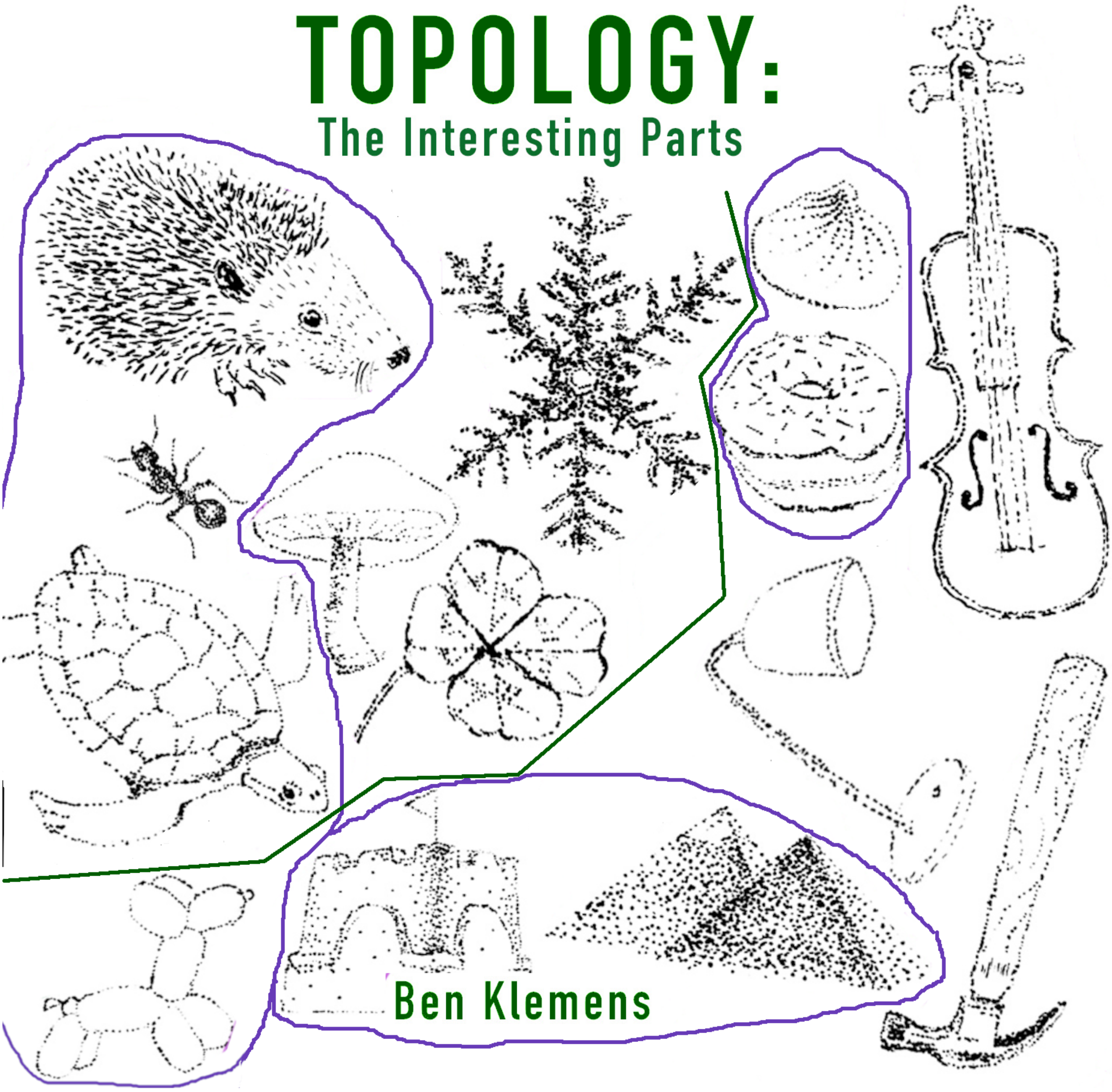
TOPOLOGY:

The Interesting Parts



Ben Klemens

This is a math book for mathematically inclined people who don't read much math. It covers Topology, the study of open sets. When I started studying it, I was surprised by how often I was surprised. The story begins with a very broad definition of *open set*, just shy of *whatever you want to call an open set*, but from that seemingly overbroad starting point, all sorts of constructions emerge almost magically. There will be stretchy doughnuts and fractals, but the bulk of the book will be an inquiry into one of the key philosophical problems of mathematics: how an infinite number of points cohere into a line or a surface. Topology has a number of distinct answers to that question, none of which depend on formalizing some concept of infinitesimal distances, revealing that it is connection, not closeness, which is vital. It is a field filled with eccentric counterexamples, each a glimpse into a world much like ours, but strangely different.

Contents

Chapter 1

Introduction: Interesting

Which is bigger, the positive integers $\{1,2,3,4,\ldots\}$ or the even positive integers, $\{2,4,6,8,\ldots\}$? The intuitive answer is that the first is larger than the second, because it consists of the evens plus the odds, while the second is just the even half. But if we think about it more, we see that they are in fact the same size, because we can write a one-to-one mapping, $1 \to 2$, $2 \to 4$, $3 \to 6$, $N \to 2N$, so that every element of the first set has exactly one corresponding element in the second, and a vice-versa one-to-one mapping in the other direction from even integers to all integers, $2 \to 1$, $4 \to 2$, $6 \to 3$, $N \to N/2$.

Or recall Zeno's Paradox: a tortoise (in the retelling of Hofstadter [1979]) crawling from zero to one meter will never fully cross the one meter mark, because first the tortoise's center of gravity must reach half a meter, then must reach $\frac{3}{4}$ meters, then $\frac{7}{8}$ meters, then $\frac{15}{16}$ meters, and to reach an infinite number of checkpoints, the tortoise's center of gravity will need infinite time. That Zeno's Paradox gives us any pause indicates that our intuition about the infinitely small is just as bad as our intuition about the infinitely large. They are in many ways the same problem, as $\{2,3,4,5,\ldots\}$ easily turns into $\left\{\frac{1}{2},\frac{1}{3},\frac{1}{4},\frac{1}{5},\ldots\right\}$.

David Foster Wallace, in *Everything and More: A Compact History of Infinity* [Wallace, 2003], pitches Zeno's Paradox as one of the core conundrums of mathematics from antiquity to the mid-1900s. The Calculus you learned in high school or college, about infinitely small differentials, clearly depends on all those infinitely small amounts adding up to something finite and definite. Wallace teased Topology in his book's title: Compactness (Chapter 8) will be a key tool by which topologists understand Zeno's paradox, and therefore a key tool in understanding the infinities in our lives.

Topology is the study of open sets. This immediately engenders the question: What is an open set? And the answer is: whatever set of points we choose. Topology is a world-building exercise in which we take a set of points, maybe the real number line, maybe the numbers $\{3, 7, 14\}$, maybe the members of the FC Barcelona futbol team, and we write down a list of groupings—whatever we want—which we declare to be open sets. The character of the world we build will depend on how we choose to invent the concept of open sets. Some will be much like the world we live in now, and others will be strangely different.

The difference between an open set and any other, like a closed set, is that the infinite union of open sets is guaranteed to also be an open set. That infinite union rule will be the (seemingly insufficient) hook we use to ask more questions of the infinitely large and infinitely small.

Those of you familiar with topology as presented in popular media will note that the question of confronting infinities and how smooth surfaces are built from points ("point-set topology") has little to do with doughnuts made of rubber sheets ("rubber-sheet topology"). There will be a limited sampling of stretchy doughnuts, in Chapter 6. But I personally find it more interesting to contemplate where those rubber sheets came from. I mean, have a look at your skin. You know from your Biology classes that it is made from individual cells; you know from your Physics classes that those cells are made from atoms linked to form chemical compounds, linked to form cells. When you look at your skin, you see none of that. Somehow, by some miracle, those individual parts cohere into a surface. Individual cells are so interlinked that they become, for our purposes, a unit. We won't cover anything so applied as biology, chemistry, or topography (the study of the features of the Earth's surface). But the same thing happens with theoretical points, as we take points at $\frac{1}{2}$, $\frac{3}{4}$, $\frac{7}{8}$, $\frac{15}{16}$, and everything in between, and we get a line, a thing we see as a coherent entity and not just an unstructured pile of points. The turtle does cross the finish line.

1.1 Surprise

Putting *The interesting parts* in the title of this book immediately raises the question of what *interesting* means. The ever-quipping Richard Feynman had some snark about how all mathematical results, once discovered and proven, are obvious. They take the assumptions, apply the rules, and get the result.

But there's still room for surprise.

Martin Gardner was an author who had a column in Scientific American on recreational math. One of his books, with kid-friendly cartoon illustrations, is entitled *The Unexpected Hanging* [Gardner, 1991]. In the lead story, the guard tells the prisoner that he's going to hang him at dawn some day between today and Sunday, and the exact day will be a surprise. *I won't get hung on Sunday, then,* the prisoner thinks, *because it wouldn't be a surprise. But if Sunday is impossible, and the call comes in on Saturday, I won't be surprised—Saturday would be the only option left.* If the prisoner knows it can't be Sunday and Saturday is as impossible as Sunday, repeat the logic, and Friday is impossible, because on Thursday night the prisoner will know that it's Friday morning. Repeat, and every day is eliminated!

The guard shows up Thursday morning and the prisoner is completely surprised.

There are different arguments to make about what went wrong with the impeccable logic. But the logic chain set an expectation—nothing—and "surprise" means an expectation defied.

Interesting math breaks our expectations. If you have a reasonable idea of what an open set looks like (I picture a blob with fuzzy edges), then it's not surprising that an open set on the real number line has no maximal element. Remember those ϵ-δ (epsilon-delta) proofs from college calculus, defining a continuous function? Of course you don't, it's a mess: "A function is continuous iff, for every ϵ, there exists a δ such that if $|x_1 - x_2| < \delta$, then $|f(x_1) - f(x_2)| < \epsilon$." I had to memorize this—in fact, you couldn't graduate from the college I went to without knowing this incantation (and knowing how to swim). It won't be until Chapter 4 that $f^{-1}(U)$ will be fully explicated, but you can already appreciate how much simpler the equivalent statement is via topology: "a function is continuous iff, when U is an open set, then $f^{-1}(U)$ is an open set." I was flabbergasted when I first saw this. How did it happen that all that complicated mess just fell away? When I was memorizing where the ϵ goes I hadn't contemplated that such a simplification could be possible.

A mathematician to whom I showed this book defended the ϵ-δ form, saying that it encapsulates a simple concept: if there is a small change in x, there is a small change in $f(x)$, and formalizing that in algebraic instead of topological terms just requires more machinery. Even this points to potential interesting: we have one definition about open sets, one about small changes, and somehow they are (given a metric as in §2.2) equivalent. How did that happen? And if the two are equivalent, but the story is simpler without infinitessimal distances, maybe the story isn't about distances at all.

I've been reading recreational math for as long as I can remember. In lieu of day care, I'd sometimes get dumped at the University of Illinois's math library, Altgeld Hall, where I'd wander the upper tiers of the stacks. The lights ran on timers that would occasionally shut off, leaving me to find my way via the light coming from lower levels through the glass-block floors. Sometimes I'd find copies of The Journal of Recreational Math and read the shortest article. Or maybe I'd find a collection of Martin Gardner's Scientific American columns and read about whatever he found to present this week. Math, when there isn't a test coming, can be fun. It can be full of toys to play with in your mind.

I myself have published one column in Scientific American (on the curse of dimensionality). In my pitch to the editor, I mentioned recreational math—not sure if I referenced Martin Gardner by name—and the editor was a little annoyed and focused on all the other parts of the pitch. But even though Mr Gardner took the term *recreational math* to the grave with him, recreational math books still exist, under various guises. They always hit a limit, though, which this book hopes to blow past: they talk about math, but halt early in doing math. I want a book that tells me interesting things, that gives me surprising toys to play with, but which trusts me to be capable of confronting the real thing. So I did what everybody else who wrote a math book did: I wrote the book I wanted to read.

Math textbooks assume you are in training to do more math, but we're here to have a good time. If you're learning to play the violin, you're spending a lot of time practicing simple scales. If you're listening for the purpose of music appreciation, you're waiting for the warm-up scales to be over so you can get to the surprising and engaging bits. That is, the pacing and focus of this book will be very different from the pacing and focus of the textbooks by authors who hope you are reading on the way to becoming a full-time mathematician.

On People

An expectation-setting aside: I think of recreational math as about math, not the colorful characters who write the math. There are bookshelves of people-heavy books—I'm in the middle of Ellenberg [2022] right now and am learning fun things about a lot of people (and a decent amount of math).

I am not going to talk about people here. First, I want to focus on mathematical objects. Second, ever since writing an article about James Sakoda, who developed an influential computational model and method in 1949, and Thomas Schelling, who re-developed the method in 1971 and claimed full credit for it, I've been sour on allocating credit to a single person who "discovered" a given mathematical object [Landau and Klemens, 2023]. In knot theory, there is the Homfly polynomial, which is named for the six(!) authors who simultaneously and independently developed it (Hoste, Ocneanu, Millet, Freyd, Lickorish, Yetter), which sounds like a victory for giving full credit, but the acronym *still* left off two people who had published on the other side of the Iron Curtain (Josef Przytycki and Pawel Traczyk), and numerous Russians who had developed it but didn't think it important enough to publish [Sossinsky, 2002, p 71]. I respect that Hausdorff did much to develop T_2-separability (§7.1.4), but if he hadn't, it is deeply implausible that nobody else would have thought of it. I will avoid naming math objects after people, but will put the names of famous mathematicians in parens, to help you know what other people are talking about when they say a space is Hausdorff.

1.2 Objects and their texture

There's a thread of mathematics that breaks concepts down into fine detail, to determine precisely what you need and don't need for any given property to hold. You don't have to memorize this incidental example, but a *magma* is a combination of a set S and a function f which takes an element of S and returns an element of S. A *monoid* is a magma with an identity element named id for which $f(id, x) = x$, for all x (also, f is associative). Think positive integers and the + operation, with the identity element of zero, as $0 + x = x$ for all x. A *Group* is a monoid where every element has an inverse: for all x there exists an x_{inv} such that $f(x, x_{inv}) = id$. In our example, add in the negative numbers, where $3 + (-3) = 0$.

You can do a lot with monoids which are not a group—just think of how much of our world runs on positive integers and addition—and there's something to be said about thinking hard about exactly what properties are needed when. But I want something completely different. How do spaces which are monoids but not groups feel different—not just like groups that don't do as much, but their own thing that will never look like arithmetic?

Topology gave me oddly behaving systems in ways I had never imagined. Chapter 7 presents a sequence of nested classifications, like how a magma is a type of monoid is a type of group, but with a much more dull naming scheme. The topology which satisfies the $T_{2\frac{1}{2}}$ property but not T_3 is an infinite bloom of mushrooms from the number line, where their stem is a single point and their head is a half-sphere that comes infinitely close to but does not touch the number line itself (see §7.1.9). My friend, that is the sort of weirdness I am looking for from mathematics. And the hierarchy itself, though badly named, was unexpected to me and raised questions that didn't have obvious answers. Which is more common, topologies where every point is a closed set, or topologies where every pair of disjoint closed sets are wrapped in a corresponding pair of disjoint open sets? Maybe that question is incoherent to you now, but by the end of Chapter 7 you'll have a solid answer.

John Conway, who cemented his position in the pantheon of recreational mathematicians by his development of the Game of Life (and surreal numbers like omega; see box in §7.3.3), took a firm stance that mathematical objects are as real as anything:

> There's no doubt that they do exist but you can't poke and prod them except by thinking about them. It's quite astonishing and I still don't understand it, despite having been a mathematician all my life. How can things be there without actually being there? There's no doubt that 2 is there or 3 is there or the square root of omega. They're very real things. I still don't know the sense in which mathematical objects exist, but they do. Of course, it's hard to say in what sense a cat is out there, too, but we know it is, very definitely. Cats have a stubborn reality but maybe numbers are stubborner still. You can't push a cat in a direction it doesn't want to go. You can't push a number, either. [Roberts, 2021, p 2]

Much of this book is about the process of poking and prodding objects too strange to exist in this world. Some examples are via a classic in the field named *Counterexamples in Topology* [Steen and Seebach, 1978], and its modernization, `topology.pi-base.org` [pi Base Community]. If the thought of a book full of objects that break assumptions sounds interesting, you're thinking like a topologist.

Once you're done inspecting the object for its own sake, you can build things with it. Stand-up comics use the problem of building furniture as a reliable go-to when they need something people have trouble with or dislike doing. First, that's not necessarily accurate, as more than enough people actually enjoy the process of taking a set of minimal parts like flat boards and nails, and combining them to form something complex and beautiful, like a bookcase. There's something a little miraculous in it. Kids build things with building blocks for the inherent joy of building, and it is only the most mean-spirited who would ask why a child wasted their time building something fanciful but impractical.

Second, mathematical constructions lack much of the tædium of physical construction. As a klutz, hammering a nail straight has never been my forté, and though I do like the process of watching a pile of boards become a bookcase, the process is stressful to me, as I know that I might derail everything by bending a nail at any moment. Of course, mathematical constructions have no such issues. I know many find the process of hammering a row of twenty nails soothing, but I get bored, and appreciate that mathematical constructions show you how one step is done, then tell you to assume that step is repeated an infinite number of times.

You'll see constructions in this book typographically distinguished from other chains of logic by being a sequence of numbered steps. The constructions are typically longer than non-constructive chains of logic, but I hope you find them to be correspondingly more interesting, as a metaphorical bookcase forms before your eyes from parts you might not expect add up to a bookcase.

1.3 Albert Einstein thought he was bad at Topology.

This is a math book for mathematically curious people who don't usually read proper math books. Expect segments at the level of chatting about math like this intro, but also look forward to segments in which we take up mathematical objects and follow through the implications of their definitions in step-by-step detail. If you lead a life steeped in mathematical notation, I expect you should be able to read this book (and may enjoy it more) if you are lightly inebriated.

Many people with a healthy curiosity don't read math because they're afraid of it. Not infrequently, I'll tell somebody that I do computational social science, and they will all but interrupt me to inform me that they are bad at math, frequently adding that they grew to fear math after a teacher somewhere along the line was deeply confusing or condescending. Being bad at math somehow became a part of their identity.

There are enough systems in the world designed to make us feel bad about ourselves. Math shouldn't be one of them.

The intent of this book is to give you a tour of a field I found to be surprisingly interesting, not to test your intelligence. I had the privilege last year of making a trip twelve time zones from where I am today, to Beijing. I can read only the most basic of Mandarin characters and speak only a few words, but I had a great time and saw many wonderful things, and at no point did I look at a wall of Chinese text and think *I must be stupid because I am struggling to read this.* This is what people do with math all the time, though: *it's not that I have no experience with this, or that the implicit underlying worldview is new to me, or that new things sometimes take time—it must be that I'm stupid.* If you take visiting a new country or learning new things to be an IQ test, you will be miserable the whole way through; if you take them as an opportunity to explore and be surprised even when things only partly make sense, moments of joy have a chance to manifest.

I'm not your therapist and can't exorcise your middle-school demons, but if the sight of mathematical symbols causes you to freeze, I can at least cheer you on and encourage you to slow down but keep reading, to lean into the passages presenting a symbol-laden logical chain instead of succumbing to the urge to skip ahead to parts that feel safer. Take it one bullet point at a time. This is a book about thinking through weird inventions, not juggling numbers, of which there will be fewer than you might be expecting (the number eleven only appears as a page number and in this parenthetical). We won't take the log of anything until the final pages of the book, and you'll be able to gloss over the two or three steps that use them if you don't remember how logs work. As language vocab lists go, the list of symbols in the mathematical language of this book is not much, and as you'll see from the list below, all of it can be read as plain English.

Mathematical writing is often characterized as an inherently alienating foreign language. The author of the Tony- and Pulitzer-winning play *Proof*, David Auburn, went to the same college I did, so we know he had to learn how to do an ϵ-δ proof. At key points in the play, he parodies the staccato rhythm of how mathematicians write, and/or uses it to express non-mathematical truths: "Let X equal the quantity of all quantities of X. Let X equal the cold. . . . The number of books approaches infinity as the number of months of cold approaches four. I will never be as cold now as I will in the future. The future of cold is infinite. . . . " [Auburn, 2001]

Such parody is common; the cynical go even further and say that that mathematical notation is designed to intimidate and exclude. But I wholeheartedly reject those accusations. The standard notation of today evolved after an unbelievable amount of thought as the means of writing which most clarifies difficult work. Check your local math history archives for an edition of Euclid or any other pre-algebraic author who wrote math using only words in complete sentences. It's maddening, and leaves one longing for xs and ys.

