain complexes in the Zig-Zag Lemma can be visualized as this commutative diagram, where every row is exact, every column is a chain complex, and every square commutes:

$$\begin{tikzcd}
 & {} \arrow[d] & {} \arrow[d] & {} \arrow[d] & \\
0 \arrow[r] & C_{n+1} \arrow[r, "\phi"] \arrow[d, "\partial"'] & D_{n+1} \arrow[r, "\psi"] \arrow[d, "\partial"'] & E_{n+1} \arrow[r] \arrow[d, "\partial"'] & 0 \\
0 \arrow[r] & C_n \arrow[r, "\phi"] \arrow[d, "\partial"'] & D_n \arrow[r, "\psi"] \arrow[d, "\partial"'] & E_n \arrow[r] \arrow[d, "\partial"'] & 0 \\
0 \arrow[r] & C_{n-1} \arrow[r, "\phi"] \arrow[d] & D_{n-1} \arrow[r, "\psi"] \arrow[d] & E_{n-1} \arrow[r] \arrow[d] & 0 \\
 & {} & {} & {} &
\end{tikzcd}$$

To prove the Zig-Zag Lemma, you should chase your way through the diagram above. Start by constructing ∂_*, then check it is well-defined, and after that, check the exactness of the long exact sequence at each group in the sequence. It is hefty work but can be quite satisfying. Everyone should do it at least once in their life.

As the following theorem shows, the long exact sequence that emerges from the Zig-Zag Lemma has a property that it plays well with functions between short exact sequences of chain complexes, the so-called **naturality property** of long exact sequences.

Theorem 18.28. *Given the commutative diagram of chain maps α, β, γ between the chain complexes of two short exact sequences:*

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{C} \arrow[r, "i"] \arrow[d, "\alpha"'] & \mathcal{D} \arrow[r, "\pi"] \arrow[d, "\beta"'] & \mathcal{E} \arrow[r] \arrow[d, "\gamma"'] & 0 \\
0 \arrow[r] & \mathcal{C}' \arrow[r, "i"] & \mathcal{D}' \arrow[r, "\pi"] & \mathcal{E}' \arrow[r] & 0
\end{tikzcd}$$

there are corresponding induced homomorphisms between the associated long exact sequences, such that the following diagram is commutative:

$$\begin{tikzcd}
\cdots \arrow[r, "\partial_*"] & H_n(\mathcal{C}) \arrow[r, "\phi_*"] \arrow[d, "\alpha_*"'] & H_n(\mathcal{D}) \arrow[r, "\psi_*"] \arrow[d, "\beta_*"'] & H_n(\mathcal{E}) \arrow[r, "\partial_*"] \arrow[d, "\gamma_*"'] & H_{n-1}(\mathcal{C}) \arrow[r, "i_*"] \arrow[d, "\alpha_*"'] & \cdots \\
\cdots \arrow[r, "\partial_*"] & H_n(\mathcal{C}') \arrow[r, "\phi_*"] & H_n(\mathcal{D}') \arrow[r, "\psi_*"] & H_n(\mathcal{E}') \arrow[r, "\partial_*"] & H_{n-1}(\mathcal{C}') \arrow[r, "i_*"] & \cdots
\end{tikzcd}$$

The following lemma is useful when analyzing diagrams like the one above.

Lemma 18.29 (The Five Lemma)**.** *Consider the following commutative diagram of groups and homomorphisms, where the rows are exact.*

$$\begin{tikzcd}
A \arrow[r, "q"] \arrow[d, "\alpha"'] & B \arrow[r, "r"] \arrow[d, "\beta"'] & C \arrow[r, "s"] \arrow[d, "\gamma"'] & D \arrow[r, "t"] \arrow[d, "\delta"'] & E \arrow[d, "\epsilon"'] \\
A' \arrow[r, "q'"] & B' \arrow[r, "r'"] & C' \arrow[r, "s'"] & D' \arrow[r, "t'"] & E'
\end{tikzcd}$$

If the rows are exact and $\alpha, \beta, \delta, \epsilon$ are isomorphisms, then γ is also an isomorphism.

The proof is a diagram chase. If you pay attention to what you are doing, you can answer the next question.

Exercise 18.30. *In the proof of the Five Lemma, not all of $\alpha, \beta, \delta, \epsilon$ are required to be isomorphisms for the conclusion to still hold. Which isomorphisms can be relaxed?*

Here is another result from homological algebra with a fun name, made famous by its appearance in a movie.

Exercise 18.31 (The Snake Lemma)**.** *Consider the following commutative diagram where the rows are short exact sequences.*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\ & & {\scriptstyle\alpha}\big\downarrow & & {\scriptstyle\beta}\big\downarrow & & {\scriptstyle\gamma}\big\downarrow & & \\ 0 & \longrightarrow & A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & 0 \end{array}$$

Show there is an exact sequence

$$0 \to \operatorname{Ker}\alpha \to \operatorname{Ker}\beta \to \operatorname{Ker}\gamma \to \operatorname{Coker}\alpha \to \operatorname{Coker}\beta \to \operatorname{Coker}\gamma \to 0$$

where Coker *stands for the cokernel of a homomorphism: the quotient of its codomain by its image.*

18.4 Useful Exact Sequences

The Zig-Zag Lemma has several important consequences. We can establish a long exact sequence involving relative homology:

Corollary 18.32 (Long Exact Sequence of a Pair)**.** *If K' is a subcomplex of a simplicial complex K, then there is a long exact sequence:*

$$\cdots \xrightarrow{\partial_*} H_n(K') \xrightarrow{i_*} H_n(K) \xrightarrow{\pi_*} H_n(K,K') \xrightarrow{\partial_*} H_{n-1}(K') \xrightarrow{i_*} \cdots$$

where the maps are induced by the inclusion maps $i : K' \to K$ and $\pi : (K, \varnothing) \to (K, K')$ and the boundary map $\partial : C_n(X) \to C_{n-1}(X)$.

Naturality implies the following theorem:

Theorem 18.33. *Given a simplicial map $f : (K, K') \to (L, L')$, there is chain map between the long exact sequences:*

$$\begin{array}{ccccccccccc} \cdots & \xrightarrow{\partial_*} & H_n(K') & \xrightarrow{i_*} & H_n(K) & \xrightarrow{\pi_*} & H_n(K,K') & \xrightarrow{\partial_*} & H_{n-1}(K') & \xrightarrow{i_*} & \cdots \\ & & {\scriptstyle f_*}\big\downarrow & & {\scriptstyle f_*}\big\downarrow & & {\scriptstyle f_*}\big\downarrow & & {\scriptstyle f_*}\big\downarrow & & \\ \cdots & \xrightarrow{\partial_*} & H_n(L') & \xrightarrow{i_*} & H_n(L) & \xrightarrow{\pi_*} & H_n(L,L') & \xrightarrow{\partial_*} & H_{n-1}(L') & \xrightarrow{i_*} & \cdots \end{array}$$

Using the Zig-Zag Lemma, we can also establish the Mayer-Vietoris Theorem, in this case, the $\mathbb{Z}$-homology version. You will recall in an earlier chapter, we proved a $\mathbb{Z}_2$ version of this theorem in a different way. Now you see the theorem can be regarded a simple consequence of the Zig-Zag Lemma. You will need to construct homomorphisms ϕ and ψ for the sequence:

$$0 \longrightarrow C_n(A \cap B) \xrightarrow{\phi} C_n(A) \oplus C_n(B) \xrightarrow{\psi} C_n(K) \longrightarrow 0$$

and to make the sequence exact, remember that plus and minus signs are your friends.

Theorem 18.34 (Mayer-Vietoris)**.** *Let K be a finite simplicial complex and A and B be subcomplexes such that $K = A \cup B$. Then there is a long exact sequence*

$$\cdots \xrightarrow{\delta} H_n(A \cap B) \xrightarrow{\phi_*} H_n(A) \oplus H_n(B) \xrightarrow{\psi_*} H_n(K) \xrightarrow{\partial_*} H_{n-1}(A \cap B) \xrightarrow{\phi} \cdots$$

Exercise 18.35. *Compute the $\mathbb{Z}$-homology of the Klein bottle, and compare it to the the $\mathbb{Z}_2$-homology of the Klein bottle.*

Exercise 18.36. *Compute the $\mathbb{Z}$-homology of every compact, triangulated 2-manifold.*

18.5 Homotopy Invariance and Cellular Homology—Same as $\mathbb{Z}_2$

As is the case of simplicial $\mathbb{Z}_2$-homology, simplicial $\mathbb{Z}$-homology is a topological invariant and homotopic maps induce the same homomorphisms on homology groups. As in the $\mathbb{Z}_2$-homology case, homotopy equivalent spaces have the same homology groups.

In the chapter on $\mathbb{Z}_2$-homology, we also introduced the concept of cellular homology, where simplices were grouped together to form cells. We showed that the $\mathbb{Z}_2$-homology groups on a simplicial complex and on an associated cellular complex were the same. The same holds for $\mathbb{Z}$-homology, that is, the $\mathbb{Z}$-homology groups of a complex K are identical to the $\mathbb{Z}$-homology groups of an associated cellular complex.

Since the proofs about homotopy invariance and cellular complexes for $\mathbb{Z}$-homology are essentially identical to those you have already done in the $\mathbb{Z}_2$-homology chapter, we will not repeat that development here, but we invite you to check them for yourself.