On the line-by-line level, you will hit walls and get lost—if you don't, it's time to find a more challenging Topology text where you will. Programmers have a joking-but-serious technique called *rubber duck debugging*, in which you explain a bug you're stuck on to an inanimate object, and in slowly breaking down the steps during the one-sided conversation, you realize the problem and its solution. When stuck on a line of math, you can readily do the same. Imagine explaining what you're not getting to a loved one, or your actual therapist, and what questions they would ask. Honestly, those lines where you get stuck and then untie the knot are the best parts, because that's where you see new things.

On the conceptual level, people with math anxiety often set the bar for fully understanding a concept as just beyond their current level of understanding. I invite you to go back to the first paragraph of this book, where we proved that there are as many even integers as there are integers. If you're a normal human, you feel uneasy about it and will never understand it the way you understand how a pair of scissors works. That's great—the use of formal methods to push ourselves past day-to-day intuition is the entire point, why we're here. Please enjoy it.

1.4 What do mathematicians do?

The core grind of the working mathematician is DTP: definition, theorem, proof, the same format Euclid and many others around the world used a few thousand years ago. Here's a sample, using a proof by contradiction:

Definition 1.1. Canonical open set
The canonical open set is the set of all real numbers between zero and one, excluding zero and one themselves. Usually written $(0, 1)$.

Theorem 1.1. No maximal element
The canonical open set has no maximal element.

Proof: Assume the canonical open set has a maximal element, x. There is, between x and 1, another value x_2 defined as the midpoint between x and one: $x_2 = (x + 1)/2$. Because x_2 is less than one but greater than x, it is in the canonical open set, but $x_2 > x$, contradicting our assumption that x was the maximal element in that set. If, for any claimed maximal element, we can construct a larger element, then there can be no maximal element. ■

Let's break down the three steps.

Definition: I've grown to appreciate that the definition step can sometimes be a more important—and interesting—step than the theorem part.

We could philosophy-of-science this and ask whether the canonical open set was invented or discovered. Like the mineral bauxite, it was always lying deep under the Earth, waiting to be separated from close friends like the set of numbers between zero and one but including both zero and one, for which the maximal element is simply one. It could be a simple invention, but if it is, then to steal a line from Eugene Wigner, it's an unreasonably effective invention, which does more than the close cousins to which we pay no heed. Bauxite, if you don't know it, is a rock from which aluminum is extracted. Over the centuries, we collectively worked out that you can take this ugly rock and pull one segment from it and use only that segment to make soda cans. Aluminum, an element, was discovered, but it had to be refined by human effort to be recognizably useful to other humans.

I used to focus on the theorems, but understand more now that it's all a unit. One of my professors in grad school, Jeff Banks, pointed out that when proving a theorem, you set the rules, so you might as well make it easy for yourself. Find the definitions that make the theorem work.

Theorem: Theorem 1.1 has more-or-less the same form as every other theorem: *now that we took the time to define an object, that object has certain useful properties, or a certain relationship to another carefully defined object.* Interestingness comes from this interplay between invented objects, and the properties about them we can discover. Math is built from an æsthetics of parsimony, that if you can get a lot of results about a simple concept, that's a win. That's also a win for interesting. The definition of *open set* we'll see below is obnoxiously simple: it's a list you defined yourself, plus the intersections and unions of those items. You made up that list. But like a magician asking you to pick a card, any list you invent that meets the rules will correctly behave the way the canonical open set behaves in appropriate situations. Like, wow.

Kids' books on math are filled with these sorts of tricks. Try this: pick any number between one and fifty. Got one? OK, then:

- add 25,
- divide by five then round to the nearest multiple of ten,
- divide by ten,
- multiply by six,
- add thirty,
- divide by three,
- round up to the next multiple of ten,
- then divide by four.

You will get five.

It feels like magic, though we know from the context of this being a math book that it's simply a question of an equation which, when written down, is baroque yet easily reduced to five. Nonfiction authors will sometimes set up your expectations wrong: it's more interesting to write "You may have thought that recreational math died with Martin Gardner, but in fact, many authors are still writing it under a different guise" than "Recreational math is still written, but not by that name." The minimalist school of writing far prefers the latter, but the former creates interesting. The weird expression I'd walked you through is equivalent to the expression "5", but one is succinct and one creates interesting.

In the mathematical context, the most interesting events are when the left and right sides of the relations are from different genres of math. Infinite sequences are hard to work with, but can we say anything about them by equating them with groups of open sets? Yes, we can; see §5.2.1.

Proof: William Thurston, a working mathematician, stresses that proofs are a social construct. "When I started as a graduate student at Berkeley, I had trouble imagining how I could 'prove' a new and interesting mathematical theorem. I didn't really understand what a 'proof' was. . . . Mathematical knowledge and understanding were embedded in the minds and in the social fabric of the community of people thinking about a particular topic. This knowledge was supported by written documents, but the written documents were not really primary. . . . " [Thurston, 1994]

Thurston is stressing a point that it's human understanding that matters. There are proofbots that will invent objects beyond human comprehension and prove all their properties; nobody cares. "We should recognize that the humanly understandable and humanly checkable proofs that we actually do are what is most important to us, and that they are quite different from [computer-programmable] formal proofs."

The format of a proof is going to remain a sequence of steps, each of which is a logical conclusion from the prior, in this book just like you did it in high school geometry. But don't forget in all the mechanics that the point is not to be convinced that the mechanism works, but to understand how and why.

1.5 Logistics

OK, here's the Topology book I always wanted, which will tell you about all the components you need to build Topology yourself, but at a pacing to linger on the interesting stuff.

The mathematical prerequisites for this book are not large, especially given that none of the topics in this book are computationally intensive. You already made it through a definition-theorem-proof sequence, and if you followed it, you are probably at the level I'm expecting.

My apologies to any topological specialists whose specialization I left out of the book. This book covers point-set topology, and is a subset of what you might cover in an undergraduate intro to Topology class.

A brief reality note

If you haven't done so, please take two minutes to head over to `bandcamp.com` and buy a copy of this book. I spoke with the owner of a used book store the other day who said that the price of a book should be the price of a good meal. I don't know what dinner costs where you are, so I've set this book to be name-your-price, with a suggested payment of $15.

I like Bandcamp for being one of the most creator-supportive platforms around, and making it easy for me to present a fully-accessible book with both text and audio. Bandcamp also has an interesting feature that the list of supporters is visible on the page. A core piece of a book pitch to a publisher or agent is the "comps" section, a list of comparable books and some evidence that they have sold well. It would be great if the next person who wants to write a reader-friendly book about math can point to the Bandcamp page for this book and say *look at the hundreds of people who supported that book.* Please take two minutes to cast your vote for writing like this.

If you'd like to tell your friends about the book, please do. The set $G \equiv$ {people good at math} $\cap$ {people good at self-promotion} is small, and, defining $I \equiv$myself, we find $I \notin G$. It's up to all of us to bring about social media filled with interesting math instead of the social-media default state of an endless stream of bad vibes. Thanks.

1.5.1 The notation

Although a vocab list is not the most exciting start, here it is, to help you read math like the pros. It's long-ish because I wanted to go beyond a terse list of definitions and include some discussion of the math you're about to encounter. Think of this as the flyover glimpsing pieces of the world you are about to delve into. If you are already familiar with all this, it will just take a minute to clarify the conventions used in this book, like what I mean by $\widetilde{S}$.

- p: A point. Generally the zero-dimensional dot on a number line or plane, but you'll see that the spaces points live in are often just a convenience to easily express how an infinite number of sets of points fit together. Nothing breaks if *a point* is a number in no underlying space, or a non-numeric but indivisible item like a member of your family tree.
- $\{\dots\}$: curly braces indicate a set of things; e.g., $\{1,2,3\}$ is the set consisting of one, two, and three. Mathematical philosophers have come to the consensus that everybody knows what a set is, and it doesn't need to be explicated further than your intuition.

  - The point p and the one-element set $\{p\}$ are closely related but conceptually distinct.
  - We will deal heavily with sets of sets, like how $\{\{p_1\}, \{p_1, p_2, p_3\}\}$ is a two-element set consisting of a one-element set and a three-element set.

- $\emptyset$: the empty set, $\{\}$. Sometimes read as *null*.
- (a, b): as a set, all points in the real number line between a and b, excluding a and b themselves. These will be referred to as u-open (the usual open) sets throughout, to contrast with all the other strange topologies we'll encounter. As per Theorem 1.1, the edges of such open sets verge on embodying Zeno's paradox, as $(0, 1)$ contains 0.9, 0.99, 0.999, . . . , but not 1 itself.
- $[a, b]$: The set of all points between a and b, including a and b themselves. Referred to as u-closed when needed. We will sometimes have half-open intervals, like $[0, 1)$, including zero but excluding one.
- $A|B$: read as *A given B* or *A such that B*, which is incoherent by itself but fits into various contexts. Statisticians might write $P(A|B)$ to express the probability that A occurs given that B has occurred.
- $\{x|x < 3\}$: Referred to as *set builder notation*, this is an amalgamation of notations to this point. This expression reads as *the set of xs such that* $x < 3$, which we might otherwise write as $(-\infty, 3)$. Before everybody had a keyboard with a pipe (|) key, this was (and often still is) written as $\{x : x < 3\}$.
- $\mathbb{Q}$: Rational numbers. All the numbers expressible as an integer divided by another nonzero integer. The noun *ratio* is embedded in the adjective *rational*.
- $\mathbb{R}$: Real numbers. All the rational numbers plus all the irrational numbers.
- $\mathbb{R}^+$: Positive reals. Zero is not positive.
- $\mathbb{R}^2$: two-dimensional reals, points on the (x, y) plane. Yes, parens are also used to indicate two-dimensional points as well as one-dimensional intervals as above; you'll work it out from context. Higher-level mathematicians don't like assuming the horizontal axis is named x and the vertical y. But when I was in high school, any time math would come up, my mother would dismiss it by saying "All I know is there's an x and a y," and then laugh. If even my math-phobic mother has internalized this notation, we might as well use it.
- $f : X \to Y$: When introducing a function, it is polite to mention

the spaces in which its inputs and outputs live. Here the inputs are in the X space and outputs in the Y space. In your algebra class, functions were all of the form $f : \mathbb{R} \to \mathbb{R}$ or $f : \mathbb{R}^2 \to \mathbb{R}$, so it wasn't worth belaboring, but here things will be much more general. We won't see functions until Chapter 4, though.

- $A \cup B$: The union of A and B, the set of points found in either one or the other (or both). Notice that it is a stylized *U*.
- $A \cap B$: The intersection, the set of points in both A and B. You can remember it as the one that isn't a stylized *U*.
- $\bigcup_{x=1}^{\infty}(x, x+1)$: the big $\cup$ with a subscript and a superscript is a compact means of writing the union of a sequence. The format for each set in the sequence is given, in this example $(x, x+1)$, and we're specifying that x begins at 1 and counts to infinity, giving us the sequence $(1, 2) \cup (2, 3) \cup (3, 4), \ldots$, adding up to the real number line past one, excluding the integers. Similarly for the intersection of a sequence, $\bigcap_{x=1}^{\infty}(-\frac{1}{x}, \frac{1}{x})$ (which is $\{0\}$).
- $\Sigma_{i=2}^{\infty}\frac{1}{2^i}$: an infinite sum, with the same indexing custom of counting i from the lower bound to the upper. In this case $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$, which approaches one as $i \to \infty$ (as i goes to infinity). Read Σ as *sigma*.
- $A \subset B$: A *is a subset of* B. The above two items, $\cup$ and $\cap$, took nouns and created compound nouns; the same stylized *U* sideways is an assertion. It looks a little like $A < B$, which is what we write when the smaller number A fits into the larger number B.
- $A \subseteq B$: Whether $A \subset A$ is a point of nomenclature debate. If an author wants to be clear, though, they'll say that A is a *strict* or *proper subset* of B if they want $A = B$ to mean $A \not\subset B$, and will use the version that sort-of incorporates an equals sign, $A \subseteq B$, if they want to allow the set to be a subset of itself. Looks somewhat like $A \leq B$.
- $A - x$: The minus sign does what you expect with sets and points. This means the set A excluding x (typically a point).
- $\widetilde{A}$: The complement of A. If the entire space is named X, then $\widetilde{A}$ is $X - A$. This will be heavily used because of Definition 3.1.
  - E.g., if $X = \{1, 2, 3, 4, 5\}$ and $A = \{1, 3, 5\}$, then $\widetilde{A} = \{2, 4\}$.
  - Sometimes written as $X \backslash A$, though in most cases it's easier to not explicitly write the overall space.
  - In logician circles, $\sim$ is sometimes read as *not*, as if our

sets were Venn diagrams and the area outside the circle A represents the group of things that lack property A.

- ϵ: The Greek letter *epsilon*. The common English equivalent is "an iota", indicating a very small amount. Not a variable, but a fixed amount in any given context; maybe think of it as 10^{-100}.
- *iff*: an English word you will find in your favorite dictionary, defined as *if and only if*. For example, $a < b$ iff $b - a > 0$.
- $\equiv$: *is defined as*. Though $\equiv$ implies to the reader that a new association is being declared and *iff* is usually reserved for inferences from other statements, the rules of logic would allow us to replace instances of $\equiv$ with *iff*. If something is some kind of object iff it has some clever property, mathematicians sometimes find benefit in using that property to define the object. For example, the condition I gave in the example for *iff* could be used as a definition: $(a < b) \equiv (b - a > 0)$. Or, observing that integers are in one-to-one correspondence with even integers, Rudin [1976] defines: infinite set $\equiv$ a set which is in one-to-one correspondence with a proper subset of itself.
- $A \Rightarrow B$: A *implies* B, or *if* A, *then* B. For example, ($A \Rightarrow B$ and $B \Rightarrow A$) $\Rightarrow A \equiv B$.
- $\{\forall, \exists, \in, \ni\}$: {*For all*, *There exists*, *In*, *Such that*}. I'm putting these all here on one line because they're often used together. E.g., we could explain that $(0, 1)$ has no maximum value by stating that $\forall x \in (0, 1), \exists y \in (0, 1) \ni x < y$, i.e., for all x in $(0, 1)$, there exists a y in $(0, 1)$ such that $x < y$. Whether such expressions are written out in English or compressed with symbols depends on the author's mood.
  - *Such that* is sometimes abbreviated to *s.t.*.
  - If you have a point p contained in a set S, use $p \in S$. If you have a set S_{sub} which is a subset of S, use $S_{sub} \subset S$.

1.5.2 Thanks

Thanks to Dave Bridgeland, Peter Gedeck, John Kaufhold, Elizabeth Landau, Aaron Schumacher, and Josh Tokle for comments and suggestions.

Chapter 2

Topologies

To define Topology, the field of study, let us define *a topology*, the structure:

Definition 2.1. Topology
A set of points X with a list of open sets of those points.

1. The full set X is on the list, as is $\emptyset$.

2. Any intersection of a finite number of open sets is also an open set.

3. Any union of open sets is also an open set.

If referring to a generic space, the whole of it will usually be referred to as X, and the list of open sets τ_x (read as *tau sub x*); then a topology is a combination (X, τ_x). Sometimes the space is taken as given and *the topology* will refer to just τ_x; you'll pick it up from the context.

Let's pause to admire how stupid simple this definition is. You make up any list of groupings—whatever you want—and we'll complete it by adding finite intersections and arbitrary unions of the elements of your list. As a technicality, we add *everything* and *nothing* to your list. This is the magician's *pick a card, any card* setup for the tricks our topologists will pull for the remainder of the text.

You have some idea of open sets from the math you've already learned, where the usual open sets are the intervals between a and b (with a and b real numbers and $a < b$), formally the set $\{x|a < x < b\}$, or more succinctly written as (a, b). The usual closed interval $\{x|a \leq$

$x \leq b\}$ is written as $[a, b]$. Following Mansfield [1963], I'm going to refer to (a, b) as u-open, the usual open sets; the set of all of those u-open sets form *the usual topology*. But the definition of *open set* implicit in the definition of topology is much, much more broad and accommodates a great deal of unusual. I'll stress this by stating a direct definition of *open set*:

Definition 2.2. Open set
Whatever set of points we want. The intersection of any finite number of open sets is also an open set, and the possibly infinite union of any open sets is an open set, and $\emptyset$ and the entire space are open sets.

A main goal is to develop definitions around the generalization of open sets which match your intuition about u-open sets. The next several chapters will basically follow that thread, asking questions of what certain expected objects will be and behave like under this definition. Infinite sequences have limits, in the case of Zeno's paradox from §1 the sequence $\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots \to 1$, and this will tie in with our open sets. We have some intuition that b, sitting just at the edge of (a, b) is a sort of limit point to the set, and often refer to $[a, b]$ as the closure of (a, b). If you give me any list of sets and declare that to be the open sets, can I find concepts of limit and closure that match (Chapter 3)?

From there we'll be able to hit what was to some extent the goal of a large portion of pre-1900s mathematics: defining continuity (Chapter 4). The real number line is just a list of individual points, yet we have some intuition of a universe where there is a smooth line between the mark at $\frac{3}{4}$ meters and the mark at $\frac{7}{8}$ meters. By the definition of a topology, you can see that the goal is to define continuity not in terms of real numbers, but in terms of open sets. Mansfield [1963, p 11] characterizes the game of topology as the problem of describing how things can stretch and bend without tearing. Continuity "is the key to the kingdom. The notions of 'stretching and bending' can be mathematically expressed in terms of functions. The notion of 'without tearing' can be expressed in terms of the continuity of these functions."

Or we can recast the question of whether the real numbers are a bag of points or a smooth line by asking whether it is "connected". Chapter 5 will present three different ways of defining that term.

The last few chapters cover an organization of topologies into an eight-level hierarchy (Chapter 7), another way of understanding infinite sequences (Chapter 8), a construction ("the hedgehog") which lets us impose a concept of distance on a topology (Chapter

9), and an application of the same tool we used to check whether infinite sequences converge to measure things in fractional dimensions (Chapter 10).

That's the book, and this entire exposition is chasing down the implications and applications of the obnoxiously simple whatever-you-want Definition 2.2. You could maybe argue that *mapping* and *preimage*, introduced in Chapter 4 are new and distinct concepts, but the rest of what is to come—closed sets, limit points, density, continuity, three concepts of connectivity, nine interlocking classes of separation properties, coverings and compactness and their implications for convergence, the torus and M-band and K-bottle—are specific configurations or variants of open sets as per Definition 2.2. It's all in there waiting to be unfurled.

But first, let's write down some examples of topologies as per Definition 2.1 and see where they take us.

2.1 Some basic topologies

My bookshelves are full, and my junk drawer overflows, but I still go to stores and browse. Apologies to those of you who work retail, but I rarely buy anything, just kill time picking things up, inspecting them, and putting them back down. Camp stores are my favorite, where people have come up with so many clever ways to make things that fold into themselves, familiar things like forks and stools and jackets in shapes I'd never considered before. This chapter is like that, but you don't have to make shelf space for any of the objects. They cram lines of infinite length into a one-by-one box, or shave off 100% of the weight of the $[0, 1]$ interval and yet still map one-to-one to that same interval, or break our expectations entirely by defining closed sets as open sets. Nonetheless, every definition of some characteristic of a topology in later chapters (like limits or closed sets) must apply to even the weirdest topologies on this list.

As above, the rules are simple: we specify any list of sets we want, but then we have to add all finite intersections and possibly infinite unions, repeating until we have a complete, self-contained list. Though I'll sometimes omit this detail in the definitions to follow, don't forget that the entire space and $\emptyset$ must be open sets in all topologies.

As promised in the introduction, these are very different from each other and will get weird. But you live with multiple topologies now. You'll see the Manhattan metric below (§2.2); it's also called the *taxicab metric*, but most Manhattanites take the subway. What is the set of points five minutes from the World Trade Center? It includes 7th Ave & 14th and 7th & 34th Streets (take the 2 or 3 line), but not 7th & 18th or 23rd street (you'd have to transfer to the 1, and good luck doing that within five minutes), some parts of Brooklyn across the East River but no part of the East River itself, a portion of New Jersey (take the Path), but very little else on the island of Manhattan itself.

Wherever you live, and whatever the quality of your nearest subway system, you maintain different concepts of how points relate beyond the strictest tape-measure distance. You have friend circles which include some people and exclude others, whom you categorize into other circles. You might be the sort of person who goes straight to building a topology when you meet somebody new at a party: *So who do you know here? How are you linked?* The points of interest in your area and the people at the party don't change, but what sets they lie in and how those sets intersect is malleable and something you can define and redefine.

The rest of this chapter will introduce various topologies, in three batches. The first are primarily a chance to get familiar with all the subtleties in the definition of an open set (Definition 2.2), like why they took pains to allow topologies to include possibly infinite unions but only finite intersections. The second batch are those generated by metrics (aka distances) and clarifies what people mean when they say that distance doesn't matter in Topology. The third batch is when things start getting weird, with topologies where the set of points is something unusual, or where the set of open sets is different enough to have significant implications, ending with my personal favorite, the seemingly simple but somehow perverse Lexicographic Topology. There will be more throughout; I doodled a little symbol, , to mark the 26 different points in this book when new topologies are introduced.