18.6 Homology and the Fundamental Group

Effective Thinking Principle. *Make Connections.* When you observe similarities in apparently different contexts, ask if there is a reason.

When we started investigating methods for recognizing holes in spaces, our first attempt led to the idea of the fundamental group. That idea involved looking at loops that surround holes and viewing two loops as equivalent if they were homotopic. If we think of a loop as a map of a circle, then a homotopy is the map of a cylinder. So two loops were viewed as the same in the context of the fundamental group if in some sense they co-bounded a cylinder.

In the case of the first homology, we again look at loops, in this case called 1-cycles. Two 1-cycles are viewed as the same in homology if they co-bound a 2-chain. Other than on the boundary, that 2-chain has the property that its 2-simplices must have boundaries that cancel out. Roughly speaking that means that the 2-chain is morally a 2-manifold with boundary equal to the two 1-cycles that are homologous.

So 2-manifolds have conceptually appeared in our thinking about the first homology of a complex. But remember that homology respects orientability, so the 2-manifolds that are appearing are actually orientable 2-manifolds. So, intuitively speaking, two loops that are equivalent in homology must co-bound an orientable 2-manifold with boundary equal to the two loops.

You may recall from the polygonal representation proof of the classification of compact, connected, triangulated 2-manifolds that the connected sum of tori were represented by sequences of edges of the form $aba^{-1}b^{-1}cdc^{-1}d^{-1}$.... Each letter corresponded to a loop in the 2-manifold. And the form $aba^{-1}b^{-1}$ suggests a commutator.

So suppose you have two loops that may be different in the fundamental group but are the same in the first homology. Then in some sense commutators of loops have made an appearance in the 2-chain that the two loops co-bound. This intuitive exploration suggests that commutators of elements of the fundamental group may correspond with trivial elements of the first homology group. That suggestion leads us to consider whether it might be true that the commutator subgroup of the fundamental group exactly captures the difference between the fundamental group and the first homology group. Satisfyingly, that connection is exactly correct.

Theorem 18.37. *Let K be a finite, connected simplicial complex. Then*

$$H_1(K;\mathbb{Z}) \simeq (\pi_1(K))/[\pi_1(K), \pi_1(K)],$$

that is, the first homology group of K is isomorphic to the abelianization of the fundamental group of K.

Pinning down the intuition that brought us to this conclusion with help us to prove this theorem. Let $\phi : \pi_1(K) \to H_1(K;\mathbb{Z})$ be the map that takes an element $[\alpha]$ of $\pi_1(K)$ to the element $[\alpha_{\#}(\mathbb{S}^1)]$ of $H_1(K;\mathbb{Z})$ where α is understood to be a simplicial map from $\mathbb{S}^1$ into K. First note that ϕ is a well-defined, surjective homomorphism.

It remains to show that $\mathrm{Ker}(\phi)$ is the commutator subgroup of $\pi_1(K)$. For this purpose, let $[\alpha] \in \mathrm{Ker}(\phi)$. Let $C^2 = \sum_{i=1}^{k} c_i \sigma_i^2$ be a 2-chain such that $\partial(C^2) = \alpha_{\#}(\mathbb{S}^1)$. Since $\partial(C^2) = \alpha_{\#}(\mathbb{S}^1)$, for each edge of any σ_i^2 that is not in $\alpha_{\#}(\mathbb{S}^1)$, that edge must be cancelled out when computing $\partial(C^2)$. So we can create an abstract 2-manifold whose 2-simplexes are the σ_i^2's where we take several copies of a simplex depending on its coefficient. Using the Classification Theorem of oriented 2-manifolds, we can recognize that α is in the commutator subgroup of $\pi_1(K)$.

18.7 The Degree of a Map

In this section we use $\mathbb{Z}$-homology to study maps from an n-spheres to itself. In some sense, every map of a sphere to itself coats itself some number of times. In the case of $\mathbb{S}^1$, a map of a circle to itself winds around some number of times forwards or backwards. Of course the fundamental group captures that fact, and the number of times around is sometimes called a **winding number**.

We would like to generalize the idea of the winding number and pin down the idea of how many times a map from a sphere to itself coats itself. Homology will allow us to formulate a useful notion of this coating process. Recall that $H_n(\mathbb{S}^n) \cong \mathbb{Z}$.

Definition. Let $f : \mathbb{S}^n \to \mathbb{S}^n$ be a continuous map. Then $f_* : H_n(\mathbb{S}^n) \to H_n(\mathbb{S}^n)$ is a homomorphism from $\mathbb{Z}$ to itself. Hence it represents multiplication by some integer, called the **degree of** f and denoted $\deg f$.

Note how this notion is tied to $\mathbb{Z}$-homology. In $\mathbb{Z}_2$-homology, this analysis would give just a binary categorization of maps.

Lemma 18.38. *If $f : \mathbb{S}^n \to \mathbb{S}^n$ is continuous, then $\deg f$ is well-defined. That is, it does not depend on the way in which we identify $H_n(\mathbb{S}^n)$ with $\mathbb{Z}$.*

Theorem 18.39. *Let $f, g : \mathbb{S}^n \to \mathbb{S}^n$ be continuous maps.*

1. *If f and g are homotopic, they have the same degree.*

2. $\deg(f \circ g) = (\deg f) \cdot (\deg g)$

Theorem 18.40. *The identity map on $\mathbb{S}^n$ has degree 1. The antipodal map has degree $(-1)^{n+1}$.*

An application of these theorems is the famous Hairy Ball Theorem.

Definition. A **vector field** on $\mathbb{S}^n$ is a continuous map $V : \mathbb{S}^n \to \mathbb{R}^{n+1}$ such that $\langle V(x), x\rangle_{\mathbb{R}^{n+1}} = 0$ for each $x \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$. The vector field V is *non-vanishing* if $V(x)$ is never the zero vector.

The following result is known as the Hairy Ball Theorem because, for $n = 2$, it can be interpreted to say that one cannot comb the hair on a billiard ball without leaving a bald spot or a cowlick. No one has ever understood why a billiard ball, which is aggressively hairless, is the metaphor of choice, but who are we to resist the wisdom of history.

Theorem 18.41 (Hairy Ball Theorem)**.** *There exists a non-vanishing vector field on $\mathbb{S}^n$ if and only if n is odd.*

18.8 The Lefschetz Fixed Point Theorem

The Brouwer Fixed Point Theorem asserts that a continuous map f from a ball to itself has a fixed point. This theorem holds for *any* continuous map. In particular, if we deform a map f by perturbing it a little bit the fixed point property persists. This analysis suggests that having a fixed point may be in some sense "robust" under homotopy. Of course, any two maps from a ball to itself are homotopic, since the ball is contractible. But for continuous maps on other spaces we may be able to obtain some related result.

For instance, consider the self-map of a circle: $g : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $g(x, y) = (-x, y)$ viewing $\mathbb{S}^1$ as a subset of $\mathbb{R}^2$. This "mirror-reversing" map clearly has two fixed points: $(0, 1)$ and $(0, -1)$. Moreover, for this map fixed points persist under homotopy: if we deform this map inside neighborhoods of those two fixed points, the resulting map will retain fixed points inside these neighborhoods. This fact can be seen by noting that in a neighborhood of one of those points, the second coordinate y is determined by the first coordinate x, so we can apply the Intermediate Value Theorem to $\Delta(x) = x - g_1(x, y)$ where g_1 is the projection function. At the endpoints of a neighborhood $\Delta(x)$ is positive when x is positive, negative when x is negative, so must have a zero somewhere inside. A zero for Δ corresponds to a fixed point for g.

On the other hand, the identity map $i : \mathbb{S}^1 \to \mathbb{S}^1$ fixes each point of the circle, but it is homotopic to the map that rotates every point by a small angle θ that has no fixed points. So in this instance, the fixed points do not persist under homotopy.

The Lefschetz Fixed Point Theorem tells us when a homotopy class of maps on a space K has fixed points, based on the computation of a number associated to the map.

We might first consider what happens to a simplicial map f: the analogue of a fixed point of f is a fixed simplex under $f_{\#}$, the induced map on chains. For a finite simplicial complex, the chain group $C_n(K)$ is freely generated by a basis consisting of the n-simplices of K. So the homomorphism $f_{\#n} : C_n(K) \to C_n(K)$ can be represented by a matrix M in that basis. The nonzero diagonal entries of M would indicate whether any simplices were sent to themselves; thus the *trace* of M is a number that captures the number of "fixed" simplices in dimension n.

This insight leads us to carefully define the notion of the trace of a homomorphism on a finitely generated abelian group.

Definition. Let G be a finitely generated abelian group, which we can express as $G = G_{free} \oplus G_{tor}$, the direct sum of a free group and a torsion group. Let $h : G \to G$ be a homomorphism. Then h induces a homomorphism on $h_{free} : G_{free} \to G_{free}$. The **trace** of h, denoted $\mathsf{Tr}(h)$, is the trace of a matrix that represents h_{free} with respect to any chosen basis of G_{free}.

Exercise 18.42. *In the definition above and using a little linear algebra, show that* $\mathsf{Tr}(h_{free})$ *does not depend on the choice of basis for* G_{free}.