The first several in this section are not going to be especially surprising, beyond a few little twists. The definition of an open set (Definition 2.2) is a straightforward definition with subtle implications. Given that I defined *Topology* as the study of open sets, it's worth considering those subtleties in detail.

2.1.1 A bundle of points

This won't happen often in this book, but there's a certain back-and-forth in developing a topology that lends itself to the style of a Platonic dialogue. Here, Socrates is teaching a talking doughnut how to build a topology.

SOCRATES: If we have a space of four points, $\{p_1, p_2, p_3, p_4\}$, what kind of topologies could we generate?
DONUT: OK, a topology is whatever I want, so the set $\{p_1\}$ is my topology. Have a nice day.
SOCRATES: Does it conform to all the rules? It looks like some things may be missing.
DONUT: You told me I can be the world-building doughnut, and an open set is whatever I want it to be. I want $\{p_1\}$ to be the only open set.
SOCRATES: But there are other conditions you have to meet. Are there missing elements from your topology?
DONUT: I guess you want the entire space—
SOCRATES: We usually notate that as X.
DONUT: —OK, X and the empty set to be in the list of open sets? So, $\{\{p_1\}, X, \emptyset\}$, is that a good topology? I think it's good enough.
SOCRATES: The definition also requires unions and intersections of all the sets in the topology.
DONUT: I only have three sets, and $\emptyset$ is barely even a set.
SOCRATES: So it'll be easy for you to tell me about its unions and intersections.
DONUT: The intersection of $\emptyset$ and anything is $\emptyset$, so that's already on my list. And if you give me *any* set S, the union of $\emptyset$ and S is just S. So intersections and unions with $\emptyset$ don't add anything.
SOCRATES: Good, now tell me about X
DONUT: The union of X and any part of X is just X, and the intersection of X and S is just S. It doesn't do anything either. This feels pedantic.
SOCRATES: I'm glad you took the time to check unions and intersections with X and $\emptyset$ now, because what you're saying applies to any topology. We've checked them once and for all.
DONUT: I'm relieved. And that means I *am* done: if I define my topology at $\tau_1 \equiv \{\{p_1\}, X, \emptyset\}$, then I have whatever sets I want, plus the mandatory *everything* and *nothing* sets, plus the intersections and unions.
SOCRATES: Good job.
DONUT: OK, that's my topology.

SOCRATES: Formally, a topology consists of a space paired with a list of open sets, so let's define the topology as (X, τ_1).

DONUT: Let's go bake cookies.

SOCRATES: You built one topology, but you know there are others. And... you eat cookies?

DONUT: You take balls of cookie dough, and you can watch the little spheres flatten down to cookies. It's fun.

Note that $\{p_1, p_2\}$, for example, is not on our talking doughnut's list, so that set is not in the topology. Yes, $\{p_1, p_2\} \subset X$, and X is always on the list of open sets, but sets can't slide into our list of open sets by being a subset of other open sets. The invitation to a topology is non-transferable.

SOCRATES: Let's add one item, so now $\tau_2 \equiv \{X, \emptyset, \{p_1\}, \{p_2\}\}$. How does that look as a topology?

DONUT: Fine, I guess. You said I already talked about how intersections and unions with X and $\emptyset$ don't add any sets that aren't already on my list and I never have to check it again.

SOCRATES: Thank you for the reminder. But what about intersections and unions between $\{p_1\}$ and $\{p_2\}$?

DONUT: They don't intersect: $\{p_1\} \cap \{p_2\} = \emptyset$.

SOCRATES: Good, so we don't have to think about the intersection. The union?

DONUT: The union is another set, $\{p_1, p_2\}$, so OK, $\tau_2 \equiv \{X, \emptyset, \{p_1\}, \{p_2\}, \{p_1, p_2\}\}$.

SOCRATES: Are we done?

DONUT: I think we're done. The intersection of the new set $\{p_1, p_2\}$ with $\{p_1\}$ is just $\{p_1\}$, which we have, and the union of the new set with $\{p_1\}$ is the new set. Same story with $\{p_2\}$.

SOCRATES: Nice, now you have two topologies on this simple four-point space. What do you think of $\tau_3 \equiv \{X, \emptyset, \{p_1, p_2, p_4\}\}$?

DONUT: I can't eat it, but if there's only one set that isn't X or $\emptyset$—

SOCRATES: We sometimes call X and $\emptyset$ *trivial open sets*—

DONUT: That sounds apropos. They feel like they're just there as some sort of trivia. If there's only one nontrivial set, then there aren't intersections or unions to worry about, and you're done.

SOCRATES: You've built another world.
DONUT: To the extent that one nontrivial set is a world.
SOCRATES: We're just getting started, my doughnut friend. And if one set isn't enough to allow interesting unions and intersections, what do you think of $\tau_k \equiv \{X, \emptyset, \{p_1\}, \{p_2\}, \{p_3\}, \{p_4\}\}$,
DONUT: The union of any two of those is a pair, like $\{p_1, p_2\}$, $\{p_1, p_3\}$, $\{p_1, p_4\}$, $\{p_2, p_3\}$, all those have to be in the topology. How many of those are there?
SOCRATES: Six: three with p_1 first, two with p_2 first, one with p_3 first.
DONUT: Wait, does order matter with these sets?
SOCRATES: Not at all. Every set is just an unordered bundle, but it's nice to keep things ordered so we can keep track. Otherwise you might count $\{p_1, p_2\}$ and $\{p_2, p_1\}$ as separate sets, even though they're identical.
DONUT: OK, makes sense. Then there are four triplets, $\{p_2, p_3, p_4\}$, $\{p_1, p_3, p_4\}, \ldots$.
SOCRATES: You can keep track of those because the first is missing p_1, the second is missing p_2, and so on.
DONUT: Yeah, that's what I was doing.
SOCRATES: Is there anything else we need?
DONUT: That seems to be everything: all the singletons, pairs, triplets. The four-element set is already in my topology, it's called X.

We started with what the mathematicians might call a kernel, the six-set list τ_k, but from that kernel sprouted all the pairs and all the triplets, and the full topology is all of those (16 sets total). We also need to check for intersections, like $\{p_1, p_2\} \cap \{p_2, p_3\} = \{p_2\}$, but given that we already have every possible combination, checking intersections won't add more sets.

Definition 2.3. Basis
A set of sets B such that every open set in a topology is a union or intersection of sets in B.

- The process of finding all the possible outcomes from a process is usually referred to as *finding the closure*, a phrase I like because of how close it is in both sound and sense to *achieving a sense of closure*. To give an example from an entirely different context, say we have the operation + and the number 1. From that, you could construct $1+1 = 2$, from which you could construct $2 + 1 = 3$ or $2 + 2 = 4$. Following to the logical conclusion, the closure of the + operator and the kernel 1 is the set of all positive integers.

- Unfortunately, the word *closure* is decidedly taken in the Topological context; see Definition 3.3.
- The typical textbook definition of a basis assumes all intersections are also in the basis, then we achieve closure by adding to the basis all unions of basis elements. My treatment here is a little different because the most natural basis is not necessarily the one where all intersections are included.
- You should also know that some authors also refer to a *point* having a basis. This is an entirely different concept, typically an infinite sequence of nested sets. This won't appear in this book until §9.1, where I'll call it a point-basis.

2.1.2 The Usual

The u-open sets are the basis for a topology over the real numbers. Consider how completing the topology with all requisite unions and intersections would work:

SOCRATES: We start with familiar open intervals like $(1,5)$ and $(4,7)$. How could those form unions or intersections?

DONUT: I already worked out how intersections and unions with X and $\emptyset$ are handled.

SOCRATES: Thanks for the reminder that those need to be included.

DONUT: Well, it matters, because if you have intervals that don't overlap like $(1,4)$ and $(5,7)$, then their intersection is $\emptyset$.

SOCRATES: True. We didn't even need to separately specify that it would be on the list.

DONUT: But your example, it doesn't need special handling either. If two intervals overlap, the overlap is an interval: $(1,5) \cap (4,7) = (4,5)$. Their union is also an interval: $(1,5) \cup (4,7) = (1,7)$. The intervals cover it.

SOCRATES: Do they? What if you have disjoint intervals, and you want their union, like $(1,4) \cup (5,7)$?

DONUT: That's not an interval, that's just two intervals taped together.

SOCRATES: Is it in our topology?

DONUT: I don't want it to be.

SOCRATES: But what do the rules say?

DONUT: OK, yeah, that this two-part union counts as an open set.

SOCRATES: This is also our first space with an infinite number of open sets, so we also have to think about infinite unions.

DONUT: Isn't that just the same thing, where you get either bigger intervals or an infinite number of intervals taped together?

SOCRATES: There are two ways it could be a little more interesting. If we take the union of $(0,1)$, $(0,2)$, $(0,3)$, $\ldots$, written more succinctly as $\bigcup_{i=1}^{\infty}(0,i)$—

DONUT: That's an interval with infinity as an endpoint, like $(0,\infty)$. Is that valid?

SOCRATES: Think of it as a nice way to write an interval with no actual end point at all. And if a sequence of endpoints doesn't diverge, then it converges, like the union $\bigcup_{i=1}^{\infty}(\frac{1}{i},1)$.

DONUT: As $i \to \infty$, $\frac{1}{i} \to 0$, yeah, so isn't the infinite union just $(0,1)$? That's already on my list.

SOCRATES: Sounds like we're done then.

DONUT: I have so many topologies!

The easiest two-dimensional analogue to the $(0,1)$ interval is the set of points at a distance less than one away from the origin, meaning everything inside a circle of radius one (but excluding the circle itself). Under the presumption that these are usually small and the custom that small values are written as ϵ, these open sets are frequently referred to as ϵ-balls. We'll have much more to say about them when we get to metrics below, but for now just bear in mind that the Usual Topology can easily be in higher dimensions, where those straight-line intervals become circles (or in 3-D, balls, or in 4-D, hyper-balls).

2.1.3 Subspace topologies

Within the space X, pick any subset Z. The main space's topology τ_x consists of a list of open sets, $\{s_1, s_2, s_3, \ldots\}$. Then $(Z, \{s_1 \cap Z, s_2 \cap Z, s_3 \cap Z, \ldots\})$ is a valid topology. In Figure 2.1, the open sets are the usual circles, but when restricted to a square subspace, the open sets are balls cut off wherever the square fell, leaving odd shapes, a mix of round and sharp edges.

Or you could even restrict the open sets in the 2-D topology to the line $y = 2x$. Then the Usual Topology induces on that line a set of open sets that look much like you'd expect: one-dimensional open intervals, centered around any point on the line, with any size. It's not complicated.

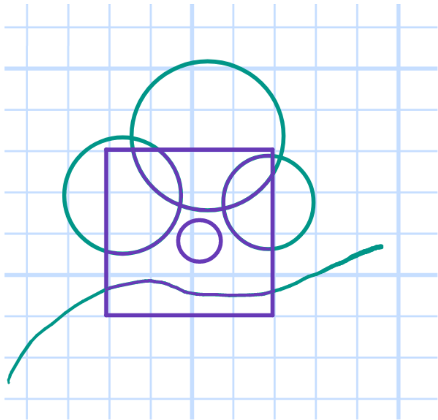


Figure 2.1: If we select a subset of the main space, the topology uses the same open sets as the parent space, but cut off within the subspace.

But bear in mind that we define open within the context of the subspace. If our host topology is the Usual Topology and our subset is $Z \equiv [0,1]$, then within the context of that subspace, an interval like $[0,0.3)$ is open, because it is formed by $(-0.3,0.3) \cap Z$. If nothing else, the entire space is always part of the topology's list of open sets, so $[0,1]$ itself must be open in Z's topology, though it's obviously not open in the parent topology.

2.1.4 Closed intervals et al.

DONUT: You said I can make anything my open intervals?
SOCRATES: Yes, whatever you want.
DONUT: OK, the closed intervals, like $[0,1]$. I want those to be open intervals. [*Donut stares defiantly.*]
SOCRATES: You want me to be uncomfortable with this, but this is perfectly fine. But we have to think through the unions and intersections.
DONUT: It's not gonna be different from the open intervals, like how the union or intersection of overlapping closed intervals are also closed intervals.
SOCRATES: True. What about something like $[a,b]$ and $[b,c]$. What is the intersection of those two?
DONUT: It's a singleton, like $[a,b] \cap [b,c] = \{b\}$. That's just another closed interval, of zero length.
SOCRATES: This is different from the Usual Topology, where singletons weren't in the list of open sets.
DONUT: So what? It's a different topology.

SOCRATES: Let's remember this. It might be important later. Tell me about the infinite union $\bigcup_{x=3}^{\infty}\left[\frac{1}{x}, 1-\frac{1}{x}\right]$, so the union of $\left[\frac{1}{3}, \frac{2}{3}\right]$, $\left[\frac{1}{4}, \frac{3}{4}\right]$, $\left[\frac{1}{5}, \frac{4}{5}\right]$, $\left[\frac{1}{6}, \frac{5}{6}\right]$, ...

DONUT: Looks like that would cover every point between zero and one, eventually.

SOCRATES: Does it cover zero or one themselves?

DONUT: No, those are outside of every open set in the infinite sequence.

SOCRATES: So what does the infinite union look like?

DONUT: Wait, it's just the canonical open set, $(0, 1)$? I have to have open sets in my topology anyway? The whole point was to not.

SOCRATES: We're just chasing down the implications. This is where it got us. . . so far.

2.1.5 The Discrete Topology

Define every point to be its own open set. Adding all unions, all combinations of points are also open sets, meaning the Discrete topology could also be given the title of the complete topology.

- We require infinite unions, but only finite intersections. If we had required infinite intersections, then the Usual Topology would have to include $\bigcap_{x=1}^{\infty}(-\frac{1}{x}, \frac{1}{x}) = (-1, 1) \cap (-\frac{1}{2}, \frac{1}{2}) \cap (-\frac{1}{3}, \frac{1}{3}) \cap \cdots$, which is a single point $\{0\}$. Do the same for every other point up and down the real line, and the Usual Topology would then become the Discrete Topology.

- When completing the closed-intervals topology above, we found that that singleton points are on the list of open sets, meaning that the closed-intervals topology we were trying to construct is actually the Discrete Topology, and every possible combination of points is an open set. You're welcome to be a contrarian and declare the u-closed sets to be open sets in your new topology, but you have to accept the consequences.

More generally, consider the case where every point p_1 is *topologically distinguishable*, meaning that for any point p_2 there is some set which includes p_1 but not p_2; more in §7.1.1.

1. If there are a finite number of points in the topology, then you can take the intersection of one open set to isolate p_1 from p_2, another open set to isolate p_1 from p_3, and so on for every point, and that finite intersection has $\{p_1\}$ as an open set.

2. Repeat for p_2, p_3, ..., and you've constructed the Discrete topology.

Compare to $\mathbb{R}$, where you can isolate a point, but only by taking the intersection of an infinite number of open sets.

That little distinction in the definition of open sets between intersections (finite only) and unions (infinite or finite) turns out to be a key distinction between open and closed sets.

2.1.6 The Indiscrete Topology

The opposite extreme from the Discrete Topology, its list of open sets has only two elements: everything and $\emptyset$. It's a valid topology, and it's sometimes worth bearing in mind the extreme cases, but there is absolutely nothing you can do with it. All the other topologies first state the space they act over and then the open sets in that space, but the indiscrete topology over $\mathbb{R}$, the indiscrete topology over $\{1, 2, 3\}$, the indiscrete topology over the unit circle in the complex plane are all identical from a topological perspective.

2.1.7 Half-closed intervals

The point set is the real line again, just like the Usual Topology, where the basic interval was of the form (a, b) for all a and b where $a < b$. Now, we'll use the same real line, but the base of the topology is the half-closed intervals $[a, b)$. Call this half-closed topology τ_{hc}. Half-closed intervals are nice because they are the smallest change we can make from the u-open intervals we grew up with and still have a topology that doesn't reduce to the Discrete everything-goes topology.

- For closed sets $[a, b]$ and $[b, c]$, their intersection is the single-point set $\{b\}$, and now you've got the Discrete Topology. But we just dodge that with half-open intervals: $[a, b) \cap [b, c) = \emptyset$.

- The infinite union of $\left[\frac{1}{2}, b\right), \left[\frac{1}{4}, b\right), \left[\frac{1}{8}, b\right), \ldots$ has a lower bound which approaches zero, but zero is never in any of the elements of the union. That is, the infinite union adds up to $(0, b)$, and the u-open sets are thus part of τ_{hc}.

Levels of infinity

The first paragraph of this book presented the chestnut that the even integers are in one-to-one correspondence with all the integers.

Rational numbers, numbers produced by dividing one integer by another, are also in one-to-one correspondence with the positive integers:

- Write each positive rational as N/D (numerator over denominator). Bundle them into groups based on $N + D$.
- There is only one rational with $N + D = 2$, which is $\frac{1}{1}$.
- There are two with $N + D = 3$: $\frac{1}{2}$ and $\frac{2}{1}$.
- There are three with $N + D = 4$: $\frac{1}{3}$, $\frac{2}{2}$, and $\frac{3}{1}$.
- There are four with $N + D = 5$: $\frac{1}{4}$, $\frac{2}{3}$, $\frac{3}{2}$, and $\frac{4}{1}$.
- You get the pattern. If there is one item for $N + D = 2$, two for $N + D = 3$, three for $N + D = 4$, . . . , then we can look up the standard formula for sums of *triangular numbers* and find that if $N + D = k$, there are $(k - 2)(k - 1)/2$ items with $N + D < k$.
- Within the $N+D = k$ group, we can order by the numerator as we did above, with $\frac{1}{3}$ first, $\frac{2}{2}$ second, and $\frac{3}{1}$ third.
- Then N/D is the $(N + D - 2)(N + D - 1)/2 + N$-th rational number in this ordering. We did it: we assigned a single integer to every rational number.
- There's the annoyance that many numbers are repeated $\left(\frac{2}{3} = \frac{4}{6}\right)$, which we could fix by a rule that every number not in lowest-denominator form be stricken. It messes up the formula, though it's conceptually simple. After that fix, we have a one-to-one correspondence between rationals and integers.

All these sets which are one-to-one with integers are referred to as *countably infinite* or $\aleph_0$ (pronounced aleph zero or aleph nought).

Above $\aleph_0$ there are still more infinite quantities, such as the real numbers: numbers expressible by a possibly infinite decimal. All the rationals are reals, but there are far, far more real numbers. In contrast to $\aleph_0$, this is often described as uncountably infinite.

Rationals can always be expressed as a decimal which eventually repeats. For example, $\frac{1}{7} = 0.142857142857\ldots$, which we'll write using an overline (the "vinculum") to indicate repetition:

$\frac{1}{7} = 0.\overline{142857}$. This pattern even holds for fractions with nicer decimal versions if we accept trailing zeros, like $\frac{1}{2} = 0.5\overline{000}$.

In our day-to-day we almost entirely deal with rational numbers, with occasional appearances by celebrity irrationals like $\sqrt{2}$ or π. That is a consequence of convenience, our frequent need to divide integer values, and our mortal abilities. If you had a scale of truly infinite precision and measured the weight of an object, would you really expect the readout to end in a nicely repeating pattern?

Irrationals vastly outnumber the rationals: there are an infinite number of ways to modify the zeros in $0.75\overline{000}$, and similarly for every other rational number. If every rational has an infinite number of irrational variants, there's no way to get a one-to-one mapping between rationals and irrationals, so we have two types of infinity, $\aleph_0$ and *the continuum*, $\mathfrak{c}$, where the second is infinitely more infinite than the first.

- Say we met a contrarian who insists that they could write down an enumerated list of real numbers, giving a one-to-one correspondence between reals, each with an infinite number of decimals, and the integers.
- I would look at the first decimal in the first item in our contrarian friend's list, and being more of a contrarian, write down one plus that number. Say the first number on the list is $0.1234321234\ldots$; I would write a 2 on my notepad.
- I would look at the second decimal in the second item in the list and write down one plus that number. Say the second number is $0.1234321245\ldots$; I would write a 3 on my notepad.
- It would take me a while, but I would write down the wrong thing for the N^{th} digit of the N^{th} number all the way down the list, thus generating a new decimal, which starts 0.23 and continues out an infinite number of decimals.
- The thing I wrote down doesn't match the first number on the list, because the first decimal is wrong. It doesn't match the second number, because the second decimal is wrong. And so on to infinity—our contrarian gave us an ostensibly comprehensive list, and we gave back a number that isn't on it.

This is the "diagonalization" argument.

Among finite, quotidian sets, if you have N items, there are 2^N ways of picking a subset of them.