Of course determining $\mathsf{Tr}(f_{\#n})$ could be very cumbersome if $C_n(K)$ is very large. Furthermore, there is no reason to expect homotopy invariance—in fact, even if f and g are homotopic, it is not necessarily true that $\mathsf{Tr}(f_{\#n})$ equals $\mathsf{Tr}(g_{\#n})$. On the other hand, this trace idea seems potentially promising. It may be useful to study the trace of the induced homomorphism f_* on homology groups, for which we know homotopy invariance holds, since we showed that $f_* = g_*$ if f and g are homotopic.

Exercise 18.43. *Construct a simple example of a map homotopic to the identity map on the triangulated circle whose induced chain map does not have the same trace as the identity chain map.*

For a given simplicial map f, is there a connection between $\mathsf{Tr}(f_{\#n})$ and $\mathsf{Tr}(f_{*n})$? One potential connection is that elements of $H_n(K)$ are cycles (modulo boundaries), and cycles are a subgroup of $C_n(K)$. Those relationships can be expressed by the short exact sequences:

$$0 \to B_n(K) \to Z_n(K) \to H_n(K) \to 0$$

and

$$0 \to Z_n(K) \to C_n(K) \to B_{n-1}(K) \to 0.$$

The map f induces homomorphisms on all these groups. The following theorem will be helpful to understand the relationship between traces of all these homomorphisms and will help us prove the Hopf Trace Formula.

Theorem 18.44. *Suppose* $0 \to A \to B \to C \to 0$ *is a short exact sequence of finitely generated abelian groups and* $f_A : A \to A$ *and* $f_B : B \to B$ *are homomorphisms such that* $i \circ f_A = f_B \circ i$. *Then there is an induced homomorphism* $f_C : C \to C$ *that makes the following diagram commutative:*

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "i"] \arrow[d, "f_A"'] & B \arrow[r, "\pi"] \arrow[d, "f_B"'] & C \arrow[r] \arrow[d, "f_C"'] & 0 \\
0 \arrow[r] & A \arrow[r, "i"] & B \arrow[r, "\pi"] & C \arrow[r] & 0
\end{tikzcd}$$

Moreover,

$$\mathsf{Tr}(f_B) = \mathsf{Tr}(f_A) + \mathsf{Tr}(f_C)$$

Theorem 18.45. *(The Hopf Trace Formula) Let K be a finite simplicial complex and let $f : K \to K$ be a simplicial map. Then*

$$\sum (-1)^i \, \mathsf{Tr}(f_{\#n}) = \sum (-1)^i \, \mathsf{Tr}(f_{*n}).$$

The expression on the right has a special name.

Definition. Let K be a finite simplicial complex, and let $f : K \to K$ be a continuous map. The **Lefschetz number** of f is:

$$\Lambda(f) = \sum (-1)^i \, \mathsf{Tr}(f_{*n}).$$

The Tr notation here is the *trace* of a homomorphism on a free abelian group, whose properties we should explore.

Theorem 18.46. *(Lefschetz Fixed Point Theorem) Let $f : |K| \to |K|$ be a continuous map on a simplicial complex K. If $\Lambda(f) \neq 0$, then f has a fixed point.*

One way to proceed is to show that if f does not have a fixed point, then the map f has a simplicial approximation $\bar{f}$ for which $\bar{f}(\sigma) \cap \sigma = \varnothing$ for every simplex σ.

Exercise 18.47. *Compute the Lefschetz number of the "mirror-reversing" self-map of a circle: $g : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $g(x, y) = (-x, y)$ viewing $\mathbb{S}^1$ as a subset of $\mathbb{R}^2$. Argue that any reversing map of a circle must have a fixed point.*

18.9 $\mathbb{Z}$-Homology—A Step in Abstraction

Looking a physical cycles led to our creation of $\mathbb{Z}_2$-homology. But once we have created a concept, that concept then becomes part of our world. So we can subject it to good practices of generalization and extension and use it as a source of analogies in the same manner as we take advantage of sights and insights in the real world. So $\mathbb{Z}$-homology is the result of a natural impulse to abstract and extend ideas we had before.

There are at least two kinds of complexity. One is complexity where steps present conceptual challenges. The other kind of complexity is where there are lots of details, but it is clear what each step means and why it follows. Our exploration of $\mathbb{Z}$-homology involved complexity of the second kind—lots of details, but essentially all of them were the result of a clear application of the strategy of moving from the $\mathbb{Z}_2$ world, which has a clearer physical interpretation, to the $\mathbb{Z}$ world where the algebraic features are more prominent.

Let's not forget to step back and see the forest. The bottom line of homology is that we can associate abelian groups with topological spaces that can help us distinguish spaces from one another. In the next chapter we take one further step of homological abstraction.

Chapter 19

Singular Homology: Abstracting Objects to Maps

We have now seen a couple of examples of homology theories—simplicial and cellular—and noted that they give the same answers on simplicial complexes when the coefficient group ($\mathbb{Z}$ or $\mathbb{Z}_2$) is fixed. And it will turn out that singular homology theory, which we discuss in this chapter, will give the same answers on simplicial complexes as well, though they hold for a larger class of spaces.

This consistency is no accident, because one can develop a homology theory axiomatically—in terms of the properties, or *axioms*, it satisfies. If the homology of any class of spaces can be determined strictly by these axioms, then any homology theory that satisfies these axioms must give the same homology groups on that class of spaces.

For example, think about some of the techniques we have used to simplify a homology computation: the long exact sequence of a pair, excision, and the fact that spaces of the same homotopy type have the same homology. To get the ball rolling with these techniques we also needed the homology groups of a specific simple space. As you will see, if we demand a homology theory satisfy these properties (the so-called *Eilenberg-Steenrod axioms*) then that is enough to compute the homology groups of all triangulable spaces.

And that is essentially why these theories will give the same answers in such cases. And that is why we use the same notation H_n for each one. And if we have multiple ways of computing homology, we can pick the one that is easiest to think about for various properties or computations we are interested in. For example, the topological invariance of homology turns out to be very easy to show in singular homology, whereas it requires work for simplicial homology. But computing the singular homology for an arbitrary space X might be complicated, while, if we know a small triangulation or cellular decomposition for that space, the simplicial homology may be relatively easy to compute.

Also, one of the main benefits of singular homology, as we shall see, is that it holds for all topological spaces, not just triangulable ones.

If you are bypassing simplicial homology in Chapter 18 and going straight to singular homology, be sure to read Section 18.3, which contains algebraic preliminaries.

19.1 Eilenberg-Steenrod Axioms

Effective Thinking Principle. *Seek Essentials.* Isolating the essential ingredients of a concept clarifies those features that are truly fundamental.

Before describing singular homology, we begin by presenting the Eilenberg-Steenrod axioms, which are properties that characterize homology groups on simplicial complexes. Then we will show that singular homology satisfies these axioms.

Definition. Let A be a subspace of the space X. We say (X, A) is a **compact triangulable pair** if and only if there exists a finite simplicial complex K, a subcomplex K', and a homeomorphism $f : (|K|, |K'|) \to (X, A)$.

Note that A may be empty, in which case the corresponding subcomplex K' is also empty.

Definition. A **homology theory** (H, ∂) **on compact triangulable pairs** is a sequence of functors and for each pair, a sequence of homomorphisms, described below.

1. There is a sequence of functors $\{H_n\}_{n\in\mathbb{Z}}$ that assigns to each compact triangulable pair (X, A) an abelian group $H_n(X, A)$ and to each continuous map $f : (X, A) \to (Y, B)$ a homomorphism $f_* : H_n(X, A) \to H_n(Y, B)$ satisfying:

   (a) if i is the identity map, the i_* is the identity homomorphism, and

   (b) for any pair of composable maps f and g, $(f \circ g)_* = f_* \circ g_*$.

2. For any pair (X, A), there is a sequence of homomorphisms $\{\partial_n : H_n(X, A) \to H_{n-1}(A, \varnothing)\}_{n\in\mathbb{Z}}$ such that each homomorphism is natural: for any map $f : (X, A) \to (Y, B)$, the following diagram commutes.

$$\begin{array}{ccc} H_n(X,A) & \xrightarrow{f_*} & H_n(Y,B) \\ {\scriptstyle\partial_n}\big\downarrow & & \big\downarrow{\scriptstyle\partial_n} \\ H_{n-1}(A,\varnothing) & \xrightarrow{(f|_A)_*} & H_{n-1}(B,\varnothing) \end{array}$$

These functors and homomorphisms must satisfy the **Eilenberg-Steenrod axioms**:

- Exactness Axiom: For inclusion maps $i : (A, \varnothing) \to (X, \varnothing)$ and $\pi : (X, \varnothing) \to (X, A)$, this sequence is exact:

$$\cdots \to H_n(A, \varnothing) \xrightarrow{i_*} H_n(X, \varnothing) \xrightarrow{\pi_*} H_n(X, A) \xrightarrow{\partial_n} H_{n-1}(A, \varnothing) \to \cdots .$$

- Homotopy Axiom: if f and g are homotopic maps, then $f_* = g_*$.

- Excision Axiom: for every (X, A) and every open subset U of X such that $\overline{U} \subset \operatorname{Int} A$, the inclusion $(X - U, A - U)$ in (X, A) induces an isomorphism

$$H_n(X - U, A - U) \cong H_n(X, A).$$

- Dimension Axiom: if P is a one-point space, then $H_n(P) \cong 0$ for all $n > 0$.

The group $H_0(P)$ is called the **coefficient group** of the homology theory.

For example, simplicial homology (either the $\mathbb{Z}$ or $\mathbb{Z}_2$ version) is a homology theory on compact triangulable pairs, setting $H_n(X, A)$ to be the relative homology groups and noting $H_n(X, \varnothing) \cong H_n(X)$. We showed the Dimension Axiom with relative ease. The Exactness Axiom for simplicial homology follows from the long exact sequence of a pair and the topological invariance of simplicial homology. We proved a version of the excision property for simplicial homology, but it isn't strong enough for the Excision Axiom, since we only showed excision is possible in the special case where removing U left a subcomplex of K. However, after seeing the proof of the excision property for singular homology, you may want to attempt the simplicial version. Similarly, you may not have fully established the Homotopy Axiom, but the singular version you will establish soon by constructing a chain homotopy may give you some ideas for how to do the simplicial version.

Here is the amazing theorem of Eilenberg and Steenrod (which we do not expect you to prove):

Theorem 19.1 (Eilenberg-Steenrod)**.** *Fix a group G. Any two homology theories on compact triangulable pairs with coefficient group G are isomorphic.*

This theorem means that for compact triangulable pairs, as long as we have homology groups with coefficients in $\mathbb{Z}$, induced homomorphisms and boundary maps that play nice with each other, and if the Exactness, Homotopy, Excision, and Dimension Axioms are satisfied, the computation of homology groups will produce the same answers.

This theorem has been generalized in various ways, such as to non-compact spaces (requiring an additional axiom) and other pairs of spaces (e.g., CW complexes).

19.2 Singular Homology

A drawback to using singular homology is that its definition is restricted to simplicial complexes. Of course this limitation has not prohibited us from proving many major results since we can, because of topological invariance, compute homology for any space homeomorphic to a simplicial complex. But topological invariance of simplicial homology was difficult to show.

Singular homology, on the other hand, will apply to any topological space, and its topological invariance will be almost trivial to show. Also, it will give the same homology groups on triangulable spaces, so once we define it, we can use it as an alternative to simplicial homology.

However, singular homology takes a little more work to define. Instead of breaking up the space X into simplices, we consider *maps* of simplices into X. The chain groups will be formal sums of such maps.

First we specify the domain of such maps. They will be simplices of various dimensions; for convenience, we place them in a common space $\mathbb{E}^\infty$, the **generalized Euclidean space**, which is defined to be the subspace of $\mathbb{R}^\infty$ (countably infinite product of copies of $\mathbb{R}$) with only finitely many non-zero coordinates. This space is metrizable, and any finite-dimensional subspace is isomorphic to $\mathbb{R}^m$ for some m.

Definition. In $\mathbb{E}^\infty$, let e_0 be the zero vector and e_i be the i-th standard basis vector which has 1 in the i-th coordinate and zeroes in every other coordinate. The **standard n-simplex** Δ_n is the convex hull of $\{e_0, e_1, \ldots, e_n\}$ in $\mathbb{E}^\infty$. The **barycentric coordinates** of a point x in the standard n-simplex are just the $n+1$ coefficients $(x_0, \ldots, x_n)$ from the convex combination $x = x_0e_0 + \cdots + x_ne_n$. The **$i$-th face of the standard n-simplex** is the set of points in Δ_n whose i-th barycentric coordinate is zero.

Definition. For $n \geq 0$, a **singular n-simplex** in a space X is a continuous map

$$\sigma : \Delta_n \to X.$$

Note that such a map could be really strange ("singular") and does not need to be one-to-one. Also, even simple spaces could have many singular simplices. For instance, if $X = R$, in every dimension n, there are uncountably many singular n-simplices.

Definition. For $n \geq 0$, the **singular n-chain group** $S_n(X)$ is the free abelian group generated by singular n-simplices. As before, we denote an eement of $S_n(X)$ by a formal sum of singular n-simplices with coefficients in $\mathbb{Z}$. Note that formal sums are always finite sums. For convenience, we set $S_{-1}(X) = \{0\}$.

Note that unlike the simplicial chain groups for finite simplicial complexes—which have only finitely many generators and are trivial above dimension n for an n-dimensional complex—the singular chain group can have uncountably many generators in any dimension, even for underlying spaces of finite simplicial complexes. So a singular n-chain group is a very large group!

It will be convenient to define the faces of a singular n-simplex $\sigma : \Delta_n \to X$. An $(n-1)$-dimensional face of σ should be a map whose domain is Δ_{n-1}, so we obtain the correct domain by prepending the map σ with a face map:

Definition. For $n \geq 1$ and $0 \leq i \leq n$, define the **i-th face map** $f_i^n : \Delta_{n-1} \to \Delta_n$ to be the affine linear map that sends the vertices e_k to e_k for $0 \leq k < i$ and sends e_k to e_{k+1} for $i \leq k \leq n-1$, so it omits e_i in the image. In barycentric coordinates, this map preserves the order of the coordinates but inserts a 0 in the i-th position, so the image of this map is just the i-th face of the standard n-simplex. Define the **i-th face operator**

$$\Phi_i^n : S_n(X) \to S_{n-1}(X)$$

to be the homomorphism specified on each singular n-simplex σ by $\Phi_i^n(\sigma) = \sigma \circ f_i^n$.

There is also a related operator that raises the dimension of a singular chain, by coning each simplex with a point. Of course the space must be star-convex with respect to that point to do this.

Definition. A subspace $X \subset \mathbb{E}^\infty$ is **star-convex** with respect to a point $x \in X$ if and only if for each point $z \in X$, the straight line between x and z is contained in X.

Definition. Let $X \subset \mathbb{E}^\infty$ be star-convex with respect to a point x. For a singular n-simplex $\sigma : \Delta_n \to X$, define the singular $(n+1)$-simplex $\mathrm{Cone}_x(\sigma) : \Delta_{n+1} \to X$ to be the map whose image is a "cone" over the image of σ to x: for each $d \in \Delta_n$, the map sends the line segment in Δ_{n+1} from $f_0^{n+1}(d)$ to e_{n+1} linearly to the line segment from $\sigma(d)$ to x in X. By extending this definition to a homomorphism of singular chains, we obtain the **cone operator**:

$$\mathrm{Cone}_x : S_n(X) \to S_{n+1}(X).$$

Exercise 19.2. *Let $X \subset \mathbb{E}^\infty$ be star-convex with respect to a point x. Verify that $\Phi_0^{n+1} \circ \mathrm{Cone}_x$ is the identity map on $S_n(X)$.*

We can now define an appropriate boundary operator by analogy with simplicial homology.

Definition. For $n \geq 1$, the **boundary of a singular n-simplex** σ is defined by

$$\partial_n(\sigma) = \sum_{i=0}^{n} (-1)^i \Phi_i^n(\sigma).$$

The boundary of a singular 0-simplex is defined to be zero. The **boundary of a singular n-chain** can be defined by extending the definition of ∂_n above linearly. Then ∂_n is a homomorphism between singular chain groups:

$$\partial_n : S_n(X) \to S_{n-1}(X)$$

for $n \geq 0$.

Definition. An element of $\operatorname{Im} \partial_{n+1}$ is a **singular n-boundary** of X: a singular n-chain that is the boundary of an singular $(n+1)$-chain. An element of $\operatorname{Ker} \partial_n$ is a **singular n-cycle** of a space X is a singular n-chain whose boundary is zero.

As expected, the boundary of a boundary is zero. The proof of this result is similar to the argument in simplicial homology.

Theorem 19.3. *For all $n \geq 0$,*

$$\partial_n \circ \partial_{n+1} = 0.$$

Hence $\operatorname{Im} \partial_{n+1} \subset \operatorname{Ker} \partial_n$.

We suppress writing the subscript n on ∂_n when there is no possibility of confusion. The sequence of chain groups and boundary maps

$$\cdots \xrightarrow{\partial} S_{n+1}(X) \xrightarrow{\partial} S_n(X) \xrightarrow{\partial} S_{n-1}(X) \xrightarrow{\partial} \cdots$$

is called the **singular chain complex**.

Definition. The **singular homology groups** (with coefficients in $\mathbb{Z}$) of a finite simplicial complex are the groups

$$H_n(X) = \operatorname{Ker} \partial_n / \operatorname{Im} \partial_{n+1}.$$

Two singular n-cycles α and β in the same homology class are called **equivalent** or **homologous**. Homologous cycles are denoted $\alpha \sim \beta$ and the homology class of the cycle α is denoted by $[\alpha]$.

By considering what the singular chain groups are, the following will be straightforward to show.

Theorem 19.4 (Dimension Axiom)**.** *If P is a point, $H_n(P) \cong 0$ for all $n > 0$, and $H_0(P) \cong \mathbb{Z}$.*

Definition. Any space with the homology groups of a point is called **acyclic**.

Theorem 19.5. *Let $X \subset \mathbb{E}^\infty$ be star convex with respect to $x \in X$. For any singular n-simplex σ,*

$$\partial_{n+1}(\operatorname{Cone}_x \sigma) + \operatorname{Cone}_x(\partial_n \sigma) = \sigma.$$

Theorem 19.6. *Show that any star-convex space is acyclic.*

Singular homology "sees" path-connectedness.