- Write down a 1 if you want to keep the first item, 0 if not. Repeat for the second, third, , N^{th} items.
- More on binary below, but you have just written down a binary number. Every in-and-out combination is another binary number.
- How many binary numbers are there between 0000 and 1111, where we have N zeros or ones? There are 2^N.
- Or, turn a number like 1001 into a binary decimal by putting a dot in front: 0.1001. It expresses the same thinking (keep the first and fourth items, leave out the second and third), but it is now a number between zero and one.
- If you forgot how binary decimals work, we'll review below, but all we need for now is that every real number between zero and one can be expressed as a binary decimal; the binary decimals cover the real numbers in $[0,1]$.
- That is, we've formed another one-to-one correspondence, this time between in-and-out selections of integers and the real numbers $\in [0,1]$.

By analogy to there being 2^4 in-and-out selections of four numbers, we write the real numbers as $2^{\aleph_0} = \mathfrak{c}$, because they correspond to the set of in-and-out selections of all integers. This book won't need them, but we can keep counting up the chain: the count of sets of real numbers is $2^{\mathfrak{c}}$, the count of sets of those sets of real numbers is $2^{2^{\mathfrak{c}}}$, and so on to a (countable) infinity of infinities.

Or here's another one: given a real number like $0.1234567892\ldots$, I can generate two numbers by reading all digits in odd positions (first, third, ...) as a number, then reading all digits in even positions (second, fourth, ...) as another number. From the above starting number, the first would be all the odds, 0.13579 and the second all the evens 0.24682. That pair of numbers is a point on the real plane, $\mathbb{R}^2$, and the mapping is entirely reversible, like how $(0.1234\ldots, 0.9876\ldots)$ would interleave to form $0.19283746\ldots$. So, we have a one-to-one correspondence between points in $\mathbb{R}$ and pairs of points in $\mathbb{R}^2$. Free your mind of expectations that correspondences have to all have the same dimension.

As a final example, how many ways are there to draw a curve from the origin point in $\mathbb{R}^2$ of $(0,0)$ to the point $(0,1)$, i.e., how many continuous functions $f(x)$ are there for $x \in [0,1]$?

- We've decided that $f(0) = 0$, but $f(\frac{1}{2})$ could be any real number; call it x_1.
- One could work out a smooth line from $f(0) = 0$ to $f(\frac{1}{2}) = x_1$ to $f(1) = 0$.
- Keep subdividing: $f(\frac{1}{4}) \equiv x_2$ could be any real number, as could $f(\frac{3}{4}) \equiv x_3$. Smoothing the curve to hit $0, x_2, x_1, x_3, 1$ is still not a complicated task.
- Keep subdividing: we eventually have an infinite sequence of $x_1, x_2, \dots$, where every single point could be any real number.

That is, we've put the set of curves into correspondence with the set of all combinations of real numbers: 2^{2^c}.

2.2 Metrics

The concept of distance doesn't exist in what we've been looking at so far, not for sets, and not even for individual points. There are several ways to invent such a thing.

- On a grid, you are very familiar with simple Euclidian distance, the shortest line between two points $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$, formally $d(p_1, p_2) \equiv \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ in two dimensions, and appropriately generalized in more. We have a Usual Topology; maybe we could call this the Usual Metric. The circle in Figure 2.2 is everything Usual-distance one from the origin.
- The *Manhattan metric* is $d(p_1, p_2) \equiv |x_1 - x_2| + |y_1 - y_2|$, where the vertical bars represent absolute value, and the overall gist is that we first have to walk along the East-West streets, then along the North-South avenues to get from point 1 to point 2, without the aid of the diagonals the Euclidian metric would measure along (that is, stay between 14$^{\text{th}}$ and 59$^{\text{th}}$ and assume Broadway does not exist). The diamond in Figure 2.2 is everything Manhattan-distance one from the origin.
- In the *Discrete metric*, $d(a, a) = 0$ for any point a, and for any

two distinct points, $d(a,b) = 1$. It's the lowest-possible-effort metric, asking whether two points are different, then refusing to do any further calculations.

The rules for a metric are three:

1. The distance from A to A is zero: $d(p_1, p_1) = 0, \forall p_1$.
2. Symmetry: $d(p_1, p_2) = d(p_2, p_1)$
3. The triangle inequality: $d(p_1, p_2) + d(p_2, p_3) \geq d(p_1, p_3)$. If you go from point A to point C directly, it's less of a walk than if you go from A to B then B to C—at best, it's the same distance.

Maybe pause here and verify that the three metrics so far (standard Euclidian, Manhattan, Discrete), all meet the rules.

Symmetry may seem pedantic, but it's a frequently broken condition. If p_1 is at the base of a hill and p_2 at the peak, then the Calories you would expend on a bike are greater going from $p_1 \to p_2$ than $p_2 \to p_1$. If you drive an automobile in actual Manhattan and use minutes of drive time as your metric, one-way streets and traffic congestion mean there is no reason to presume that the drive times $d(p_1, p_2) = d(p_2, p_1)$ for any pair (p_1, p_2) anywhere in the city. My personal favorite function which fails the symmetry condition is K-L divergence. Given two probability distributions, it measures the distance from one to the other, but in a manner such that $A \to B \neq B \to A$.

2.2.1 Metric topologies

But this is a topology book, so let's use the metrics to generate topologies. We do so by defining an ϵ-ball around a center point c as the set of points p such that $d(c,p) < \epsilon$. With real numbers and the standard Euclidian metric, we've just generated the Usual Topology. In $\mathbb{R}$, those are intervals around a center point like $(c - \epsilon, c + \epsilon)$, in $\mathbb{R}^2$ those are the interior of circles, in $\mathbb{R}^3$ the interior of spheres. At a single point, there are an infinite number of such balls, from size approaching zero to size approaching infinite, and such balls exist around every point.

- For the Manhattan metric in $\mathbb{R}^2$, our ϵ-balls look like fuzzy-edged diamonds, not spheres like Euclidian ϵ-balls, though we still call them balls, and we still have an infinite number of them centered around every point.

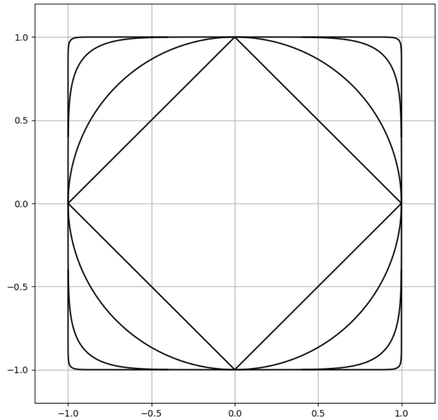


Figure 2.2: Some ℓ-norms: The unit ball in the Manhattan metric (ℓ_1) is the diamond. In the Euclidian metric (ℓ_2) the unit ball is the circle. The ℓ_6 norm is a rectangle with rounded corners, and the ℓ_{40} norm approaches a simple square.

- Or more generally, given a point p and a distance ϵ, define the set $\{x|d(p,x) < \epsilon\}$. That's the ϵ-ball in the context of whatever distance function you might be working with. Declare all of those, for all p and all ϵ, to be open sets, and you have induced a topology from a metric.
- Well, you almost have. Don't forget to include all unions and intersections. We tend to focus on the simple circles and spheres, but if you have a clump of six or seven simple open balls in some odd arrangement—maybe a balloon animal or looped into a necklace—that amalgam is also an open set.

Now that we have a topology, we can ignore the source metrics and focus on the open sets. Consider the set of open sets generated by the Euclidians and the set of open sets generated by Manhattanites.

- Pick an open set in the Manhattan-metric topology, such as every point (p_x, p_y) around a focal point (x, y) such that $|x-p_x|+|y-p_y| < 3$. That's a definite set of points, though do note the use of *less than* instead of *less than or equal*, giving us soft-edged boundaries just as with the canonical open set.

- This would not be the case if we had a hard boundary, but as it is, there is always a *Euclidian* ϵ-ball around any point in the Manhattanite set small enough that it is entirely within the set.
- If every point in a set is part of a Euclidian ϵ-ball, then you could construct the set from an infinite union of open sets in the Euclidian topology, then that set is in the Euclidian topology.

So the open-diamond topology is a subset of the open-circle topology. All the same logic holds if we start with a Euclidian-topology open set, call it E: every point is inside a Manhattan-topology open set within E, so E is part of the Manhattan topology. Dropping down to basic logic: When you have two sets A and B, and $A \subseteq B$ and $B \subseteq A$, then $A = B$. Applying that here, both metrics generate the same topology.

Both Euclidian and Manhattan metrics are in the same family, meaninglessly named the ℓ-norm family. Everything in the family has the form $d(p_1, p_2) = (|x_1 - x_2|^b + |y_1 - y_2|^b)^{\frac{1}{b}}$, including the Euclidian metric when $b = 2$, and Manhattan when $b = 1$. See Figure 2.2 for some examples. For $b \to \infty$, you get the *sup norm*, where $d(p_1, p_2) = \max(|x_1 - x_2|, |y_1 - y_2|)$.

Theorem 2.1. All ℓ norms are alike
All topologies generated by an ℓ-norm are equivalent to each other.

You've probably read about how topologists don't care about distances, or about how everything can be freely stretched and adjusted, and this theorem gets to the core of what that means. It's not that one metric transforms into another via appropriate transformation, but that each generates a truly identical set of open sets.

The Discrete metric produces one of two topologies: if $0 < \epsilon < 1$ you get the Discrete Topology, because if $d(a, b) = 1$ for all $a \neq b$ then no ϵ-ball spans two points; if $\epsilon > 1$ you get the Indiscrete Topology, because if $d(a, b)$ is either zero or one, an ϵ-ball of size 1.1 around a reaches all points in the whole space.

No, metrics didn't get us far in inventing new topologies, but it's worth taking the time to note that the concept of distance is somewhat separate from the open sets issue we want to talk about here. Your intuition might be that the best way to resolve Zeno's paradox is some means of clarifying what small distances mean (that's my intuition, at least), but that's not what the topologists will be doing at all: Which ϵ-balls overlap depends on closeness in the usual metric sense, but once we generate those open sets, we can otherwise throw away the metric.

In the opposite direction, given some arbitrary topology, is there some metric that generates it? Could you pin the points in the topology to points in $\mathbb{R}^N$ such that the open sets generated by the metric match those in the arbitrary topology? Chapter 9 will ask that question and give a straightforward answer: yes, if the topology is T_2, second-countable, and completely regular. The text between here and there will put you in a position to understand what that's supposed to mean.

2.3 Some less basic topologies

One reader of this book was saddened by how many topologies are over the space of plain old real numbers, which on the one hand shows some lack of imagination, given all the things in our world which could be grouped and subsetted, but on the other hand shows incredible imagination, that the same numbers can behave remarkably differently depending on the groupings we impose on them.

This section presents a few more topologies which go much further afield from the open intervals in the introductory list of basic topologies above. The first few in this section define "points" differently: a point is an entire graph or an infinite sequence. The last goes back to $\mathbb{R}$, but does surprisingly unusual things with it.

2.3.1 Graphs

The pedantic will recognize that lines along an x and y axis are *plots*, while a *graph* is a network of nodes and edges. One could build one incrementally, starting with a graph consisting of a single node, then adding another node and an edge between them, then adding another node and edge to produce a chain of 1—2—3. If there can be at most one link between any given pair of nodes then there are no choices to be made up to this point. But the next step could be adding just an edge linking nodes 1 and 3 to form a triangle; adding node 4 and an edge to node 3, producing a 1—2—3—4 chain; or adding node 4 and an edge to node 2, forming a three-pointed star centered at node 2. A *point* in this space will be a single graph, with all its nodes and edges. Then a topology over this space will be a list of sets of graphs.

Define an open set to be one graph plus all graphs which could be formed by adding an edge (possibly with a node), or subtracting an edge (possibly removing a node) from that base graph. Figure 2.3 shows the sorts of open sets that result.

- Almost all sets as defined here therefore include one graph, say it has n edges, and some number of graphs with $n+1$ and $n-1$ edges.
- But only *almost all*, because if the starting point is a one-node graph, there are no edges to remove. There is only one other graph in the set starting from the trivial one-node graph, with two nodes linked by one edge.
- The triangle at left is a complete graph: every edge between the three nodes is present. If we don't have labels on the nodes, a complete graph can expand in only one way, by adding a node and an edge anywhere, and can contract in only one way, by subtracting an edge. A graph with 30 nodes has a capacity of 435 edges, given 30 source nodes which could point to 29 destinations. If complete, even that has only one way to expand and one way to subtract.
- Other graphs have far more neighbors. The four-node sideways-T-shaped graph next to the triangle can expand four different ways.
- I'd described building a graph via a stepwise process, and the topology records how that building happens. You can trace a path from the root graph of just one node to the next graph with two nodes, which are in the same set, to the next graph with three nodes, which is in the same set as that two-node graph. Between any two graphs, there is a chain of open sets.

We've converted questions about growing networks into topology questions. Other topologies would encode very different information. Start over: define open set n as the set of all graphs with n nodes. This is a valid basis for a topology, but much less interesting, as there is no overlap between any open sets in our basis.

2.3.2 Arithmetic sequences

The space for this topology will be $\mathbb{Z}$, the integers $\dots, -2, -1, 0, 1, 2, \dots$. Let o and s be integers, and let $N_{o,s}$ be the set of points $\{o + xs | x \in \mathbb{Z}\}$. Here, o is the offset and s is the amount we step by. For example, $N_{2,5} = \{2 + 5x | x \in \mathbb{Z}\} = \{\dots, -13, -8, -3, 2, 7, 12, 17, \dots\}$. Then each $N_{o,s}$, $\forall o, s \in \mathbb{Z}$, is an open set in our topology.

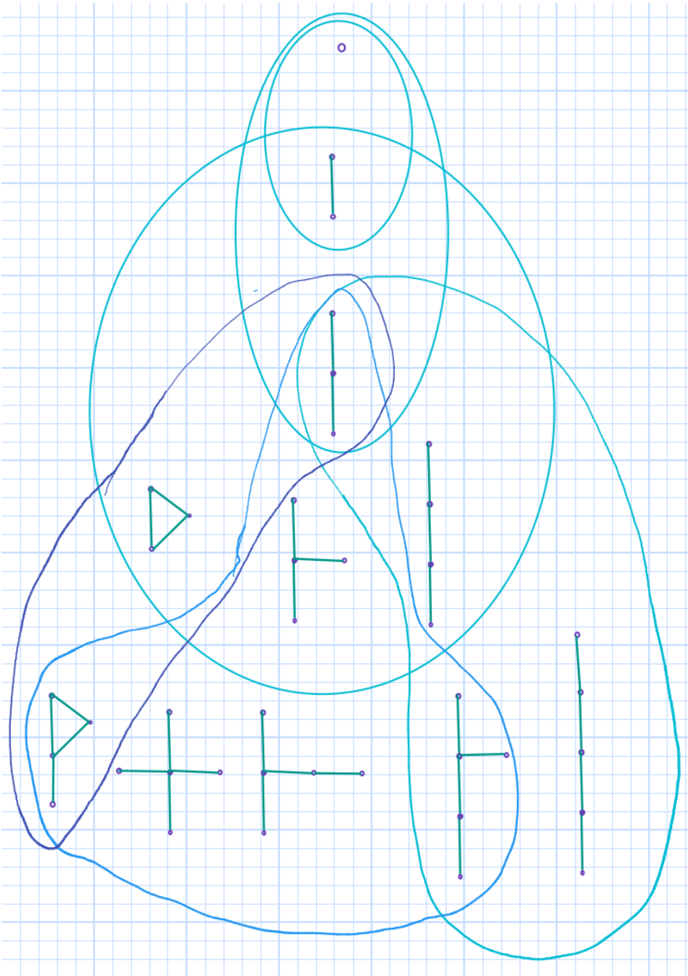


Figure 2.3: Graphs, and the open sets they lie in. Each graph is the center of one open set, including all graphs that can be built from adding or removing one edge—except the trivial graph of one node and no edges, which has no smaller graph. Sets centered around the graphs in the last row are no shown.

- If you have a few offsets o with the same step s, you can cover the space. For example, $N_{0,3} = \{\dots, -6, -3, 0, 3, 6, \dots\}$, and $N_{1,3} = \{\dots, -5, -2, 1, 4, 7, \dots\}$, and $N_{2,3} = \{\dots, -4, -1, 2, 5, 8, \dots\}$, and all the integers are in one or the other. The easiest such case is $N_{1,2}$, which is all the odd integers, and $N_{2,2}$, which is all the evens.
- It might be amusing for you to verify that the intersection and the union of $N_{a,b}$ and $N_{c,d}$ is a sequence also of the form $N_{e,f}$, by solving for e and f in the cases of union and of intersection. Of course, for cases like odds and evens in the prior bullet point, the union is $\mathbb{Z}$ and the intersection is $\emptyset$. But multiples of three alternate even and odd, so $N_{0,3}$ intersects $N_{0,2}$ in a definite pattern.
- By the way, we're using $\mathbb{Z}$ because the German for *numbers* is *Zahlen*.

We've imposed a shape and texture to the monotone number line. Odds and evens are disjoint, so if you want a chain of overlapping open sets along you can step from odds to evens, you have to go through some other sequence, like walking from evens (the set $N_{0,2}$) to multiples of three (the set $N_{0,3}$) to odds (the set $N_{1,2}$). The sequence $0 + 6n$ and the sequence $1 + 6n$ are also disjoint, but you can't get from one to the other via multiples of three: $0 + 3n$ will never meet $1 + 6n$ and $0 + 6n$ is entirely a subset within $0 + 3n$; and $1+3n$ is a superset of $1+6n$, and $1+3n$ will never meet $0+6n$. Where a regular person might say *Six is three times two*, a topologist could say *for $i = 0, 1$, $N_{i,6}$ is a subset of $N_{j,2}$ and of $N_{j,3}$ for $i = j$ and is disjoint from those sets when $i \neq j$.*

If you were a number theorist, this is how you'd start asking questions of numeric sequences using the topological tools you'll see over the course of this book. Later, we'll use this topology to prove that there are an infinite number of primes.

2.3.3 ◯◯ The Outer-thirds Set

I had an actual dream while ideating about this book, about Trinity Island. There was an East and a West part with a valley in between. It was the sort of place where the coffee shops made ordering complicated. Small, medium, large? Unprocessed, half-caffeinated, decaf? No milk, half, or something more milk than coffee? Oat, almond, or soy milk? The cashier would write your decisions on your cup, like 0220. The apparel shops were all well-stocked, with sections for men, women, and everybody. Even within the men's

Figure 2.4: The Outer-thirds Set: start with $[0,1]$, then remove the middle third. Next step: of the two segments remaining, remove their middle thirds. Next step: of the four segments remaining, remove all middle thirds. Repeat infinitely.

section, there were male-leaning, uncoded, and female-leaning apparel. We went back years later, and climate change had left the valley permanently flooded, leaving an East island and a West island. People were gruff and polarized. The coffee shop had reduced your options so they could move people faster: small or large, caffeinated or no, milk or no, though my order still turned out to be 0220. The weirdest part was that the locals didn't notice a difference. Even after the losses it still felt like a lot of latitude to move about.

1. Start with the interval $[0,1]$, then remove the middle third, leaving two intervals, $[0,\frac{1}{3}] \cup [\frac{2}{3},1]$. The top of Figure 2.4 is the whole interval, then the next row down is with the middle third removed.
2. Repeat, taking out the middle thirds again, so writing $\frac{1}{3}$ as $\frac{3}{9}$ and $\frac{2}{3}$ as $\frac{6}{9}$, the interval $[0,\frac{3}{9}]$ gets cut into $[0,\frac{1}{9}]$ and $[\frac{2}{9},\frac{3}{9}]$, and $[\frac{6}{9},1]$ gets cut into $[\frac{6}{9},\frac{7}{9}]$ and $[\frac{8}{9},1]$.
3. Take those four intervals, $[0,\frac{1}{9}]$, $[\frac{2}{9},\frac{3}{9}]$, $[\frac{6}{9},\frac{7}{9}]$ and $[\frac{8}{9},1]$, and remove the middle thirds to produce eight intervals. Repeat, infinitely.

What's left is the Outer-thirds Set. As a historic note, the development of the Outer-thirds Set was a turning point in Western mathematics, because of the properties discussed below. But after some digressions about measures, we're going to use it as an eccentric topology. It will reappear to cause trouble throughout.

- The Outer-thirds Set is usually called *The Cantor Set*.
- *Stigler's Law*, first articulated in the academic literature by Robert K Merton, states that everything named after a person is named after the wrong person.
- Henry J Stephen Smith, in an 1874 paper on integrability of discontinuous functions, offers a construction where $[0,1]$ is divided into m segments, then the first $m-1$ segments are

divided into m segments again, and, with $m = 3$, we wind up with a sequence of divisions that looks much like the middle-third sequence of Cantor's Set (published 1883), albeit with the unbroken thirds at the end and not the middle. Even the most revolutionary ideas have precedent somewhere.