Theorem 19.7. *For a space X, show that $H_0(X)$ is a free abelian group with a generator for every path-connected component of X.*

19.3 Topological Invariance and the Homotopy Axiom

If $f : X \to Y$ is a continuous map, there is a natural homomorphism called the **chain map** between singular n-chains:

$$f_\# : S_n(X) \to S_n(Y)$$

which we can specify by its action on singular simplices: $f_\#(\sigma) = f \circ \sigma$. The chain map plays nicely with the boundary operator.

Theorem 19.8. *Let $f : X \to Y$ be a continuous map. Then for any chain $c \in S_n(X)$, $\partial(f_\#(c)) = f_\#(\partial(c))$. In other words, the diagram:*

$$\begin{array}{ccc} S_n(X) & \xrightarrow{f_\#} & S_n(Y) \\ \partial\big\downarrow & & \big\downarrow\partial \\ S_{n-1}(X) & \xrightarrow{f_\#} & S_{n-1}(Y) \end{array}$$

commutes.

Because of this property, the image of a cycle is a cycle and the image of a boundary is a boundary. So there is an **induced homomorphism** $f_* : H_n(X) \to H_n(Y)$ defined by $f_*([z]) = [f_\#(z)]$.

Exercise 19.9. *Check that the induced homomorphism is well-defined and a homomorphism.*

The next two theorems are the so-called **functorial properties** of the induced homomorphism.

Theorem 19.10. *The identity map $i : X \to X$ induces the identity homomorphism on each homology group.*

Theorem 19.11. *If $f : X \to Y$ and $g : Y \to Z$ are continuous maps between topological spaces, then $(g \circ f)_* = g_* \circ f_*$*

These functorial properties are far more easily established than their simplicial counterparts! Thus you can show the topological invariance of singular homology rather easily.

Theorem 19.12. *If $f : X \to Y$ is a homeomorphism, then $f_* : H_n(X) \to H_n(Y)$ is an isomorphism between singular homology groups.*

Recall that two maps are homotopic if, roughly speaking, we can deform one map into another continuously over time. It should come as no surprise, then, that homotopic maps f and g induce the same homomorphism in homology, since we expect that the image of a cycle under f would be homologous to the image of that cycle under g. The key is to produce a chain that witnesses that the image cycles are homologous.

Let's think about this situation a little more generally. Suppose f and g are maps from X to Y. Ignoring for the moment any homotopy between f and g, let's just ask under what conditions they induce the same homomorphism in homology.

If z is a chain in X that is an n-cycle, then the homology class of z satisfies $f_*([z]) = g_*([z])$ as long as $f_\#(z) - g_\#(z)$ is the boundary of a $(n+1)$-chain in Y. The main problem is how to find such a chain for every cycle.

As we have done with many proofs so far, we ask if we can define such a chain by doing it for each simplex. Thus we are looking for an operator $D : S_n(X) \to S_{n+1}(Y)$ such that if σ is a singular simplex in X, then $D\sigma$ is an $(n+1)$-chain in Y that ties together $f_\#(\sigma)$ and $g_\#(\sigma)$. So we expect the boundary of $D\sigma$ to involve $f_\#(\sigma)$, $g_\#(\sigma)$ and some n-chains that connect the corresponding boundaries of $f_\#(\sigma)$ and $g_\#(\sigma)$. Since ∂ commutes with $f_\#$ and $g_\#$, this quest is the same as looking for an n-chain whose boundary comprises the $(n-1)$-chains $f_\#(\partial\sigma)$ and $g_\#(\partial\sigma)$. But that is the problem of defining D one dimension lower. So we may as well be inductive, and set this term to be $-D\partial\sigma$. Then we can require D to be a homomorphism such that for a chain $c \in S_n(X)$, we seek $D : S_n(X) \to S_{n+1}(Y)$ such that:

$$\partial Dc = f_\#(c) - g_\#(c) - D\partial c.$$

Note that if z is a cycle, then the final term disappears and $\partial(Dz) = f_\#(z) - g_\#(z)$. Thus Dz is the desired chain that shows $f_*(z) = g_*(z)$.

A homomorphism $D : S_n(X) \to S_{n+1}(Y)$ such that

$$\partial D + D\partial = f_\# - g_\#$$

is called a **chain homotopy** between chain maps $f_\#$ and $g_\#$. Our discussion shows that when a chain homotopy exists between chain maps, they induce the same homomorphism in homology.

You have actually seen a chain homotopy before. Recall from Theorem 19.5 the cone operator Cone_x on a star-convex space satisfies

$$\partial \,\mathrm{Cone}_x + \mathrm{Cone}_x \,\partial = \mathsf{id}$$

where id stands for the identity operator. What this equation actually means is that the identity map id on chains is chain homotopic to the zero map on chains. So for cycles, every cycle differs from the empty chain by a boundary, i.e., is a boundary itself. This perspective shows why star-convex sets are acyclic (Theorem 19.6).

To prove the Homotopy Axiom for singular chains, you'll need to construct a chain homotopy.

Theorem 19.13 (Homotopy Axiom). *If f and g are homotopic maps from X to Y, then they induce the same homomorphism in homology.*

To show this fact, you should construct a chain homotopy D between $f_\#$ and $g_\#$ using the homotopy $F : X \times I \to Y$ that exists between f and g. Start by defining it on a singular simplex $\sigma : \Delta_n \to X$. The intuition is that the chain $D\sigma$ should "cover" the prism $F(\sigma(\Delta_n) \times I)$ in Y, but of course σ is singular and could have a crazy image in X so it's not obvious what simplices to use. So instead of trying to write $\sigma(\Delta_n) \times I$ as a sum of singular simplices, you might view it as the image of a "model space" $\Delta_n \times I$ under the map $\sigma \times i_I : \Delta_n \times I \to X \times I$, where i_I is the identity map on $I = [0, 1]$, and decompose the model space $\Delta_n \times I$ as a sum of simplices and push that sum forward by $F_\# \circ (\sigma \times i_I)_\#$. There are two potential strategies to decompose the model space into simplices. One is to construct a special triangulation of $\Delta_n \times I$. Another is to appeal to the acyclic nature of the model space and construct a chain homotopy inductively by dimension. The relation $\partial Dc = f_\#(c) - g_\#(c) - D\partial c$ provides the inductive step.

19.4 Relative Singular Homology

Relative homology can be developed for singular homology, just as we did with simplicial homology.

Definition. Let A be a subspace of a topological space X. By viewing singular simplices in A as singular simplices in X, the singular chain group $S_n(A)$ can be viewed as a subgroup of the chain group $S_n(X)$. Then we can define the quotient group:

$$S_n(X, A) = S_n(X)/S_n(A)$$

which we call the **group of relative singular chains of X modulo A**. An element in $S_n(X, A)$ can be represented by a chain involving only singular simplices whose images do not lie completely in A.

Note that the boundary map $\partial : S_n(X) \to S_{n-1}(X)$ restricts to $\partial : S_n(A) \to S_{n-1}(A)$ so that taking boundaries of chains in A stays in A. This observation produces a boundary map on relative chains.

Theorem 19.14. *There is a boundary map*

$$\partial : S_n(X, A) \to S_{n-1}(X, A)$$

such that $\partial_n \circ \partial_{n+1} = 0$ for all $n \geq 0$.

By analogy with the usual homology, $\operatorname{Ker}\partial_n$ is a subgroup of $S_n(X, A)$, consisting of relative n-cycles. A **relative n-cycle** can be represented by an n-chain in X whose boundary is a chain in A. And $\operatorname{Im}\partial_{n+1}$ is a subgroup of $S_n(X, A)$, consisting of relative n-boundaries. A **relative n-boundary** can be represented by an n-chain in X that, together with a n-chain in A, forms the boundary of an $(n+1)$-chain in X. The theorem above shows that the relative boundaries form a subgroup of the relative cycles, so we can define the **relative homology group** $H_n(X, A)$ as a quotient—the "relative cycles mod relative boundaries":

$$H_n(X, A) = \operatorname{Ker}\partial_n / \operatorname{Im}\partial_{n+1}.$$

Exercise 19.15. *Check that if $A = \varnothing$, the empty set, then $H_n(X, A) = H_n(X)$, the usual homology.*

Exercise 19.16. *Let X be the complex consisting of a triangle and all its faces. Determine $H_n(X, A)$ for all $n \geq 0$.*

Exercise 19.17. *Let K be a triangulation of an annulus, and let K' be the subcomplex consisting of the inner and outer edges of the annulus. Find a relative 1-cycle in $C_1(X, A)$ that is not a relative 1-boundary.*

Exercise 19.18. *Let K be a triangulation of a Möbius band, and let K' be its boundary. Determine $H_n(X, A)$ for $n \geq 0$.*

Now that we've defined the relative homology $H_n(X, A)$, we can consider simplicial maps on pairs that induce homomorphisms on relative homology. Suppose A is a subcomplex of X and B is a subcomplex of Y. We write

$$f : (X, A) \to (Y, B)$$

to denote a continuous map $f : X \to Y$ for which $f(A) \subset B$.

Theorem 19.19. *Given a continuous map $f : (X, A) \to (Y, B)$ there is an associated chain map $f_\# : S_n(X, A) \to S_n(Y, B)$ and induced homomorphism $f_* : H_n(X, A) \to H_n(Y, B)$.*

To understand the relationship between $H_n(X, A)$ to $H_n(X)$ and $H_n(A)$ we should examine the relationship between the associated chain groups. There is an obvious one from the definition of $S_n(X, A)$ as a quotient, namely, the short exact sequence:

$$0 \to S_n(A) \xrightarrow{i} S_n(X) \xrightarrow{\pi} S_n(X, A) \to 0$$

This short exact sequence leads to a long exact sequence in homology.

Theorem 19.20 (Long Exact Sequence of a Pair)**.** *Let A be a subspace of X. Then there is a long exact sequence:*

$$\cdots \xrightarrow{\partial_*} H_n(A) \xrightarrow{i_*} H_n(X) \xrightarrow{\pi_*} H_n(X, A) \xrightarrow{\partial_*} H_{n-1}(A) \xrightarrow{i_*} \cdots$$

where the maps are induced by the inclusion maps $i : A \to X$ and $\pi : (X, \varnothing) \to (X, A)$ and the boundary map $\partial : S_n(X) \to S_{n-1}(X)$.

19.5 Excision

The excision property for homology seems quite plausible when you think about it: a relative cycle in $H_n(X, A)$ should be representable by a chain that is mostly outside of A or near the boundary of A. So if we 'excise' out U, a portion of A away from the boundary, it shouldn't affect the relative homology group.

The excision property was easy to establish for simplicial homology when removing an open set U from A left a subcomplex behind (see Theorem 18.22). Removing U only removes simplices inside A, and so such simplices can be safely removed from chains of relative cycle and produce a relative cycle that is still homologous to the original cycle.

However, for singular homology, the problem is that singular simplices can be wild, and some of their images may not lie completely inside A or $X - U$. If so, then if we excise an open set U from A, we may rightly worry that a relative cycle in $S_n(X, A)$ involving such simplices may not be homologous to any relative cycle in $S_n(X - U, A - U)$.

In order to show that we do not need to worry, we want to show first that singular simplices can be subdivided sufficiently so that their images lie completely in X or $A - U$, and then show that we do not lose any homology classes when we perform such subdivisions, i.e., removing the 'big' simplices from the singular chain groups will not affect the homology of the chain complex.

We will need a subdivision operation, closely related to barycentric subdivision on simplices.

Definition. Let X be a topological space. We define the **barycentric subdivision operator on singular chains** $\mathsf{SD} : S_n(X) \to S_n(X)$ inductively as follows. If σ is a 0-simplex, define $\mathsf{SD}\,\sigma = \sigma$. Assuming SD has been defined for singular chains of dimension less than n, let $i : \Delta_n \to \Delta_n$ be the identity map (called the **standard singular n-simplex**), let b be the barycenter of Δ_n, and recall that Cone_b is the cone operator that takes singular chains and cones them to b. Then for any simplex $\sigma : \Delta_n \to X$, define

$$\mathsf{SD}(\sigma) = \sigma_{\#}(\mathrm{Cone}_b(\mathsf{SD}(\partial i))).$$

In words: we are taking the boundary of the standard singular n-simplex, applying the subdivision operator there (since it has been defined for dimensions less than n), and coning to the barycenter (which is possible since Δ_n is convex). This procedure produces a barycentric subdivision of the standard singular n-simplex. Then we push this subdivision forward to X by the map that σ induces on chains (applying σ to each singular simplex in the chain).

Exercise 19.21. *Verify that* SD *is a chain map, commuting with* ∂*, and verify that it is natural, which means that for any continuous map* $f : X \to Y$*, the following diagram commutes.*

$$\begin{array}{ccc} S_n(X) & \xrightarrow{f_\#} & S_n(Y) \\ {\scriptstyle \mathsf{SD}}\downarrow & & \downarrow{\scriptstyle \mathsf{SD}} \\ S_n(X) & \xrightarrow{f_\#} & S_n(Y) \end{array}$$

.

Given a cover of a space X, we will use the barycentric subdivision operator to reduce the size of the simplex images so that they lie completely inside one element of the cover. It may seem strange to refer to 'size', since X is not necessarily a metric space. However, the domain of each simplex is a metric space and this fact may help you prove the next theorem.

Theorem 19.22. *Let* $\mathcal{U}$ *be an open cover of a space* X*. For any singular simplex* σ*, there exists an* m *such that each term of* $\mathsf{SD}^m(\sigma)$ *has an image that lies within one of the elements of* $\mathcal{U}$*.*

Thus we have a method to make sure simplices can be subdivided enough to fit inside any cover. Next you will want to ensure that no homology class is lost when you ignore big simplices.

Theorem 19.23. *There is a chain homotopy between* SD^m *and the identity map on* $S_n(X)$*; in other words, there exists a homomorphism* $D_X : S_n(X) \to S_{n+1}(X)$ *such that*

$$\partial D_X \sigma + D_X \partial \sigma = \mathsf{SD}^m \sigma - \sigma$$

for every singular simplex σ *of* X*. Moreover, this chain homotopy is natural, meaning it commutes with maps of spaces: if* $f : X \to Y$*, then* $f_\# \circ D_X = D_Y \circ f_\#$*.*

As usual in these kinds of proofs, you will want to construct the chain homotopy inductively by dimension, first on a model space (the standard singular simplex), and then push forward.

Theorem 19.24. *For each* $n \geq 0$*, the induced homomorphism*

$$\mathsf{SD}^m_* : H_n(X) \to H_n(X)$$

is an isomorphism.

Theorem 19.25. *Suppose U and A are subspaces of X such that $\overline{U} \subset \operatorname{Int} A$. Then the inclusion map of $(X - U, A - U)$ in (X, A) induces an isomorphism*

$$H_n(X - U, A - U) \cong H_n(X, A).$$

Note that in singular excision, U is not required to be open, so this result is stronger than needed for the Eilenberg-Steenrod axiom.

Now that we have established the Eilenberg-Steenrod Axioms of Exactness, Homotopy, Excision, and Dimension, we illustrate an application that uses all four in a nice way—computing the homology groups of spheres.

Recall that S^n is the n-dimensional sphere and is the boundary of an $(n+1)$-dimensional ball. Spheres can be built up from two n-balls by gluing them along their boundaries. Or, said another way, if you take a n-sphere and cut it with a plane through its center, you get two hemispheres B^n_+ and B^n_-, each of which is a closed n-ball if you include their boundaries. But their common boundary is the $(n-1)$-sphere S^{n-1}. See if you can use these ideas to compute the homology of spheres.

Theorem 19.26. *For the $n \geq 1$, show that $H_n(S^n) \cong \mathbb{Z}$ and $H_n(S^k) \cong 0$ if $n \neq k$.*

Some hints: consider the pair (S^n, B^n_-). If you could excise the interior of B^n_-, it would leave you with the pair (B^n_+, S^{n-1}), but you can't quite do that because the excised set needs to have its closure contained in the interior of B^n_-. Can you get around that? Then compare $H_n(S^n, B^n_-)$ with $H_n(S^n)$ and $H_n(B^n_+, S^{n-1})$ with $H_{n-1}(S^{n-1})$ using the long exact sequences.

19.6 A Singular Abstraction

This chapter was the final step that we will take in starting with a concrete approach to finding holes in spaces by surrounding them with physical sets and step-by-step abstracting that process. The step of abstraction in using maps of simplices rather than simplices themselves is in some sense a rather natural process. Singular homology refers to uncountably many maps, but soon we realize that the geometric significance of those maps is captured in the original intuition that started the ball rolling.

The Eilenberg-Steenrod Axioms and Theorem present us with yet another example of the strategy of developing mathematics by seeking essentials and then following the consequences of our choices. One indication that the choices of essentials were well made is the fact that the motivating homology theorems are subsumed in the generality.

The steps of abstraction that we have seen in the development of homology theories provide us with a good metaphor for one manner in which mathematics develops. It also reminds us of the importance of returning regularly to motivating examples and insights in order to understand both the origins and the generalizations more deeply.

Chapter 20

The End: A Beginning—Reflections on Topology and Learning

We have reached the end of a beginning—the end of an introduction to a subject whose end is not in sight. Topology explores and exposes the essence of matters mathematical.

The first part of this book demonstrated how the heart of familiar mathematical ideas and objects can be expressed in set-theoretical terms. We started with the basic idea of counting and extended the idea to the concept of the cardinality of infinite sets. Arguably, one of the great triumphs of human thought is the counterintuitive realization that infinity comes in more than one size. Cantor's theorems prove that infinity is not a monolithic everything-ness, but instead infinity comes in infinitely many different sizes. Cantor's theorem is one jewel in a gem-encrusted crown of the exploration of the infinite.

We next defined what we mean by a topological space. That definition was intended to capture the set-theoretic essence of familiar spaces. The wisdom of the choices made in creating that definition was made clear in the richness of all the chapters that followed.

After defining a topological space, we developed concepts that reflected our intuition and experience with familiar mathematics, but framed them in the context of the world of topological spaces. That exploration let us see familiar spaces such as the Euclidean spaces and familiar ideas such as continuity in a new light. We saw connectedness and the concept of distance reflected through the lens of topological spaces. Seeing the familiar with more nuance became the pole star of our strategy for exploring the consequences of taking a topological perspective.

We are most interested in those parts of our world with which we are most familiar. In the domain of mathematics, the most familiar spaces are the Euclidean spaces and objects in them. So the entire last part of this book was an exploration of objects that are made from basic Euclidean pieces.

Two ways to construct Eucidean-based spaces are to think of spaces that are locally Euclidean (creating manifolds) or to think of spaces built from rectilinear building blocks (creating simplicial complexes).