Binary and trinary

Let's take a beat to remember how binary, decimal, and other base-n systems work.

- We do this instinctively, but a number like 283 breaks down to 200 + 80 + 3, whose pattern is most clear if we spell out the powers of ten: $2 \cdot 10^2 + 8 \cdot 10^1 + 3 \cdot 10^0$.
- This also works after the decimal point, bearing in mind that $10^{-1} = \frac{1}{10}$ and $10^{-2} = \frac{1}{100}$. E.g., the number $23.76 = 2 \cdot 10^1 + 3 \cdot 10^0 + 7 \cdot 10^{-1} + 6 \cdot 10^{-2}$.
- Binary works by replacing all the 10 bases with 2 bases. For example, $101.1 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} = 5\frac{1}{2}$.
- Since the digital revolution, we've all gotten used to binary, but one could do this with any base, like trinary, where $12.02 = 1 \cdot 3^1 + 2 \cdot 3^0 + 0 \cdot 3^{-1} + 2 \cdot 3^{-2} = 5 + \frac{1}{9}$.

Now that you're fluent in trinary, we'll be looking at the set of points between zero and one, inclusive, expressible in trinary using only zeros and twos.

- Putting that set of numbers into one-to-one correspondence with binary numbers couldn't be easier: just replace all the twos with ones, like 0.2202 becomes 0.1101.
- When you've done so, you've now got all binary decimals between zero and one, which means all of $[0, 1]$.
- But on the trinary side, there are holes, because all those numbers that need a one to be expressed, like 0.12, were excluded. The range from $0.1\overline{000}$ to $0.1\overline{222}$ is the range between $\frac{1}{3}$ and $\frac{2}{3}$. I.e., it's the part that we threw out in the first step of generating the Outer-third set.
- At the next level, anything between 0.01 and 0.02 will need a one to be expressed, by which I mean everything between $\frac{1}{9}$ and $\frac{2}{9}$. Same with anything between 0.21 and 0.22, meaning everything between $\frac{7}{9}$ and $\frac{8}{9}$.

- As a relevant aside, you know how $0.\overline{999}$ is another way to write 1?
  - If $x \equiv 0.\overline{999}$, then $10x = 9.\overline{999}$.
  - Subtracting $10x - x$, everything after the decimal cancels out: $10x - x = 9.\overline{999} - 0.\overline{999} = 9$.
  - If $10x - x = 9x = 9$, then $x = 1$.
  - The same holds for later digits, like how $0.72\overline{999} = 0.73$.
- You can do the same in trinary, where $0.\overline{222} = 1$. That means elements like $\frac{1}{9} = 0.01$ really are in our set, because we can write them as $0.00\overline{222}$.
- But maybe you see where this is going: every point in the Outer-thirds Set, as constructed by removing middle thirds from ever-shorter lines, can be expressed as a trinary decimal using only zeros and twos, and vice versa.

Measure theory can get weird, but it starts simple: the range $[0,1]$ has measure (total length) one. So does the open interval $(0,1)$. If you subtract one from the other, you are left with two points, 0 and 1, and there is no trickery in measuring them: we took a length of measure one and subtracted a length of measure one, leaving two points with measure zero.

With the Outer-thirds Set, we started with $[0,1]$ and removed one segment of length $\frac{1}{3}$, two of length $\frac{1}{9}$, and so on, for a total length

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \sum_{i=1}^{\infty} \frac{2^{i-1}}{3^i}.$$

I won't present a proof, but that's a convergent sum, which totals one. From a line segment of measure one, we removed segments with total measure one, meaning that the Outer-thirds Set has measure zero. This is how Cantor blew up mathematics. We have the same number of points, but in one configuration the set has length one, and in another it is a bunch of disconnected points with length zero. Instead of a world where some sets are bundles of points and some are lines, we have something in between—Chapter 10.2 finds a compromise between calling it zero-dimensional like a bundle of points and one-dimensional like a line by a calculation which concludes that it is 0.631-dimensional.

We've constructed a countably infinite set of points, with an easy one-to-one correspondence with all points in $[0,1]$ (changing all 2s to 1s, like 0.2202 to 0.1101), but which itself has measure zero.

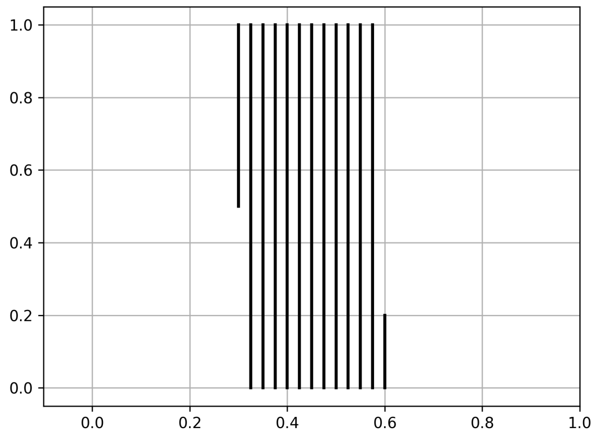


Figure 2.5: A schematic of the open set between $(0.3, 0.5)$ and $(0.6, 0.2)$ in the Lexicographic Topology.

This is itself a great feat of interesting. But this is a topology textbook, so we're going to add it to our list of topologies as our first subspace topology: take the Usual Topology, then restrict the set of points to the Outer-thirds Set, and the set of open sets to the intersection of the usual open sets with the Outer-thirds Set.

- One more interesting bit: As in our usual base-10 decimals, a number ending in a repeating pattern, possibly $\overline{000}$, is rational, but there are infinitely more irrationals whose expression never repeats. Because the endpoints are all cuts made at a rational number, you might expect that there are an uncountable number of points which are not endpoints of segments of the Outer-thirds Set, and so in the interior. We'll see below that there are in fact *no* interior points. Chapter 3 will introduce limit points, however, which will allow us to say something about how points in the set clump together.

2.3.4 ○○ Lexicographic ordering

When you alphabetize a list, only the first letter matters—unless two words both start with the same letter, in which case only the second letter matters.

We could order $\mathbb{R}^2$ like this, first using the x-axis and then the y-axis only to break ties. For two given points, $(x_1, y_1) > (x_2, y_2)$ iff either (1) $x_1 > x_2$, or (2) $x_1 = x_2$ and $y_1 > y_2$. Set the space for the Lexicographic Topology to be the one-by-one box $[0,1]\times[0,1]$. It's all very simple, the numeric equivalent of the sort of alphabetization you learned when you were maybe seven years old—what could go wrong?

Each pair of points p_1 and p_2 generates what we'll call a basic open set, $\{q|p_1 < q < p_2\}$. This looks like how open intervals are defined in the Usual Topology, but this is not a subset topology or otherwise related to the Usual—we are effectively redefining what $<$ means, so the open sets will look completely different. We broke the concept of distance entirely:

- The distance from $(0.5, 0.5)$ to a point just above it like $(0.5, 0.6)$ is sensible and not actually broken, and we can call it 0.1.
- Between $p_1 \equiv (0.5, 0.5)$ and $p_2 \equiv (0.6, .05)$, though, we have all points of the form $(0.55, y)$, with $y \in [0,1]$, so by our usual expectations that's length one. We also have all points of the form $(0.56, y)$, $(0.57, y)$, and in fact an infinite number of lines of height one between p_1 and p_2. Add them up, and the distance between p_1 and p_2 is infinite.

So we have a complete order, where for every two points we can say which is greater than the other, and we have a sensible concept of distance on any vertical bar. But if p_1 and p_2 aren't on the same vertical bar, there is infinite distance between them.

For our open sets, instead of balls, visualize vertical stripes, starting at (x_1, y_1), including all values above that point, every point with an x-value between p_1 and p_2 regardless of y-value, then topped off with every point south of (x_2, y_2).

Then, don't forget to take all unions and intersections and include the complete one-by-one box and $\emptyset$ in your topology.

- Put the ϵ-balls in the Usual Topology out of your mind. Figure 2.5 gives a schematic of the open set between $(0.3, 0.5)$ and $(0.6, 0.2)$. It starts at $(0.3, 0.5)$ and included everything above that point with $x = 0.3$, then includes every vertical bar with $0.3 < x < 0.6$, then finishes with everything with $x = 0.6$ but $y < 0.2$. I'm calling the figure a schematic because there are of course an infinite number of vertical bars in the body of the open set, not just the few I've drawn.
- What looks like a straight horizontal line, say from the point $(\frac{1}{4}, \frac{1}{2})$ to the point $(\frac{3}{4}, \frac{1}{2})$, is actually an infinite sequence of dis-

connected points. In this topology, an ϵ-ball with $\epsilon < 1$ will never contain two of the points in that horizontal line, because it takes a step of one to wrap around from bottom to top.

- What would an ϵ-ball in this world look like? The ball of radius $\frac{1}{4}$ around the center point $c = (0.3, 0.5)$ seems natural: it's the vertical line between $(0.3, 0.25)$ and $(0.3, 0.75)$.
- The ball of radius $\frac{1}{4}$ around a center point $c = (0.4, 0)$ goes up to $(0.4, 0.25)$, and down to... nowhere, because any candidate point less than c is on another vertical line, and as above, there are an infinite number of vertical lines of height one between that point and c.

It still feels weird to me. Everything is perfectly ordered, using an ordering scheme you learned in grade school, and yet at regular intervals there is a drop-off where no points seem to be. Some call this topology *the long line*, because we usually visualize fully ordered lists in a line, but this line has infinite distance between any two points where $x_1 \neq x_2$.

This chapter introducing topologies has gone full circle, so to speak. We started with the Usual Topology, which produces what you expect and are familiar with within the subspace $[0, 1] \times [0, 1]$, and ended with the Lexicographic Topology, which assigns new open sets to exactly the same bunch of points. How points of interest connect is vital, and is what Topology is about. Manhattan is what it is for everybody, but People who travel around it by taxis have what is in many ways an entirely different conception of what Manhattan is from the ones who travel by subway.

Chapter 3

Closed

The definition of *closed set* is even simpler than the definition of *open set*:

Definition 3.1. Closed set
A set S in a space X is closed iff its complement, $X - S$, is open.

- Notation reminder: complements are written with an over-tilde, $\widetilde{S} \equiv X - S$.
- For example, if (∞, a) and (b, ∞) are open sets, as they are in the Usual Topology, and if $a < b$, then the complement to the union of those, $[a, b]$, is a closed set, as we would expect.
- But to stress that we're starting anew on what *closed* means, note that this definition easily allows a set to be both open and closed, in which case it is given the awkward name of *clopen*. Both $\emptyset$ and the whole of X are open by definition, and $\widetilde{X}$ is $\emptyset$, and $\widetilde{\emptyset}$ is X, so these two sets are always clopen in any topology.
- If we had a trivial topology consisting of two points, $X \equiv \{a, b\}$, where both $\{a\}$ and $\{b\}$ are open sets, both are clopen. More generally, in the Discrete Topology (§2.1.5), every set is clopen, because the union of all points but p is an open set.
- A set can be neither open nor closed. Say that our space is $\{p_1, p_2, p_3\}$, and our list of open sets is $\{\{p_1, p_2\}, \{p_3\}\}$. Then both $\{p_1, p_2\}$ and $\{p_3\}$ are clopen, and all other not-complete and nonempty combinations, like $\{p_2\}$ by itself or $\{p_2, p_3\}$, are neither open nor closed.
- By the way, topologists have one more way of bringing sets into the hierarchy of open sets: a *neighborhood* is a set of any

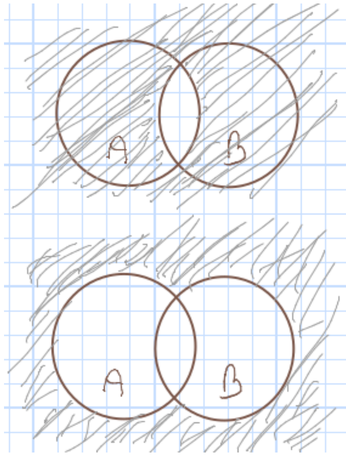


Figure 3.1: Top: the complement to the vesica piscis, both $\widetilde{A \cap B}$ and $\widetilde{A} \cup \widetilde{B}$.
Bottom: the outer regions, both $\widetilde{A \cup B}$ and $\widetilde{A} \cap \widetilde{B}$.

type which contains an open set.

Intersections and unions
See Figure 3.1 for some diagrams confirming that the complement of an intersection is the union of the complements ($\widetilde{A \cap B} = \widetilde{A} \cup \widetilde{B}$), and the complement of a union is the intersection of the complements ($\widetilde{A \cup B} = \widetilde{A} \cap \widetilde{B}$) (de Morgan).

Our definition of *open set* had the smallest of asymmetries, that the finite intersection of a set of open sets must also be open, and the possibly infinite union of open sets is open. If we wanted to talk about the complements of open sets, i.e., the list of closed sets, the possibly infinite intersection of any of those will be closed, and the finite union of closed sets is also closed. It is a thin distinction.

If we wanted to, we could build a topology by writing down whatever list of closed sets we want, then carrying on to find all finite unions and arbitrary intersections, then taking the complement of all of those as the list of open sets. By the logical identities above, that list will be appropriately complete.

You may be familiar with how the complement of an open set is closed from real algebra, where we have much stronger expectations of closed sets. There's no Zeno's paradox shenanigans at the edge of $[0,1]$, as 1 is in the set and there's a clean break. Even the notation of $(0,1)$ versus $[0,1]$ hints at a fuzzy edge versus a sharp edge.

But at this point in the narrative our definition of *open set* is *whatever you want to call an open set* (plus finite intersections and infinite unions), meaning our definition of *closed set* is *the complement to whatever you want* (plus infinite intersections and finite unions). We're still just making it up, and yet at the end of this chapter you'll have another definition of *closed* grounded in the distinction between sharp versus round edges. It is, to me, still a little astonishing that the two approaches are identical.

3.0.1 On the infinitude of primes

That there are an infinite number of primes has been known for millennia. If there weren't, if there were a finite list of primes, $p_1, p_2, \ldots, p_n$, then we could take the product $P \equiv p_1 \cdot p_2 \cdot \cdots \cdot p_n$, and add one, and $P+1$ is not evenly divisible by any of the primes on our list. There's your one-sentence proof that the idea of primes being finite is self-contradictory, but as either an exercise in recreational math or as practice in bouncing between open sets and their closed-set complements, we can prove that there are an infinite number of primes via the Arithmetic Sequence Topology [Furstenberg, 1955].

The Arithmetic Sequence Topology was defined in §2.3.2: for any integer offset o and integer step s, the open sets of our topologies are named $N_{o,s}$, the set of points $\{o + xs | x \in \mathbb{Z}\}$ (where $\mathbb{Z}$ is the integers).

I gave the example of the evens, $N_{0,2} = \ldots, -4, -2, 0, 2, 4, \ldots$ and the odds, $N_{1,2} = \ldots, -3, -1, 1, 3, 5, \ldots$. Of course, the evens and odds don't overlap, and in fact, the complement of the evens is the odds, and vice versa. In the language of this chapter, the complement to the evens is an open set (the odds), so the evens are a closed set; same with the odds. In fact, both are clopen. For multiples of three, the integers break down into the three sets $N_{0,3}$, $N_{1,3}$, and $N_{2,3}$, so the complement of $N_{1,3}$ is $N_{0,3} \cup N_{2,3}$, so $N_{1,3}$ is closed. You could construct the complement of any open set $N_{i,j}$ from its sibling open sets $N_{k,j}$ where $k < j$ but $k \neq i$, meaning every open set is actually clopen.

- Of course, the open sets are more than just the N-sets we wrote down, but also their intersections and unions. Earlier, I'd suggested thinking through what those look like, but the main point is that, except cases where the intersections are $\emptyset$, these composite sets always have an infinite number of elements in them, at a regular interval. For example, $N_{0,2}$ intersects $N_{0,3}$ at every multiple of six (i.e., $N_{0,2} \cap N_{0,3} = N_{0,6}$). If a number n is in an intersection or union of $N_{a,b}$ and $N_{c,d}$, then $n + b \cdot d$ will also be in the amalgam—in fact, every $n + z \cdot b \cdot d$ will be, $\forall z \in \mathbb{Z}$.

- Unions don't have quite so even a rhythm, but they are still certainly infinite, like how the positive part of $N_{0,2} \cup N_{0,3}$ is $\{2, 3, 4, 6, 8, 9, 10, 12, \ldots\}$.
- Every number is either a prime number or a multiple of a prime—except one and negative one, which get skipped by the multiples of $2, 3, 5, 7, \ldots$.
- Turning that observation into more formal notation, that means $\mathbb{Z} - \{-1, 1\} = N_{0,2} \cup N_{0,3} \cup N_{0,5} \cup \cdots = \bigcup_{p \in \text{primes}} N_{0,p}$.
- The definition of an open set dictates that the finite intersection of open sets is open; then almost by definition, the finite union of closed sets is closed. If the number of primes were finite, then the right-hand side of that equation is the finite union of closed sets, a closed set.
- An aside: If there were an infinite number of primes (spoiler: there are), we don't know anything about the infinite union of closed sets. The definition just doesn't address that situation. Those $N_{0,p}$s are also open sets, and we do know the infinite union of open sets is an open set. But that's not how we're going to corner ourselves into a contradiction, so let's go back to the presumption of a finite number of closed sets, which we know to be closed.
- So if the number of primes were finite, then the right-hand side of the above equation is a closed set, meaning the left-hand side, $\mathbb{Z} - \{-1, 1\}$ is a closed set, meaning its complement $\{-1, 1\}$ is open. But go back to that first bullet point, where every open set has an infinite number of points in some kind of regular interval.

Something broke: with a finite number of primes, $\{-1, 1\}$ has to be an open set, but all nonempty open sets have an infinite number of elements. The only conclusion is that no, the primes aren't finite. As promised, we arrived at the same conclusion that we arrived at in one sentence at the head of this section, but we learned some things about sequences in the process, notably that their combinations are also regular infinite sequences, as are their complements. For those properties to be consistent, there have to be an infinite number of primes generating an infinite number of regular sequences.

3.1 Limits

The definition of *closed set* at this point is very different from the one you learned when you first learned about closed intervals, when you were told that a closed interval is one which includes its end points, like $[0,1]$. The goal of this chapter is to make this jump: how can we coherently explain that iff a set is the complement of a set on a list we made up with almost no rules, then it includes its end points?

First, we'll need to formalize what we want from those end points:

Definition 3.2. Limit of a set
A point x is the limit of a set A iff every open set containing x intersects with $A - x$.

When you saw the word *limit* maybe you were expecting something about an infinite sequence converging to its limit. If you were, take the name as foreshadowing. For now, let's get used to the definition with some easy examples.

- We can start with the canonical open set in the Usual Topology, $(0,1)$, and its eastern edge, one. Any u-open set around one must include some values less than one, and thus intersect $(0,1)$. So it checks out: one is a limit point of $(0,1)$.
- For the closed set $[0,1]$, one is still a limit point in a very identical manner: every u-open set around one intersects with $[0,1] - 1 = [0,1)$.
- Any interior is also going to be a limit point. For example, any u-open set around a half certainly intersects $(0,1) - \frac{1}{2} = \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right)$.

Let me tell you about my first-person perspective as a limit point. I'd described myself as a computational social scientist in the intro. I had a professor once (Michael Chwe) who defined *economics* as *the application of math to the social sciences*. But the great majority of economists restrict themselves to a smallish subset of methods, so I go by this economist-adjacent title of computational social scientist, and I let my membership to the American Economic Association (AEA) lapse. By no means does that make my life economist-free, and in fact every social circle I find myself in has an AEA member somewhere in it (or it at least feels that way sometimes). The AEA is huge and the great majority of its members will never know I exist, but from my perspective, it is unavoidable and I can't separate myself from the AEA. I'm as AEA-adjacent as possible without actually being listed as a member of the organization itself.

3.1.1 Points adrift

If a point isn't in any open set, then it is a limit point of every open set. This holds by applying the definition very literally: if there are no sets containing x, the clause about *every open set containing* x *intersects with* $A - x$ in the definition applies vacuously.

That tells us something about the topological worldview: if something isn't in an open set, it is at sea, with no fixed location, adjacent to everything at once. This is not a situation you'll encounter often, as it's an oddity topologists will avoid when designing a topology, but it isn't against the rules and is perhaps a sometimes elucidating example to consider.

3.1.2 Limits of the Outer-thirds Set

As an exercise in limit-point finding, what would a limit point of the Outer-thirds Set look like?