We started the second part of this book by exploring 2-manifolds—that is, surfaces. Our description of those surfaces was driven by an urge to classify those spaces, which are all the same locally, but differ globally. In the case of compact, connected 2-manifolds, we successfully showed that every such object was built from combining simple 2-manifolds: the 2-sphere, the torus, and the projective plane.

The success in classifying 2-manifolds actually opened doors to new collections of questions and mathematical concepts and tools for answering those questions. One challenge we faced in distinguishing 2-manifolds was to pin down the difference between spaces that intuitively appeared different. What is the fundamental difference between a torus and a double torus? How can we pin that down? Answering those questions led us to develop some of the most powerful strategies for understanding and distinguishing topological spaces.

Associating a group with a space in a specified manner turned out to be an extremely effective way to create distinctions among spaces. The first idea we pursued came from the observation that we can encircle a hole in the plane, for example, by going around the hole some number of times. It required an effort of specificity and clarity of purpose to arrive at the definition of the fundamental group. Putting loops into equivalence classes based on homotopies created ideas that were extended and modified in many ways.

Fundamental groups and their relationships to covering spaces provided us with a satisfying study, but the fundamental group did not help us to deal with spaces with higher-dimensional holes. So we invented yet another method of associating groups with a space—in this case, a group for every dimension, the homology groups.

We saw that there is a direct connection between the fundamental group, which reflects structure captured by 1-dimensional loops, and the first homology of the space—namely, the first homology is merely the abelianization of the fundamental group. But the homology groups allow us to distinguish spaces whose differences are based on higher dimensional structure.

Homology was created by starting with, in some sense, the simplest way we could imagine to capture the presence of a 'hole' in a space, namely, thinking about putting an object around the hole. The physical strategy led to $\mathbb{Z}_2$-homology. The consequences of $\mathbb{Z}_2$-homology are enormous.

You proved many of the highlight theorems of topology using Z_2-homology. You proved the No Retraction Theorem, the Brouwer Fixed Point Theorem, the Borsuk-Ulam Theorem, the Ham Sandwich Theorem, the Jordan-Brouwer Separation Theorem, and many others. These theorems justify our characterization of Z_2-homology as a topological superhero.

One of the principal themes of mathematics is that once we have met with success, extensions and variations will lead to yet further successes. So it was with homology. After the success of Z_2-homology, we were driven to extend the idea to create Z-homology and singular homology. Those extensions and abstractions of the physical basis of Z_2-homology were created by taking all the theorems and techniques that we had constructed in the physical Z_2-homology world and see how we could abstract them a bit. In fact, the abstractions to Z-homology and singular homology followed the development of Z_2-homology closely, but then went beyond.

The new homology theories in their turn allow us to prove yet more fascinating theorems. The concept of the degree of a map allows us to prove theorems such as the Hairy Ball Theorem, which tells us which dimensional spheres admit non-zero vector fields. The stronger homology theories allow us to prove new fixed point theorems such as the Lefschetz Fixed Point Theorem.

The whole of our exploration of topology was the result of employing effective strategies of thinking that turn our minds in the direction of concept creation. Creating new mathematics or other new ideas is not magic—insights arise by employing practices of mind that inevitably lead to insight. In a real sense, the whole of topology is an example of the fabulous results that come about by thinking hard about basics. Soon the quest for understanding the essentials blossoms into a forest of beautiful ideas. The part of topology you have seen here is just a bit of undergrowth in that lush jungle—there is always more to come.

We hope your interactions with this book included many pleasant struggles and triumphs. Topology presents us with many delights. We have served you a bountiful feast of delectable treats, but we hope you leave hungry for more. We wish you a lifetime ahead of joyful grappling with ideas from topology and beyond.

Appendix A

Appendix - Group Theory Background

A.1 Group Theory

Definition. A **group** is a set G along with a binary operation $G \times G \to G$, typically denoted by $\cdot$ or juxtaposition satisfying the following three conditions:

1. There exists an element $1_G \in G$, called the **identity element** such that $g \cdot 1_G = 1_G \cdot g$ for all $g \in G$.

2. For every $g \in G$ there exists an element $g^{-1} \in G$, called the **inverse** of G, such that $g \cdot g^{-1} = g^{-1} \cdot g = 1_G$.

3. For all $g_1, g_2, g_3 \in G$ the **associative property** holds: $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$

The notation above for the group operation is generally called multiplicative notation. An alternative is so-called additive notation. The operation is denoted by '$+$', the identity element is denoted by 0_G, and the inverse of $g \in G$ is denoted by $-g$. When there is only one group in question (or when all the groups in question have the same identity element), we will typically drop the 'G' subscript from the notation for the identity.

Additive notation is usually reserved for *commutative* or *abelian* groups:

Definition. A group G is **commutative group** or **abelian group** if $g_1 \cdot g_2 = g_2 \cdot g_1$ for all $g_1, g_2 \in G$.

Often instead of using '$\cdot$' to denote the group operation, we use juxtaposition. In other words, $xy = x \cdot y$. We often use the verb 'to multiply' to indicate the group operation.

Definition. The **trivial group** is the group that contains only one element, namely the identity. In other words $G = \{1_G\}$ or $G = \{0_G\}$ (depending on which notation is being used).

Definition. Let G be a finite group. Then the cardinality, $|G|$, of the set G is called the **order** of G.

Definition. Let A be a set of n elements. Then a **permutation** is a bijective function from A to itself. Usually we use positive integers to describe A, that is $A = \{1, \ldots, n\}$. Let $\{a_1, \ldots, a_m\} \subseteq A$, then we use $(a_1 a_2 \ldots a_m)$ to represent the function that takes a_i to a_{i+1} for $1 \leq i \leq m-1$ and a_m to a_1 (and preserves all others). Such a permutation is called an *m-cycle*. A 2-cycle is called a **transposition**.

Exercise A.1.

1. *Show that the set of all permutations on n elements forms a group with the group operation of function composition.*

2. *Show that any permutation can be written as a composition of disjoint cycles.*

3. *Show that any m-cycle can be written as a composition of transpositions.*

Definition. The group of all permutations on the first n positive integers is called the **symmetric group**, denoted by S_n.

Exercise A.2. *What is the order of S_n?*

Note that S_n is not an abelian group for $n \geq 3$.

Definition. A permutation is **even** if it can be written as the composition of an even number of transpositions and *odd* otherwise.

Exercise A.3.

1. *Show that an n-cycle can be written as the composition of $n-1$ transpositions. Thus a 3-cycle is an even permutation and a 4-cycle is an odd permutation!*

2. *Show that the group of even permutations is a subgroup of S_n.*

Definition. The group of even permutations is called the **alternating group**, denoted by A_n.

Exercise A.4. *What is the order of A_n?*

Definition. The symmetry group of a regular n-sided polygon (under composition) is called the **dihedral group**, denoted D_n

Exercise A.5. *Show that if we let a represent a reflection along a line passing through the polygon's center and a vertex, and b a rotation of $2\pi/n$ around its center, then*

$$D_n = \{1, b, \ldots, b^{n-1}, ab, \ldots, ab^{n-1}\}$$

Exercise A.6. *Show that in D_n as above, we have $ab = b^{n-1}a$, and thus D_n is not abelian for $n > 2$.*

Definition. A **subgroup** of a group G is a subset H that forms a group under the operation it inherits from G. We write $H < G$.

Since the symmetries of a polygon induce permutations on its vertices, it is easy to see that $D_n \cong H \subset S_n$, and $H \neq S_n$.

Exercise A.7. *Show that D_n is isomorphic to a proper subgroup of S_n.*

Exercise A.8. *Under what conditions, if ever, is D_n is isomorphic to a subgroup of A_n?*

Definition. Let $g \in G$, a group, and H be a subgroup of G. Then the **left coset** of H by g is

$$gH := \{gh | h \in H\}.$$

We can define the **right coset** Hg similarly.

Exercise A.9. *Let $g, g' \in G$. Then either $gH = g'H$ or $gH \cap g'H = \varnothing$.*

Definition. Let H be a subgroup of G, then the **index** of H in G, denoted $[G : H]$, is the number of left cosets of H in G.

Theorem A.10 (Lagrange's Theorem)**.** *Let G be a finite group, and H a subgroup. Then the cardinality $|H|$ of H divides the cardinality |G|of G and*

$$[G : H] = \frac{|G|}{|H|}$$

Definition. A subgroup H of G is called a **normal subgroup** of G (denoted $H \lhd G$) if $gHg^{-1} = H$, where $aHb := \{ahb | h \in H\}$.

Multiplying a group or an element on the left by one element and on the right by its inverse is called *conjugation*, so a normal subgroup is one which is unchanged (set-wise) by conjugation.

Theorem A.11. *Let $H \lhd G$ be a normal subgroup. Then its left and right cosets coincide for all $g \in G$, in other words $gH = Hg$ for all $g \in G$.*

Definition. The **direct product** $G \otimes H$ of two groups G and H is the set $G \times H$ with the group operation defined by $(g, h) \cdot (g', h') = (gg', hh')$. When the groups are additive we call this **direct sum** and write $G \oplus H$.

Definition. A function $f : G \to H$ is a (group) **homomorphism** if $f(g \cdot g') = f(g) \cdot f(g')$ for all $g, g' \in G$.

In other words f preserves the group structure in the image of G.