- Recall (§2.3.3) that only numbers expressible in trinary using only zeros and twos are in the set.
- Call a potential limit point x; we want every ϵ-ball around that point to intersect the Outer-thirds Set somewhere besides x.
- Say that x's trinary expression does have a one in the N^{th} decimal point. *Trinary decimal* is weird, though; maybe *trecimal* would be better.
- There is some junk after that point, in trecimals $N + 1$ and beyond. If there weren't, and x has a trecimal ending in $1\overline{000}$, you will recall that we can rewrite that as a trecimal in the Outer-thirds Set ending in $0\overline{222}$—the 1 was a mirage.
- That there's junk after the one in whatever trecimal point means you'd have to subtract that junk to drop to a number that has a zero instead of the one. Rather than calling the amount junk, foreshadow where this is going by calling that amount you'd have to subtract ϵ_1.

- What would you have to add to x to raise the N^{th} decimal from its current value of 1 to 2? Call that upward adjustment ϵ_2.
- Take the smaller of the two ϵs, and that is the diameter of an ϵ-ball where both up and down, every number in the ball has a 1 in the N^{th} position, and therefore is not in the Outer-thirds Set. We tried to make x a limit point, and failed.
- Now consider a new point x inside the Outer-thirds Set: every decimal is either a zero or two. Can we find some other point in The Set within a given ϵ-ball of size $\frac{1}{3^N}$? No problem: at trecimal point $N+5$ if you see a 0 replace it with a 2; if you see a 2 replace it with a 0. We've checked all the boxes on the form: the new number is within the ϵ-ball, is in The Set, and is different from x.

Like the $[0, 1]$ interval, every point in the Outer-thirds Set is a limit point of the set, and the set contains all its limit points. But unlike the canonical closed set, between every pair of limit points of the set, you will easily find a number with a 1 in its trecimal expansion somewhere, which as above is not a limit point (it has an ϵ-ball around itself with no Outer-thirds Set members). This hints at how limits are not necessarily about closeness, the way they taught it in Calculus, but about what is connected to what.

- In an interesting little expression of a value judgment, a set like the Outer-thirds Set which equals the set of its own limit points is called *a perfect set.*

3.2 Closure

Here we move on to another definition, whose name gives away exactly where this is going.

Definition 3.3. Closure
The closure of a set A, annotated $\overline{A}$, is the union of the set and all of its limit points.

Above, 1 was a limit point of $(0, 1)$, as is 0, so the closure of the canonical open set is $[0, 1]$, as we expect. Having settled the easiest case, let's move on to something more unusual.

3.2.1 The Topologist's Sine Curve

This is a subspace of $\mathbb{R}^2$, so a subset topology as per §2.1.3. The set of points is the origin, $(0, 0)$, plus $f(x) = \sin(1/x)$, or more formally the

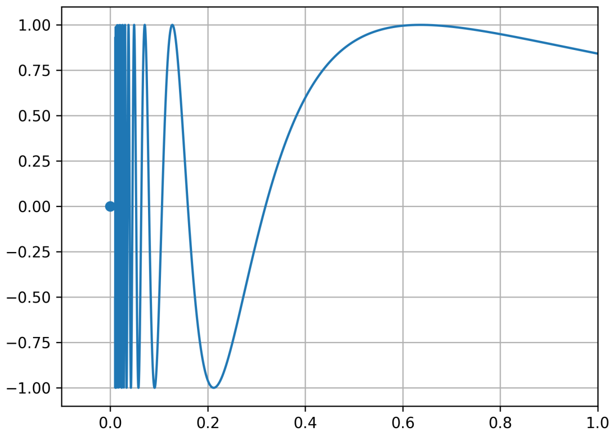


Figure 3.2: A plot of $\sin\left(\frac{1}{x}\right)$ for $x > 0$, plus the origin.

ordered pairs $\{(x, \sin(1/x)) | x \in (0, 1]\} \cup (0, 0)$. The set of open sets is the subspace topology you get from the intersection of the usual ϵ-balls from the Usual Topology with this sine curve, plus the origin.

- You know what $\sin(x)$ by itself looks like: one up-and-down cycle as x goes from 0 to 2π, then another cycle between 2π and 4π, repeating out to infinity.
- But we want $\sin(1/x)$. If $\frac{1}{4\pi} \approx 0.0796$ and $\frac{1}{2\pi} \approx 0.15915$, then $\sin(1/x)$ takes a full cycle over the $\frac{1}{4\pi}$ span between them of 0.0796 units of x. With $\frac{1}{6\pi} \approx 0.0531$, the 4π to 6π cycle takes only 0.0265. The 6π to 8π step happens in an x-span of 0.0133. As x gets smaller, $\sin(1/x)$ cycles faster and faster.
- There are so many doughnuts in topology, but here's a croissant: if you look around the origin, you find an infinite number of layers; see Figure 3.2.

What is the closure of the Topologist's Sine Curve?

- The bulk of the curve behaves much like any non-curved interval of the real line: every point on the path is a limit point, and the vacillation at infinite speed of $\sin(1/x)$ doesn't change that.
- Regarding the origin, we can think this through by considering the points crossing the x axis: $\frac{1}{\pi}, \frac{1}{2\pi}, \frac{1}{3\pi}, \ldots$.

- The origin is a limit point of that sequence: given any ϵ-ball around it, there is an N large enough that $\frac{1}{N\pi}$ is less than ϵ.
- But the sine curve is more than just the x-axis. A sine curve traverses from $y = -1$ up to $y = 1$, and along any horizontal level between those boundaries, our sine curve will register an infinite number of points. It's not just the origin that is a limit, but any point along the entire vertical axis from the point $(0, -1)$ up to the point $(0, 1)$ which is in the closure of the curve. (You'll find some authors who define the Topologist's Sine Curve as this closure, by the way.)

3.2.2 Truth in labeling

Yes, the closure of any set is closed:

- The discussion in §3.1.1 pointed out that any point that isn't in any open set must be a limit point of any given set A, and so in $\overline{A}$. Let's focus on the points not in $\overline{A}$.
- If a point x is not in $\overline{A}$ then it is not a limit point of A. Checking the definition of *limit point* (and bearing in mind the last bullet point), it tells us x is in at least one open set which doesn't intersect $\overline{A}$.
- The same could be said for *all* points not in $\overline{A}$. Take the (possibly infinite) union of those open sets which contain points not in $\overline{A}$ and don't intersect $\overline{A}$, and you have an open set which covers the complement of $\overline{A}$.

We've arrived at the prior definition of *closed*: if the complement of $\overline{A}$ is open, then $\overline{A}$ itself is closed. Or do this in the other direction, using the closed complement definition to show that limit points of a closed set cannot live outside of the set:

- If A is closed, its complement $\widetilde{A}$ is open, and by definition doesn't intersect A.
- This means every point x in $\widetilde{A}$ can't be a limit point of A, because $\widetilde{A}$ is an example of an open set containing x but not intersecting A.

Stating those last two points as a theorem:

Theorem 3.1. Limits $\subset$ closed
Any closed set A contains all of its limit points.

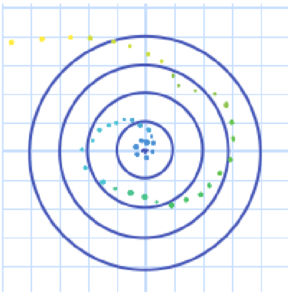
Or to summarize the whole discussion: any set which is a closure is closed; every set which is closed is a closure. We can use this iff relation as an alternative definition to replace Definition 3.1:

Definition 3.4. Closed
A set is closed iff it contains all its limit points.

So, yes, the *pick a set, any set* definition and the *contains its edges* definitions exactly match. This will become even more of a feat when we match up this still-abstract definition of limit points with the sort of *limit of an infinite sequence* definition everybody else uses.

3.3 Limits and infinite sequences

We have intuition about the meaning of a limit, that it is what a sequence like $0.9, 0.99, 0.999, \ldots$ is approaching but never reaches. Using just open sets, without any Calculus-speak about infinitesimals, we developed something that matches that intuition.



But what happens when you have an actual infinite sequence of points? We'll generalize this later, but for now let's consider metric spaces (§2.2), where distances are well-defined and the set of open sets is generated by the metric as the set of open balls of all sizes. An infinite sequence $x_1, x_2, \ldots$ is convergent in that context iff for any arbitrarily small ϵ, there is some N such that every value after N, meaning $x_N, x_{N+1}, x_{N+2}, \ldots$ is within ϵ of the point of convergence (Cauchy). In the doodle, even though the sequence of points sometimes drifts inward toward the convergence point and sometimes out, for any given ϵ-ball there is a point of no return after which it remains inside the ball.

Theorem 3.2. Closed iff convergence in a metric
Given a space X with a metric and the metric-induced topology (§2.2.1). A set A is closed iff any convergent infinite sequence of points in A converges to a point also in A.

- If the point of convergence is outside of a closed set A, meaning the complement of A is open, then there is some ϵ-ball around the point of convergence outside of A.

- We defined convergence to mean that for any ϵ-ball, there's a position in the sequence after which every point is in that ball. The above bullet point, where we have an ϵ-ball excluding any and all points in A, ruins that entirely. So if A is closed, the point of convergence can't be outside of A.

Now let's go in the other direction, and take as given that all infinite sequences of points in A converge to a point inside of A. Is that sufficient to show that A is closed?

- Every ϵ-ball around the sequence's limit contains some points in the sequence, meaning points in the set A. Flip back to Definition 3.2: if every open set containing a point p intersects a set $A - p$, then it is a Definition 3.2-limit point of A. It checks out: under our setup here, the limit of the sequence is a Definition 3.2-limit point of the set.
- Definition 3.1 said that iff a set contains all its Definition 3.2-limit points, then it is closed.

So yes, with a metric, our abstract open-set definition of limits matches the concept of a limit as the convergence point of a sequence, making Theorem 3.2 a sort of special-case restatement of Theorem 3.1.

We assumed away divergent sequences, and if you've read much Real Analysis you know there's a lot of hand-wringing about whether any given sequence will converge at all. Chapter 8 will address that aspect.

It would also be nice to distance ourselves from metrics and still talk about convergence, and there's a reasonable way to do so:

Definition 3.5. Convergent sequence

A sequence of points $x_1, x_2, \ldots$ converges to a point x in a topology iff, for any open set U such that $x \in U$, there exists an n such that $x_m \in U$ for all $m > n$.

In situations where there is an infinite nesting of sets within sets, this looks much like the definition of a convergent numeric sequence. If you have $(-1,1)$, $\left(-\frac{1}{2}, \frac{1}{2}\right)$, $\left(-\frac{1}{3}, \frac{1}{3}\right)$, $\ldots$, you could pick one point from each set, $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$, and get a convergent sequence. In both metric and non-metric cases, for every open set there exists a point of no return in the sequence were all further points in the sequence are going to also be inside of that set.

I'd introduced this discussion of convergence using the metric-oriented definition of convergent sequences, but without a metric, all of the discussion about the equivalence between our definitions of limit point and convergence point of a sequence still holds, replacing *ϵ-ball* with *whatever open set is on the list you made up.* Once again, the metric was an easy way to generate an infinite number of open sets at once, but we don't need distances *per se.*

3.4 Dense

Sometimes, the closure of a set is the entire space.

Definition 3.6. Dense
A set $A \subset X$ is dense iff the closure of A is X (i.e., $\overline{A} = X$).

You will recall (§2.1.7) that rational numbers are expressible as a fraction of two integers, and their decimal expansion always ends in a repeating sequence—even a simple one like $\frac{3}{4} = 0.75$ actually ends in repeating zeros, $0.75\overline{000}$.

- But consider an ϵ-ball around an irrational number, k. By referring to ϵ-balls, we are indicating that this segment is about the Usual Topology over $\mathbb{R}$. Recall that ϵ is not a variable, and for any given ϵ-ball it takes some fixed value, something small like 10^{-100}. For any ϵ, there is some power p such that $\epsilon < 10^{-p}$.
- Sticking with $\epsilon = 10^{-100}$ for expository purposes, everything in that ball has the same first 99 digits in its decimal expansion as the center point k. Any variation has to be smaller.
- But at smaller values, we get *all* the variation. Within our ϵ-ball of size 10^{-100}, every possible combination of digits is traversed over the 101^{st} digit and beyond.
- That will include pleasant patterns like $\overline{000}$ and $\overline{555}$ for the 101^{st} digit and beyond. Which is to say that within that ϵ-ball around k, there must be rational numbers.
- So we have the set of rationals, $\mathbb{Q}$, and a point k, and every open set containing k also intersects $\mathbb{Q}$. That's the definition of a limit point, so we conclude that every irrational number is a limit point of $\mathbb{Q}$.

Then the closure of $\mathbb{Q}$ includes all rationals and irrationals: $\mathbb{R}$. That $\mathbb{Q}$ is dense in $\mathbb{R}$ is another way in which infinities are trippy. Between any two irrationals there lies a rational, and by the same logic in the bullet points above between any two rationals lies an irrational. At first glance, relying on our finite-things intuition, this would imply a one-to-one relationship, a rational-irrational-rational-irrational alternation, but we know that there are infinitely more irrationals than rationals—$\aleph_0$ versus the continuum, $\mathfrak{c}$.

- This is an example of how every set has a closure, not just those on a topology's list of open sets. The set $\mathbb{Q}$ is neither open, nor closed, nor clopen, it simply is.
- A set which has a countable dense subset is called *separable*. This use of the word has nothing to do with the uses you will see later in this book, like the separation of components as per Definition 5.1, or the definition of *separation* in Definition 7.12, or the use of separation through the entirety of Chapter 7. The same people who came up with all the weirdness of Topology itself couldn't come up with four different words. We can laugh it off as a running gag.

A fact that really stuck in my head about dense sets is that, on the real line, between any two points in a dense set, there is another point in the set. If there weren't, then call the distance between those two points 2ϵ, and the point in the center has an ϵ-ball around it with no points in the allegedly dense set. Somebody would say something innocuous like *the coffee shops sure are dense in this neighborhood*, and my mind would start spinning out trying to picture thin little coffee bars crammed into the alleys between other bigger coffee shops, and coffee carts set up between the coffee bars, *ad infinitum*.

3.4.1 Interior

We can go from the canonical open set $(0, 1)$ to its closure $[0, 1]$, so it makes sense that we can (usually) go in the other direction:

Definition 3.7. Interior
The interior of a set A is the union of all open sets in A.

For our canonical open set, this definition works: the interior of $[0, 1]$ is $(0, 1)$.

- The closure of the rationals, $\mathbb{Q}$, was $\mathbb{R}$, but the interior of $\mathbb{R}$ is $\mathbb{R}$ itself.
- Meanwhile, given that it consists of isolated points, the interior of $\mathbb{Q}$ is $\emptyset$.

Moving on to the next logical step, the Outer-thirds Set as a subspace topology of the Usual Topology.

- Say there is an ϵ-ball in the set, with width ϵ.
- We constructed the set by removing middle thirds. Before removing anything, the widest interval entirely inside The Set is one, then after the middle-third removal, the widest interval inside The Set is $\frac{1}{3}$, then after the next removal, the widest interval is $\frac{1}{9}$, et cetera.
- At some iteration, the widest interval is less than ϵ, the width of the ϵ-ball from the first bullet point. So, no, any given ϵ-ball can't be entirely in the set.

Then, the interior of the Outer-thirds Set is empty.

Chapter 4

Continuous

There aren't bridges; the farms in the West Country all have boats that ship their goods across the wide river to the East Country. As an import inspector in the East Country, you know a lot about the West Country. The boats don't all go to a single port, but to whatever dock across the river is convenient, and you've observed that they generally travel in a straight line East. Though, sometimes everybody in an area at the West all converge on a single dock in the East. You've observed that deliveries in the southern areas generally lean toward coffee while goods in the northern areas generally lean toward wheat, though there's always a mix of both. From this you naturally conclude that the farms in the West Country grow more coffee in the South and more wheat in the North. You know that any two adjacent docks will have a generally similar coffee-wheat mix. At the main docks where several ferries all arrive at the same spot, you get a survey of what's growing in a wider part of the West Country, though you still expect the arrivals to this dock will look somewhat like the arrivals at the next dock over where fewer shipments arrive.

One season you find that one dock is receiving all coffee, and the next dock to the immediate north is receiving all wheat. Something happened. You initially suspect there's an anomaly in the river such that not everybody is simply shipping East, plus or minus, but there's some new break in the middle where the coffee-heavy farms are leaning South and the wheat-heavy farms leaning north. After asking around among the shippers, you find that isn't the case, which leaves only one possibility, a change in what is growing: instead of a mix of both crops, the farmers to the South switched to just growing coffee and the farmers to the North are just growing wheat.

Some evenings, as the sun sets over the river, you wonder what it would be like to be an export inspector from the West Country. You'd be able to know what the farms in the West Country are doing with great certainty, but what would you know about the East Country? Just nothing, except that they receive goods from the West.

To this point in our topological journey, we've looked at almost nothing but sets. This is the first time functions are appearing, and we're going to leverage what we know about open sets to ask the same question of functions: when are they smooth? The standard answer from the Real Analysis books is that a function is smooth when a very small movement in the inputs induces a very small movement in the outputs. That answer is about movement and metrics, but maybe it's not about distances at all. Maybe metrics are a canard, and we are better off describing smoothness without them.

But set aside continuity for now and let's turn to the simple concept of a function. I'll discuss three common ways to look at them. The first is as a verb, implying action and movement. If $f(x) = 2x$, we take the input, and we double it. We took action and did a thing to an input.

The second is closer to a table of values—a map, dotted lines from points on the West coast of the river to points on the East coast. Write down the number 3, and next to it write the number 6. The result of the function evaluation is a simple fact of the universe, which doesn't require any active work on our parts. One day, we'll all be gone, lost to a climate so hot clouds don't form anymore, or our Sun expanding in its final stages to consume our planet and then burn out. The future of cold will arrive for us all, and twice 3 will still be 6.

The hardcore computer geeks prefer the second form of function as static map. This perspective is called *functional programming*, which initially confused me because every program uses what are called functions, to print junk to the screen or modify tables or look things up in those same tables. But the functional programmers take functions that execute verbs as impure, and their ideal program is a set of simple functions in the sense of fixed, immutable mappings from inputs to the outputs. By abstracting away (in some ways even denying) the operations the computer undergoes, a swath of compsci annoyances disappear. There are no variables that are set, then change in value in a progression, there is no need to think about the order in which different functions are evaluated, so evaluations can be done in parallel, or whatever order the computer feels is most efficient. With nothing changing state, no temporal aspects to coördinate, there are no unexpected interactions with other functions. In short, time disappears.

The third approach, the one you took as a goods inspector in the East Country, is to view functions as a link or conduit between spaces. If you've had much Linear Algebra, one approach to thinking about matrix multiplication is as a funhouse mirror transformation, where you start with a triangle, multiply, and get a taller or rotated or inverted triangle. This is a book about open sets, so an open set in the pre-function space is connected to some set in the post- space about which we could ask questions (e.g., is it open?), and given an open set in the post-space we can do the inspection to get a sense of what the pre-function set was that transformed into this result (e.g., is it open?). Especially with continuous functions, the nature of what the receiving space sees says much about the nature of the sending space. If there is a discontinuity in what the sending space is putting out, a continuous function will transmit that discontinuity to the receiving space. You don't notice this in Real Analysis because the sending space is always $\mathbb{R}$ or $\mathbb{R}^2$, but you're in Topology country now: what if the sending space is the Outer-thirds Set, or is otherwise full of holes? We can use continuous functions as a tool for learning about spaces.

4.1 Open preimage

The inverse of a function $f(x) = y$, written $f^{-1}(y) = x$, is reading the table in reverse, just swapping the *from* and the *to* columns. If $f(x) = 2x$, then $f^{-1}(x) = \frac{x}{2}$, and if $f(3) = 6$, $f^{-1}(6) = 3$.

The only asymmetry is that we will require that for every x for which the function is defined, $f(x)$ has exactly one value. In the $f(x) = 2x$ example, the reverse direction also had one pre-value for each post-value, but this is not at all guaranteed. The standard example is $f(x) = x^2$: if $f(x) = 4$, we don't know if x is two or negative two. At the extreme, $f(x) = 3$ maps every value to three, and yet that still counts as a function. The inverse is a mess, being that $f^{-1}(x)$ is the empty set for $x \neq 3$, and is all of $\mathbb{R}$ for $x = 3$.

This is where they left it in your Algebra class, but the topologists are a little more loose about it, because they're willing to consider how entire sets get mapped.

- If U is an open set, $f(U)$ is a well-defined concept, the set of values you get when you bundle together $f(x)$ for all $x \in U$. The big shift in perspective is that we're going to use functions that map source points to target points, but we're going to ask questions about how sets map to sets.

- Given U, the forward projection via the function is *the image* of U in the function-space, the set $f(U)$. Given the set $f(U)$, the set that produced it is *the preimage*, U.
- I'll notate a function between arbitrary source and destination spaces with arbitrary topologies as $f : (X, \tau_x) \to (Y, \tau_y)$, and let lower-case x indicate elements of X and lower-case y indicate elements of Y.
- N.b. that some values of y may not have preimages, meaning the set of elements that point to y is $\emptyset$. That means a post-image set U could have bits with no preimage, and it's no problem. If $F^{-1}(y_1) = 3$ and $f^{-1}(y_2) = \emptyset$, then $f^{-1}(\{y_1, y_2\}) = \{3\} \cup \emptyset = \{3\}$.