Definition. A bijective homomorphism $f : G \to H$ is an **isomorphism**, in which case we say G is isomorphic to H and write $G \cong H$.

But what about when the homomorphism is not bijective?

Definition. The **kernel** of a homomorphism $f : G \to H$ is

$$\operatorname{Ker} f := \{g \in G | f(g) = 1_H\}$$

Theorem A.12. *An onto homomorphism $f : G \to H$ is an isomorphism if and only if* $\operatorname{Ker} f = \{1_G\}$.

Theorem A.13. *Let $f : G \to H$ be a homomorphism from a group G to a group H, then* $\operatorname{Ker} f \lhd G$ *and* $f(G) < H$.

Definition. A normal subgroup $N \lhd G$ has its left cosets equal its right cosets: $gN = Ng$. Therefore the set

$$G/N := \{gN | g \in G\}$$

of all left cosets of N is a group with the group operation

$$(gN) \cdot (g'N) := gg'N$$

This group is called the **quotient group** of G by N.

Definition. Let H be a subgroup of G. Then the **normalizer** of H in G is $N(H) = \left\{g \in G | gHg^{-1} = H\right\}$.

Note that $N(H)$ is a subgroup of G, $H \lhd N(H)$, and that it is the largest subgroup of G in which H is normal, meaning that any subgroup of G containing H in which H is normal must be contained in $N(H)$.

Theorem A.14 (First isomorphism theorem)**.** *Let $f : G \to H$ be a homomorphism with* $\operatorname{Ker} f = N$. *Then $f(H) \cong G/N$.*

Definition. Let $g \in G$. Then $\langle g \rangle$ the **cyclic subgroup generated by** g is the subgroup formed by all powers of g:

$$\langle g \rangle := \{g^n | n \in \mathbb{Z}\}$$

where $g^n = \overbrace{g \cdot g \cdots g}^{n \text{ times}}$ if $n > 0$, $g^0 = 1$, and $g^{-n} = \overbrace{g^{-1} \cdot g^{-1} \cdots g^{-1}}^{n \text{ times}}$ for $n \in \mathbb{N}$.

Note that with additive notation $\langle g \rangle = \{ng | g \in G, n \in \mathbb{Z}\}$, where

$$ng = \overbrace{g + g + \cdots + g}^{n \text{ times}}$$

for $n \in \mathbb{N}$, $g^0 = 0$, and $-ng = \overbrace{g + g + \cdots + g}^{n \text{ times}}$ for $n \in \mathbb{N}$.

Definition. If $G = \langle g \rangle$ for some $g \in G$ we say G is a **cyclic group** with *generator* g.

Note that cyclic groups are abelian. If $G = \langle g \rangle$ and there exists $n \in \mathbb{Z}$ such that $g^n = 1$, then there exists a smallest $n \in \mathbb{N}$ such that $g^n = 1$. This n is the order of G.

Theorem A.15. *A cyclic group that is infinite is isomorphic to* $\mathbb{Z}$*.*

Theorem A.16. *A finite cyclic group of order* n *is isomorphic to* $\mathbb{Z}_n$*, the integers with addition* $\mod n$*.*

Definition. A group $G \cong \overbrace{\mathbb{Z} \oplus \mathbb{Z} \oplus \ldots \oplus \mathbb{Z}}^{n \text{ times}}$ is called the **free abelian group of rank** n. G has a generating set of n elements of infinite order, one for each $\mathbb{Z}$ factor.

Definition. Let G be a group and $S \subseteq G$. Then the smallest subgroup H of G containing S is called the **subgroup generated by** S. If $H = G$ then we say G is **generated** by S, or that S generates G.

Note that the set of generators of a group is by no means necessarily unique. We can view the subgroup H generated by S as the set of all possible products $g_1 g_2 \ldots g_n$ where $g_i \in S$ or $g_i^{-1} \in S$. We can also view H as the intersection of all subgroups of G that contain S.

Exercise A.17.

1. *Verify that the dihedral group* $D_n = \{1, b, \ldots, b^{n-1}, ab, \ldots, ab^{n-1}\}$ *is generated by* $\{a, b\}$*.*

2. *Show that the symmetric group* S_n*, for* $n \geq 2$*, is generated by the set of* 2*-cycles:* $\{(12), (23), \ldots, (n-1, n)\}$.

3. *Show that the symmetric group* S_n*, for* $n \geq 2$*, is generated by the pair of cycles* (12) *and* $(12 \ldots n)$*.*

Definition. A group is **finitely generated** if there exists a finite subset S of G that generates G.

Theorem A.18 (Classification of Finitely Generated Abelian Groups)**.** *Let* G *be a finitely generated abelian group. Then* G *is isomorphic to:*

$$H_0 \oplus H_1 \oplus \ldots \oplus H_m$$

where H_0 *is a free abelian group, and* $H_i \cong \mathbb{Z}_{p_i}$ $(i = 1, \ldots, n)$ *where* p_i *is a power of a prime. The rank of* H_0 *is unique and is called the* ***rank*** *of* G*. The orders* $p_1, \ldots, p_m$ *are also unique up to reordering.*

Definition. A **commutator** in a group G is an element of the form $ghg^{-1}h^{-1}$. The **commutator subgroup** G' is the subgroup generated by the commutators of G.

Theorem A.19. $G' \lhd G$*, and is the smallest subgroup for which* G/G' *is abelian. In other words, if there is a subgroup* $N \lhd G$ *such that* G/N *is abelian, then* $G' \subset N$*.*

Definition. If G is a group, and G' is its commutator subgroups, the group G/G' is called the **abelianization** of G.

Theorem A.20. *Isomorphic groups have isomorphic abelianizations.*

There is a useful notation for groups that, roughly speaking, uses the fact that if we know a set of generators and the "rules" (called "relations") to tell when two elements are the same, then the group (up to isomorphism) is determined by a list of these generators and relations. What follows is a very non-technical description of the generator-relation notation for groups.

For example, in D_n (as described above) it is enough to know that there are two generators a and b, of order 2 and n respectively, and that they satisfy $ab = b^{n-1}a$. These facts determine a complete list of elements $1, b, \ldots, b^{n-1}, ab, \ldots, ab^{n-1}$. The expressioin $ab = b^{n-1}a$ can be written as $aba^{-1}b = 1$, and the word $aba^{-1}b$ is called a *relation*. By a and b's order, we also know that a and b satisfy $a^2 = 1$ and $b^n = 1$. The two letters a and b, together with the above relations completely determine the group D_n, and thus we write:

$$D_n = \langle a, b | a^2, b^n, aba^{-1}b \rangle$$

Similarly, we can write the cyclic group of order n as:

$$C_n = \langle a | a^n \rangle.$$

We can write the infinite cyclic group as:

$$C_\infty = \langle a | \quad \rangle.$$

We can denote the free abelian group of rank n as:

$$F_n^{\text{ab}} = \langle a_1, \ldots, a_n | a_i a_j a_i^{-1} a_j^{-1}, i \neq j \in \{1, \ldots, n\} \rangle$$

Exercise A.21. *Confirm that the lists of generators and relations given above completely determine the groups.*

We should note that since the relations $g \cdot g^{-1} = 1$, $g \cdot 1 = g$ and $1 \cdot g = g$ hold for any $g \in G$, as they are implicit in the definition of a group, such relations are not included in the list of relations. In general a group G can be written as $G = \langle L | R \rangle$, where L is a collection of generating elements (often called 'letters') and R is a collection of relations between them (that is words in the letters which are trivial in the group). This is called a **group presentation** G.

This notation is very useful, especially when dealing with the fundamental group and Van Kampen's theorem. The problem with this notation, however, is that it is *very* difficult, in general, given two groups with this notation, to tell if the groups are isomorphic or not, or even if two words represent the same group element. A group that can be presented with a finite number of generators and a finite number of relations is called **finitely presented**.

Theorem A.22. *Suppose that L is a set and R is a collection of words in L. Then there is a group whose presentation is $\langle L|R\rangle$.*

Exercise A.23. *What is a group presentation for an arbitrary finitely generated abelian group? for the symmetric group?*

Definition. For $n \in \mathbb{N}$, the **free group on n letters** is group F_n whose presentation is $\langle e_1, \ldots, e_m|\varnothing\rangle$. That is, F_n is the group which is generated by n letters with no relations between them. In general any group which can be given a presentation without generators is called free. We say that n is the **rank** of the free group F_n.

Theorem A.24. *For $m \neq n$, $F_m \neq F_n$.*

An important construction for creating a new group from other groups is the *free product*. We take two disjoint groups and construct a new group from the elements in each. The only relations are those from the original group, except that we assume that the two identities are equal (and also equal to the identity in the free product). The formal definition is below.

Definition. Let G and H be disjoint groups. Let $G = \langle S, R\rangle$ and $H = \langle T, Q\rangle$ be presentations. Then the **free product of G and H** is $G * H = \langle S \cup T, Q \cup R\rangle$.

Lemma A.25. *We have a natural injections $G \to G * H$ and $H \to G * H$ so that G and H may be viewed as subgroups of $G * H$.*

Theorem A.26. *Let G and H be disjoint groups. Each element in $G * H$ has a unique expression of the form $g_1h_1 \cdots g_nh_n$ where $g_1, \ldots, g_n \in G$, $h_1, \ldots, h_n \in H$, and g_1 and h_n are allowed to be the identity, but no other letter is.*

Index