We're engaging in the study of open sets, so perhaps the most natural question (already asked in a parenthetical above), is whether, given an open set U, $f(U)$ is also open. The topologists prefer to do it the other way around, from the perspective of the receiver looking at the sender, for reasons we'll discuss as we think through the key definition:

Definition 4.1. Continuous
A function $f : (X, \tau_x) \to (Y, \tau_y)$ is continuous iff, for every open set $U \in \tau_y$, $f^{-1}(U) \in \tau_x$ is open.

I drew you a typical-looking curve in Figure 4.1. Along the y-axis, we have a typical-looking open interval marked. Its preimage is the set of x values which have corresponding y-values in the open interval. If the vertical range is an open set, so is the horizontal.

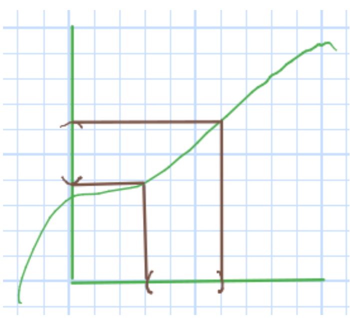


Figure 4.1: The open set along the y-axis has an open set on the horizontal axis as a preimage.

Or let's consider a blatantly discontinuous case, like $f(x) = 0, \forall x < 0$, and $f(x) = 1, \forall x \geq 0$. In the Usual Topology, the open ball $(1 - \epsilon, 1 + \epsilon)$ in the post-image space has a preimage of $x \geq 0$, which is not a u-open set. The problem is that, at the break at $x = 0$, we have to assign $x = 0$ to either the left side (so $f(x) = 0$) or the right (so $f(x) = 1$). Regardless of what we do, we've induced a preimage for that side of the discontinuity which is not open.

4.1.1 The closed version

If an open set U has open preimage $f^{-1}(U)$, then both $\widetilde{U} \equiv Y - U$ and the preimage complement $X - f^{-1}(U)$ are closed. So a continuous function pre-maps closed sets to closed sets. This is too easy a jump to be a theorem, so let's call it a lemma:

Lemma 4.0. *A function $f : (X, \tau_x) \to (Y, \tau_y)$ is continuous iff, for every closed set $U \in \tau_y$, $f^{-1}(U) \in \tau_x$ is closed.*

4.2 Why does the definition go backward?

To answer this question about direction, we need a function that hits the core asymmetry between functions and inverse functions head-on; we'll use $f(x) = 0$. If you have an open ball in the function space that includes zero, then the preimage, the set of x values that map to that open ball, is $(-\infty, \infty)$, which is open by definition. If your open ball does not include zero, then no x values map to it, and $\emptyset$ is also defined to be open. So, our definition passes: for any open U, $f^{-1}(U)$ is open.

But in the forward direction, for any ϵ-ball in the x-space, maybe $(3 - \epsilon, 3 + \epsilon)$, its image is zero, which is a single point, and singleton point-sets are not open in the Usual Topology. That didn't work at all.

For the forward $f(x)$-to-x direction to work, we have two requirements.

Definition 4.2. Injective (one-to-one)
Every y has at most one x such that $f(x) = y$.

That is, there is no ambiguity when we back-trace from the Y space to the X space. To rehash the standard example, $f(x) = x^2$ is not injective—given $f(x) = 4$, was $x = 2$ or $x = -2$?

Definition 4.3. Surjective (onto)
Every y has at least one x such that $f(x) = y$.

That is, the image of X covers the space Y without any holes. For example, if X and Y are $\mathbb{R}$, then $f(x) = \sqrt{x}$ is not surjective—what value of $x \in \mathbb{R}$ gives us $f(x) = -3$?

- It took me a long time to remember which was which. To *inject* as with a needle has a very specific source and target, and you'd draw a needle as a single straight line, not a bunch of lines pointing at a target. The Latin-derived prefix *sur-* is usually interpreted as *onto*, implying the source covering the destination, like a table cloth.

Of course, if you combine the condition that y has at least one corresponding x with the condition that y has at most one corresponding x, you get that y has exactly one corresponding x.

Definition 4.4. Bijective (one-to-one and onto)
Every y has exactly one x such that $f(x) = y$.

And in this case, yes, if $f : X \to Y$ is continuous, then $f^{-1} : Y \to X$ is also continuous. We can define continuity forward again: if f is one-to-one and onto, then it is continuous iff for any open set $U \in X$, $F(U)$ is also open.

4.3 The obfuscated version

Let's compare this with the epsilon-delta definition that they made you learn in Calculus class. Your Calculus textbook assumed a metric right from the start, and we know from §2.2 that ϵ-balls in a metric are the open sets that form a basis for a topology. So rephrasing the above, for every ϵ-ball in the function space U, there is a corresponding ϵ-ball in the preimage space, $f^{-1}(U)$. Figure 4.1 was a diagram of open sets; it's also the diagram for ϵ-balls. If a point $x \in f^{-1}(U) \subset X$, then its post-image is in $U \subset Y$.

- The set-builder form for an ϵ-ball in the post-image space around $f(p)$ is $\{f(x) : |f(x) - f(p)| < \epsilon\}$.
- The ϵ-ball will live in the post-function space, and we will want a corresponding ball in the preimage x-space. To keep things clear, we can refer to the x-space ball with another Greek letter, giving a δ-ball of $\{x : |x - p| < \delta\}$.
- Where before we had said that the preimage $f^{-1}(U)$ of an open set U is an open set, we now say that the preimage of the ϵ-ball matches the x-space δ-ball: δ−ball $= f^{-1}(\epsilon$−ball$)$.
- Moving the function to the other side: $f(\delta$−ball$) = \epsilon$−ball, which we can read to say that if a point is in the δ-ball, then its forward image is in the ϵ-ball: if $|x - p| < \delta$, then $|f(x) - f(p)| < \epsilon$.

With that, we can rephrase the above rephrasing as:

Definition 4.5. Continuous, with a metric
A function $f(\cdot)$ is continuous at a point p iff, for all $\epsilon > 0$, there exists some $\delta > 0$ such that, if $|x - p| < \delta$, then $|f(x) - f(p)| < \epsilon$.

- The awkward $\forall \epsilon > 0, \exists \delta > 0$ clause makes clear that open sets in the post-space have a corresponding open set in the pre-space, but not necessarily vice versa. Note that δ comes before ϵ in the alphabet, hinting at how we're stepping back from post-image to pre-image.

This is quite a muddy way to say *for every open set in the function space, its preimage is open*, but it will sometimes be useful.

Now that we've derived the definition of continuity with a metric, the big lesson might be that the metric only added unnecessary complications. Mathematicians glorify simplicity and parsimony, and the statement *open sets premap to open sets* is decidedly simpler than the ϵ-δ mess of Definition 4.5, but beyond simple æsthetics the universe may be trying to tell us something. Our Calculus teachers taught us to think about smoothness in terms of a metric, where tiny movements on one side lead to tiny movements on the other, but the topological approach advises that the metric might not be the real driver of smoothness. Using the lemma presenting the closed-set version of continuity (Lemma 4.0), it is instead that well-defined sets with limit points always pre-map to well-defined sets with limit points.

4.4 Funky preimages and paths

I want to go back to the anomaly you observed as a goods inspector for the East Country, where you noticed a split between wheat and coffee deliveries where there had been a more continuous range. If you have a continuous mapping from West to East, that doesn't guarantee smooth outputs on the East if there aren't smooth inputs on the West. I took the time to write out the function specification in Definition 4.1, *A function* $f : (X, \tau_x) \to (Y, \tau_y)$ *is continuous iff...*, because the source topology matters: The destination inherits the traits of the source.

- The function $f(x) = 0, \forall x < 0$, and $f(x) = 1, \forall x > 0$ has a hole in it: there is simply no value for $f(0)$. This is a function $f : \mathbb{R} - 0 \to \{0, 1\}$.
- But this is continuous. The preimage $f^{-1}(U)$ if U contains 1 is the open set $(0, \infty)$, and the preimage if U contains 0 is the

> open set $(-\infty, 0)$, and if it contains both the preimage is the open set $\mathbb{R} - 0$, which is the complete preimage space and therefore open—and is open in $\mathbb{R}$ anyway. However you slice it, your preimages are open sets.

So, if you have a disconnected source and destination, you can have a continuous function between them.

If you have a continuous function to a disconnected destination topology, the only way that can work is that the preimage must be a discontinuous topology as well. There's nothing wrong, but this clarifies that continuity is not what gives us our favorite properties, but continuity from a source with our favorite properties.

4.4.1 Paths

But that's not what you expected from continuity. You think of it as representing smooth movement along the x-axis inducing smooth movement along the y-axis. For this to work, you have to start with something smooth along the x. In a sense, the function isn't really the goal at all, but we want to characterize the target space itself as smooth, and if there is a continuous function from a smooth source to the destination, then that means the destination space can inherit the smoothness of the source.

Definition 4.6. Path
Given a topology (X, τ), a path from x_1 to x_2 is a continuous function $f : [0, 1] \to X$, with $f(x_1) = 0$ and $f(x_2) = 1$.

At the beginning of this chapter, time disappeared, but we can nonetheless use the definition of continuity to define smooth motion. At zero, you started somewhere, at one, you ended somewhere else, and because the function is continuous, there are no points where a step in the $[0, 1]$ space corresponds to a sudden jump in the X space.

We have recovered time: given a path in X, you could put a ticker on it starting at zero on one end, then counting up to one at the other end. Part of why this works is that, with $[0, 1] \in \mathbb{R}$ and the Usual Topology in place, the pre-space of the path function is very well-behaved—it's no Outer-thirds Set or Topologist's Sine Curve.

Within the subspace of $[0, 1]$, $[0, 0.2)$ is an open set, so if we have a continuous and bijective function f, then $f([0, 0.2))$ is an open set in whatever the target space is. It could be $\emptyset$, but for the sake of the narrative let's assume it is some nontrivial bunch of points. We've recorded where everything from the west coast in the $[0, 0.2)$

region went to on the east coast; now keep the same breadth but move north by just a little bit and see where goods from $(0.01, 0.21)$ are going. The region $f((0.01, 0.21))$ is some bunch of points, which has to have some overlap with the first bunch, because $f([0, 0.2)) = f([0, 0.01]) \cup f((0.01, 0.2))$ and $f((0.01, 0.21)) = f((0.01, 0.2)) \cup f([0.2, 0.11))$. The bulk of it, the image of $(0.01, 0.2)$, is shared by both images. When we crawl forward just a little on the $[0, 1]$ space, we morph just a little in the target space from one open set to the next mostly overlapping set.

We'll revisit paths in §5.5, where they will be used to define a concept of contiguity (can you draw a line on it?). Paths will be useful to describe smooth transformations from one object to another, and even make a surprise appearance in Definition 7.8 as a means of describing how a set continuously deforms into a point. In Chapter 6, we'll use them to understand spaces which wrap around in odd ways. If you decide to pursue Algebraic Topology, it's all about transformations from one path into another (i.e., a path through the space of paths). In all of these cases, you'll see frequent use of the phrase *there exists a continuous function*, because what the smooth link between the source and destination looks like is not at all important, just that there is one of some sort. The goal is not a function, but getting some of the pleasant properties of $\mathbb{R}$ to rub off on the target topology.

Chapter 5

Connected

There's a funny English expression, "you can't get there from here," which you might hear if you were to ask directions for a point across a river from where you are. *You can't get there from here, you'll have to go back a piece down the road, cross the bridge, then ask somebody there.* Of course, the statement isn't literally true—here on Earth you can get anywhere from anywhere—but it expresses a thinking that here and where you want to go are in distinct sets of points, and that you need to find some connecting point that will join those sets before you can continue.

In mathematical spaces, the question of whether you can get there from here is more difficult, as there are spaces that may be less easy to traverse than the roads and trails covering our planet. You might picture two mathematical islands, like if your space is $(0, 1)$ and $(2, 3)$ and nothing in between, you certainly can't get from 0.5 to 2.5. But 0.9, 0.99, and 0.999 are distinct, discrete points—why are they connected enough to form a line? Why isn't Zeno's turtle stepping across an infinite number of disconnected islands?

There are many ways to approach the question of how individual, discrete points add up to a line or surface; this chapter will cover three. There's the cliché about how when all you have is a hammer, everything is a nail; we know all topologists have is open sets, so all three types of connectedness are going to be defined in terms of how open sets overlap. But expect a little more nuance than simply saying that if two open sets intersect then they are connected. For example, the topology over the space $\{p_1, p_2\}$ with open sets $\tau_2 \equiv \{\{p_1\}, \{p_1, p_2\}, \{p_2\}, \emptyset\}$ will be neither globally connected, nor locally connected, nor path connected.

5.1 Globally connected

The definition of *connected* is so simple and mysterious it deserves to be written as a haiku.

Definition 5.1. Connected

Ignore X and null.
The space is connected iff
No set is clopen.

It's mysterious because there's no immediate relationship between what we think of as connection and the existence of a non-trivial set that is both closed and open. But the link can be made and all elucidated.

Start with the opposite condition, disconnected. If a set S is clopen, it is open and its complement $X-S$ is open, so we're defining the space as splittable iff we can cover it using two non-overlapping open sets.

- Exactly two is not binding: If there exists a covering with a hundred open sets, then the union of the first 99 is also an open set, and that 99-union plus the hundredth open set are a two-set split.

The definition meets the desideratum of doing what we expect with intervals on the real line. If we try to split $(0, 1)$ into two non-overlapping open sets, like $(0, \frac{1}{2})$ and $(\frac{1}{2}, 1)$, we don't quite cover the space, because $\frac{1}{2}$ itself is missing.

- Splitting the composite subset $(0, 1) \cup (2, 3)$ into two open sets is as easy as reading the definition of the subset. You will recall this as the two-island space I used in the intro to this chapter as an example of an obviously disconnected space.
- The half-closed topology (§2.1.7), where sets like $[a, b)$ are τ_{hc}-open, is very much not connected. For any x, the space divides into $(-\infty, x)$ and $[x, \infty)$.
- In the two-points toy example in the intro, with space $\{p_1, p_2\}$ and open sets $\{\{p_1\}, \{p_1, p_2\}, \{p_2\}, \emptyset\}$, both $\{p_1\}$ and $\{p_2\}$ are clopen, so the topology is not connected.
- A function like $f(x) = x^2$ which you can draw without lifting your pen generates a set of points, formally $\{(x, y)|y = x^2, \forall x \in \mathbb{R}\}$, and that set isn't splittable. If you wanted to try disconnecting the space at some value of x, maybe $(-\infty, x)$ and $[x, \infty)$,

it won't work—one of the two sets will be closed and the other open, just as with efforts to split $(0, 1)$ as a standalone space.

5.1.1 Is the Topologist's Sine Curve connected?

Recall the Topologist's Sine Curve, §3.2.1, drawn by the function $y = \sin(1/x)$ plus the origin, $(0, 0)$. As x shrinks to zero, $\sin(1/x)$ vacillates up and down over shorter and shorter intervals of x.

- Notably, y is zero at $1/\pi, 1/2\pi, 1/3\pi, \ldots$.
- For any ϵ-ball around the origin, there is some value of N such that $\frac{1}{N\pi}$ is in that ball.
- That is, there is no way to separate the origin from the over-active sine curve by an open set around the origin. Any open set around the origin allegedly disconnecting it from the main curve will in fact include some portion of the main curve.
- Just as the set of points traced out by $f(x) = x^2$ isn't splittable, the main function $f(x) = \sin(1/x)$ isn't splittable. There's nowhere left to put two clopen sets.

So, the topology is connected. You can't drive a wedge between the parts.

5.1.2 Building a chain

Picture two points, and a chain of open sets connecting the two. Given that people draw both links of a chain and open sets as loose ovals, you'd have something that does indeed look like a chain, with one loop around the first point, then the next oval partly overlapping that first, then another oval partly overlapping the second, and so on until the chain of overlapping loops reaches the other point.

Definition 5.2. Chain
A chain between two points p_a and p_b is a sequence of nontrivial open sets (i.e., excluding X and $\emptyset$) where $p_a \in U_1$, $U_1 \cap U_2 \neq \emptyset$, $U_2 \cap U_3 \neq \emptyset$, ... and for the last set U_ω, we have $p_b \in U_\omega$.

That would be another way to think about connectedness: can we build a chain from any one point to another? The haiku definition at the head of this chapter matches this conception:

Theorem 5.1. Connected iff chained
Iff the topology is connected, for any set of nontrivial open sets which cover the space (i.e., every point is in one of those sets) and any two

points, there is a chain connecting the points using the sets in the covering.

- The theorem said you need to be able to build a chain for *any* covering of the space. If you have a pair of clopen sets which form a covering, then there's a missing link between the two islands.
- If *any* covering has a chain, no covering could be splittable into a pair of clopen sets.

The use of *iff* means this chain concept could be used as a definition of globally connected. The haiku definition—no clopen sets—is so simple and easy to check it seems like a more natural pick of the two definitions, but the chain is perhaps more descriptive of what we want from connectedness, more directly answering the question of whether we can always get there from here.

5.2 The Box and Product topologies

The model here is $\mathbb{R}^2$, where we took one set of $\mathbb{R}$ and called it the horizontal axis, then took another copy of $\mathbb{R}$ and called it the vertical axis, and now we have a two-dimensional space where each point has a two-part coördinate (x, y). Given any two topologies (X, τ_x) and (Y, τ_y), we can generate a new one by the same sort of product, giving a topology whose space $X \times Y$ is all the ordered pairs of the points of X and the points of Y, and whose open sets are all the ordered pairs of the elements τ_X and τ_Y—and their unions and intersections.

Cross products like this are often used to represent multiple types of information. A t-shirt vendor might have a list of snarky messages, which can be printed onto any color of shirt, forming (message, color) pairs. The messages may have relations which put them into groups forming a topology in message-space; the colors are related by the color wheel (which you should bear in mind when you meet $\mathbb{S}^1$ in Chapter 6).

- As per the left side of Figure 5.1, any one cross-product of simple intervals is a fuzzy-edged rectangle, which is why the thing we're on our way to defining is called the Box Topology. But a topology includes all unions and intersections, so the Box topology has abundant open sets made from multiple boxes at haphazard positions.

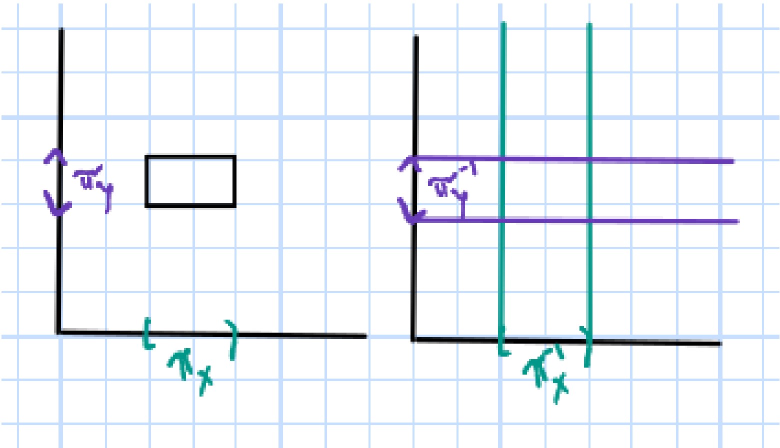


Figure 5.1: Left: A box is the cross-product of two intervals, or read the intervals as the projections π_x and π_y of the box onto the x- and y-axes.
Right: The projection functions have inverses, an infinite column for π_x^{-1} and infinite band for π_y^{-1}. Form a box as the intersection of the two sets.

- I'll stress this more below, but note that we've introduced a new type of operation on sets here. Up to this point, everything has been about intersections and unions, but the cross-product of a set of t-shirt slogans with a set of colors is neither the intersection nor the union of those sets.

And why stop at two? Why not generate something like $\mathbb{R}^3$ by crossing another topology with our pair, giving $(X\times Y\times Z, \tau_x\times\tau_y\times\tau_z)$? And why stop at three? Why not generate an infinite-dimensional topology, be it $\mathbb{R}^\infty$ or an infinite cross product of any of the exotic topologies above? There are very reasonable reasons, covered below.

The idea of this infinite cross product of $\mathbb{R}$, herein The Box Topology, is appealing, though. Given any observation in $\mathbb{R}$ (the temperature, a person's body weight, your attention span in seconds), you could observe it N times over the course of N periods to produce a sequence of observations, the kind of time series you are used to seeing on plots where the x-axis is time, and the y-axis is temperature, weight, whatever. But you could instead write this sequence of observations as values on the first, second, third, . . . dimension, so that all the observations of one person's weight over N periods generates a single point in $\mathbb{R}^N$. Rather than the laborious problem of drawing an infinite sequence as an infinite number of points, you need only draw a single point in $\mathbb{R}^\infty$. Below, we'll work out how to achieve this delightful efficiency, representing an infinite sequence via a single open set.

Topologists get nervous here—Starbird and Su [2019] say some will refer to the Box Topology as $\mathbb{R}_{\text{bad}}$. It is a valid topology, but the key C-words I used for chapter titles in this book (continuous, connected, compact) don't work with an infinite cross-product: the infinite cross product of continuous functions isn't continuous (how do you do an ϵ-δ proof when you have to add up an infinite number of ϵs?); compactness in Chapter 8 will require a finite set of sets (what does that mean in infinite dimensions?); and I'm bringing it up in this chapter because we'll see below that the infinite product of connected sets isn't connected.

Concerned individuals will add the caveat that we can build a cross product of an infinite number of spaces, but for any open set, only a finite number of those spaces may themselves be finite, the others must be infinite—though for different open sets, which dimensions are finite can vary. For example, the open set which is all of $\mathbb{R}$ on dimension one, $(-3, 5)$ on dimension two, $(-2, 6)$ on dimension three, and all of $\mathbb{R}$ on every other dimension is valid in this Product Topology, because only two dimensions are restricted. We'll see this restriction makes things much more well-behaved, perhaps to the point of being more boring.

- You can take the product of anything with anything. If you want to generate a product topology of the Outer-thirds set with itself, feel free—or wait until this happens §10.2.
- When topologists (and myself, below) write of *the* Product Topology, they mean the infinite product of Usual Topologies on $\mathbb{R}$.
- It's always *the* Box Topology because topologists will only use it over the Usual $\mathbb{R}$ as an example of bad things happening in infinite dimensions. In finite dimensions, the box method and the product method are identical—see the following box—and the *product* term is preferred.

Projection functions

I skimmed past the standard way to build the Product Topology the first few times I read it because it sounded pedantic to me.

But it does have a logic to it, which makes the Product Topology seem less arbitrary.

The projection function $\pi_x : \mathbb{R}^n \to \mathbb{R}$ takes a multi-dimensional point and returns just the x coördinate; the projection function π_y similarly returns just the y coördinate. As in the left side of Figure 5.1, the π_x projection function is a lamp you hold over a set in 2-D space, where the shadow cast onto the x-axis is the projection. Point the lamp at an object horizontally, and it casts a shadow π_y onto the y-axis.

Given a point or interval on the real line, like the one on the right side of Figure 5.1, the inverse function π_x^{-1} would preimage that to all points in $\mathbb{R}^2$ with that x-value and any y-value—an infinite column. This would also work in 3-D or 50-D, where the inverse is the set of points entirely unrestricted on all other dimensions but with the given x-values. With a 1-D interval like $(0, 2)$ on the x-axis generating a π_x^{-1} preimage of a two-unit-wide column, and a 1-D interval like $(2.5, 4)$ generating a π_y^{-1} preimage of a 1.5-unit-tall horizontal band, the intersection is a 2×1.5 rectangle. In three dimensions, the other dimension is unrestricted, so we have a rectangle of infinite depth.

We can use this to define all the rectangles, in a manner that follows the track of all our other topologies. Let the basis of our topology be all the π_x^{-1} preimages of open sets and all the π_y^{-1} preimages of open sets. Take the intersection of those open sets to generate the rectangular open sets.

In 2-D, this is just taking the cross product in the manner that didn't require inventing projection functions and inverting them. But now we're on the standard track of specifying an easily described basis, then taking *finite* intersections and infinite unions of open sets on our list.

And that's the distinction between the ad hoc but entirely valid cross-product method that built the Box Topology and this intersection of pre-projections method: if you have an infinite-dimensional space, the *finite* intersection of pre-projections will always be unrestricted on an infinite number of dimensions.

As an aside to another subfield, one way to read Category Theory is that it is an effort to rebuild mathematics entirely via functions. Projection functions work into the project because they're a clever way to define cross products using only functions. Given the functions π_x and π_y, define the unique thing

t for which $\pi_x^{-1}(t) = 3$ and $\pi_y^{-1}(t) = 2$ as the 2-D point $(3, 2)$. The concept of multi-dimensional product is derived from projection functions. So, that's nice.

If (X, τ_x) is connected and (Y, τ_y) is connected, the product topology constructed by the cross product of the two is connected—if you can't drive a wedge to split the X dimension or to split the Y dimension, why would you be able to split one or the other when you cross them. This meets our intuition, but everything will break when we repeat an infinite number of times.

5.2.1 Sequences of reals to boxes

When I picture a convergent sequence, I picture a dotted line on the usual $\mathbb{R}^2$ plane, a sort of time series, starting at zero, then curving upward at a slower and slower pace, until it levels off at some value. A divergent sequence is one which never slows down enough to reach a steady level. The sequence $\{\sqrt{x}|x = 1, 2, 3, \ldots\}$ is divergent, even though its slope gets flatter and flatter as x rises. I can picture taking the slope of a convergent line and gently pushing it upward until it crosses the boundary from just barely convergent to just barely divergent. But Theorem 5.2 below will tell us that that sort of gentle nudging is not possible. There is a fundamental break between sequences which approach an arbitrarily large value and sequences that go to infinity.

Above, I had enthused about the idea of treating an infinite sequence as a single point in $\mathbb{R}^\infty$, with the N^{th} value setting the position on the N^{th} dimension. This is the Usual Topology, though, where singleton points are not open sets, so we have to puff those points up to be open sets. Let o_x be the open set $\left(x - \frac{1}{2}, x + \frac{1}{2}\right)$ on the Usual Topology. If we have a sequence of numbers $x_1, x_2, \ldots$, we can represent that as a single open set by putting o_{x_1} on the first dimension, o_{x_2} on the second, and so on. The cross-product of the first two sets is the interior of a 1×1 square centered at (x_1, x_2); the cross-product of the first three is a cube around (x_1, x_2, x_3); and so on out to infinity. With each infinite sequence a single open set, a multiplicity of infinite sequences-as-sets becomes the basis for a topology, and we have turned questions about the structure of infinite sequences into questions about the structure of a topology.

- The set $(-\infty, \infty)$ is open, so our Box Topology even includes open sets which are infinite on some dimensions; same with $\emptyset$.

The list of open sets which are infinite or $\emptyset$ on some dimension is a well-defined list, which I wall call $\mathcal{I}$, which you might think stands for *infinite*, but below read it as *ignore this part*. Note that The Product Topology is a subset of $\mathcal{I}$. The list of open sets which have a finite maximum and minimum on all dimensions is equally well-defined; it is the complement $\widetilde{\mathcal{I}}$.

- The Product Topology is a strict subset of $\mathcal{I}$, because $\mathcal{I}$ can include sets which are finite on all even dimensions but infinite on all odd dimensions, but the Product Topology doesn't allow those.
- The union of every set in $\mathcal{I}$ is an open set, as is the union of all elements of $\widetilde{\mathcal{I}}$, meaning both sets are clopen and the Box Topology splits between the two.
- We'll rename $\widetilde{\mathcal{I}}$ to $\mathcal{S}$ for *sequential*, because we can read an open set in $\mathcal{S}$ as a one-value-per-dimension infinite sequence: because each dimension has finite upper and lower bounds, there is a well-defined point that is the mean between the upper and lower edges (technically, their closure or outer limit points). For $(0, 1)$ it is $\frac{1}{2}$; for $(0, 3) \cup (19, 20)$ it is $(0 + 20)/2 = 10$.

So $\mathcal{I}$ and $\mathcal{S}$ split the space (and $\mathcal{I}$ further sub-splits into Product Topology and rest-of-$\mathcal{I}$), and every open set in $\mathcal{S}$ has an associated infinite sequence, formed by listing the midpoint on every dimension, though every infinite sequence has infinitely many associated representations. There's a lot going on.

- As discussed above, any given open set within the subset $\mathcal{S}$ of the Box Topology represents an infinite sequence of real numbers, which may be bounded or unbounded.
- The infinite union of all bounded open sets is an open set, call it B; the infinite union of all unbounded open sets is an open set, call it U. Both of these are the union of an infinite number of open sets; that is, B and U are open sets.
- A sequence is either bounded or it isn't; the only third option is the set of sets which allow no sequence at all, $\mathcal{I}$. That is, $B \cup U \cup \mathcal{I}$ covers the Box Topology.

Then we have a split, a handful of open sets which cover the space.

Theorem 5.2. Box-cutting
The set of open sets representing convergent or divergent sequences is not connected.

This was a real revelation to me when I first saw it, reading a textbook in Rock Creek Park one day. I'd thought you could walk from a finite to an infinite sequence; if you could just nudge a finite sequence up enough, you could smoothly cause it to be an infinite sequence. But that's not possible. There's a gap, no guarantee that a chain can link any given bounded sequence to any given unbounded sequence. Because the set of bounded sequence Bs is clopen, it contains its limit points and as per §3.3 any convergent sequence of elements of B converges to something in B, not U. If it seems obvious given the neat division of the topology into subregions of $B \cup U \cup \mathcal{I}$, that means we used the right tool. Imagine proving this via ϵs and δs, where you'd prove that for all metrics and any unbounded sequence $u \in U$, there does not exist some sequence of bounded sequences which converge to u.

- The Box Topology is the cross product of Usual Topologies, and the Usual Topology is connected, so this is the counterexample showing that the infinite product of connected spaces is not necessarily itself connected.
- But the Product Topology, which bans open-set representations of infinite sequences, is connected. Any two open sets intersect, or if they don't, they both intersect another set which is infinite on the dimensions where the sets we want to connect are finite.

5.3 The Intermediate Value Theorem

The name is a call-back to your Calculus class, where this was stated in terms of real-valued functions: if $f(x)$ is continuous on the range (a, b) then it takes on all values between $f(a)$ and $f(b)$. For example, if $f(a) < 0$ and $f(b) > 0$, then there exists some x in this range where $f(x) = 0$. This is one of those things that seems obvious, but when you try to nail it down it's hard to articulate why exactly this has to be the case. I checked the Wikipedia page for The Intermediate Value Theorem while writing this, and the proof using Real Analysis was a wall of ϵs and δs.

Recall that for a function $f : (X, \tau_x) \to (Y, \tau_y)$, continuity means that any open set $U \subset Y$ pre-maps to an open set in the X space, $f^{-1}(U)$.

We want connectedness to carry forward through a continuous function, so if X is connected then Y is as well, and we'll get that, though we have to add one caveat: the function has to be surjective, meaning that every value $y \in Y$ has some $x \in X$ such that $f(x) = y$ (Definition 4.3). One could in fact redefine the target space as the set $Y \equiv \{f(x)|x \in X\}$, gaining surjectivity by construction. Given the addition of that condition, the result that continuity carries connectedness forward makes intuitive sense.

- Surjectivity means that for any $U \subset Y$, $f^{-1}(U)$ is not $\emptyset$. For a proper subset of Y, its preimage also can't be all of X, because it is $f(X)$ which is Y, not some subset thereof.
- There was a lemma (Lemma 4.0) which stated continuity in terms of closed sets: if $f(U)$ is closed the pre-image U is closed.
- Putting it together, for a surjective function, nontrivial clopen sets must pre-map to nontrivial clopen sets.
- So if X has no clopen sets, then Y can't have any clopen sets, because those Y-clopen sets have nothing to preimage back to.

Theorem 5.3. The Intermediate Value Theorem
Given a globally connected topology (X, τ_x) *and a surjective function* $f : (X, \tau_x) \to (Y, \tau_y)$*. Then* (Y, τ_y) *is connected.*

That's the whole game. As an application, if your function is $f : \mathbb{R} \to \mathbb{R}$, and $S \in X$ is a connected subspace $S \equiv [a, b]$, then the theorem tells us $f(S)$ is a connected interval, and if $f(a) < 0$ and $f(b) > 0$, then that closed interval $f(S)$ must include zero.

By thinking in terms of open sets and not individual points, not a single ϵ or δ was harmed. All we needed was that if the X space has no clopen sets, the Y space can't either, and that ϵ-balls form a topology over $\mathbb{R}$. Once again, it's not about doing accounting on minuscule distances, which are a distraction from the easier path of looking for how the connections between points are themselves connected.

5.4 Locally connected

If Topology were Topography, connectedness as defined to this point would determine whether a space was a single landmass or had distinct continents and islands. As per all the discussion about the Box Topology, infinite sequences live on a different landmass from finite. But on a single landmass, we're not guaranteed that you can get there from here. Reasonable people differ on how to define what it means to be connected, so we're going to have a

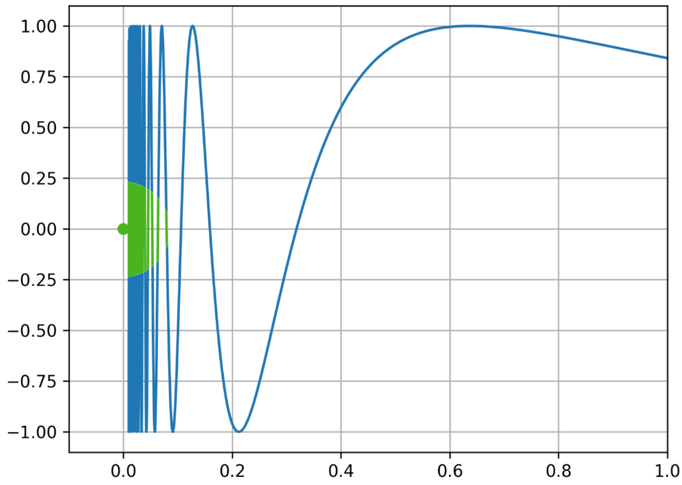


Figure 5.2: An ϵ-ball is a scoop out of the space. Follow the curve and it comes in and out of the scoop.

few definitions of connectedness which you might apply to a single landmass.

Recall from §3 that a *neighborhood* is a set, not necessarily in our approved list of open sets, which contains an open set. All open sets are neighborhoods; not all neighborhoods are open sets.

Definition 5.3. Locally connected at a point
A topology is locally connected at a point x iff every neighborhood of x contains a (Definition/Haiku 5.1) connected open set containing x.

Here, we're taking the subspace topology consisting of the main topology intersected with a neighborhood, and asking whether that subspace is connected. If this condition holds for every point in a space, then we'll call the space locally connected.

- You can't split a Usual ϵ-ball into two clopen sets, and within any set and around any point x in the Usual Topology, there is an ϵ-ball to be had. So, the Usual Topology is locally connected.
- Figure 5.2 shows an ϵ-ball around the origin of the Topologist's Sine Curve. The sine curve oscillates in and out of the ball, so this subset of it can be broken down into segments:

the part that intersects the oscillation centered at $\frac{1}{10\pi}$, the little segment that intersects the oscillation at $\frac{1}{11\pi}$, and similarly for every other oscillation around a small enough x-value.

- Put all points in the ϵ-ball associated with oscillations $\frac{1}{100\pi}$ and smaller into one set; all points in the ϵ-ball associated with oscillations $\frac{1}{99}\pi$ and larger in another set (or use $\frac{1}{1000\pi}$ or $\frac{1}{10000\pi}$ or whatever is appropriate for your ϵ-ball). You've just disconnected the contents of the ϵ-ball into two segments—it's not connected.

Using the topology-as-topography metaphor, maybe we can take the sine curve as a stream. If you had the entire thing, you could sail a canoe along it from any non-origin point to any other, and get arbitrarily close to the origin. But if you were restricted to a ball around the origin, you would have only a walled-off series of streamlets to navigate along, each more-or-less parallel to the other. Where with no restrictions you could have drifted up and around from one oscillation to the next, with the restriction you will need to drag your canoe across regions not in the set to get from one streamlet to the next, possibly an infinite number of times. It's exhausting, and there is no ϵ-ball around the origin small enough that you won't have disconnected streamlets.

Global connectedness is about situations where we can't get a wedge between components, and local connectedness is closer to the micro-scale question of whether a turtle can step from any given side of a set to the other. They're entirely distinct: You just saw that the Topologist's Sine Curve is globally connected but not locally connected, and it's trivial to construct locally connected topologies which are not globally connected: let the space be $(0,1) \cup (2,3)$ and the topology be the Usual Topology restricted to this subset.

5.5 Path connected

Here's one more popular conception of *connected*, which embodies the simple intuition that if something is connected, you could draw lines on it.

Definition 5.4. Path connected
A space X is path connected iff for any two points $x_1, x_2 \in X$, there exists a path, a continuous function $f : [0,1] \to X$ such that $f(0) = x_1$ and $f(1) = x_2$.

You'll recall paths from §4.4.1 in the chapter on continuity. I'd stressed in that chapter that the destination of a continuous function inherits the traits of the source. If the source space is always the completely regular, perfectly normal, globally connected Usual Topology, and it is possible to map a continuous function from that source to a destination, that tells us something about the destination.

Can we draw a line from the origin $(0,0)$ to some point on the rest of the Topologist's Sine Curve? We can restate this in terms of paths: can we find a continuous mapping from $[0,1]$ to the Topologist's Sine Curve, where $f(0)$ is the origin point and $f(1)$ is some point on the main curve? Because $[0,1]$ is part of $\mathbb{R}$, continuity is less free-wheeling than in general topological spaces, looking more like the $\epsilon-\delta$ form I'd derided above: for every output-space region, there exists a small enough input-space interval such that if the inputs are constrained in that region, the outputs are constrained in the original region we started with.

- If you got your ice cream scoop and drew out an ϵ-ball at the origin as in the pistachio-colored segment of Figure 5.2, if $\epsilon < 1$ you would pull out an infinite number of little line segments scooped from the middle of an infinite number of oscillations.
- The tops and bottoms of those oscillations are not in the scoop, though. If you trace the path of the sine curve, it would enter and exit the ϵ-ball at every oscillation.
- Does there exist a δ-ball such that if a point on the preimage $[0,1]$ is within that ball, the value it maps to on the sine curve is always within the ϵ-ball? Not at all, because the curve is oscillating in and out of the pistachio-colored segment.

Above, we saw that we can't drive a wedge between the origin and the curve, that every open set containing the origin must contain some part of the curve, and we concluded that the origin and the curve are connected. But here we see that such connection is still not sufficient to allow us to draw a continuous path.

So how do path, local, and global connection relate?

Path connected isn't enough to give us locally connected. Draw a line from the origin in the Topologist's Sine Curve down to negative two, and wrap it around to connect to the sine curve somewhere around $x = 1$ or $x = 2$. That is, make the curve path-connected by just drawing a path, but go the long way around. It resolves path-connected, but doesn't change anything about the last bullet point above that showed that locally connected doesn't work.

To tie path and local connectedness together, we'll need a slightly stronger version of path connectedness:

Definition 5.5. Locally path connected
A space is *locally path connected* iff it has a basis of open sets where each set in the basis, taken as a subspace, is globally and path connected.

I'll omit further discussion, but this is directly comparable to plain local connectedness.

Theorem 5.4. Locally path connected is stronger than locally connected
If a space is locally path connected, then it is locally connected.

On to path connected versus globally connected. The Intermediate Value Theorem is about functions from one space to another, paths are functions, so we can use the IVT to say things about them.

- If a space is path-connected, that means there's a mapping $f : [0, 1] \to X$.
- The interval $[0, 1]$ is globally connected: it has no nontrivial clopen sets.
- Then the IVT tells us that we can't have $f(0)$ in one clopen set and $f(1)$ in another, because those clopen sets would have clopen preimages in $[0, 1]$. So no clopen sets in $[0, 1]$, no clopen sets in X, and X is as globally connected as $[0, 1]$ is.

Theorem 5.5. Path-connected is stronger than globally connected
If a space is path connected, then it is globally connected.

That means that if something isn't globally connected, there are missing paths. Back on our infinite sequences as open sets on the Box Topology, can we draw a path connecting an open set representing a divergent infinite sequence to one representing a convergent one? No, because they are on separate components of a disconnected space, so Theorem 5.5 tells us all our efforts at finding a connecting path will fail.

Before we digress away from these conceptions of connected, let's try one last example, to keep us all uneasy about all this. Recall that the Outer-thirds Topology is what you get when you read a binary number as trinary, then change all the ones to twos. If we annotate reading a base n number as base m as $P_{n,m}$ (the *type-punning function*), then the mapping from $[0, 1]$ to Outer-thirds Set is $f_1(x) \equiv 2P_{2,3}(x)$, and the mapping in reverse is $f_2(x) \equiv P_{3,2}(x/2)$. Both the base-punning step and the multiplication (by two or $\frac{1}{2}$) steps are continuous, so both composite functions are continuous. A continuous mapping from $[0, 1]$ to the Outer-thirds Set: that's what we call a path.

In the language here, the Outer-thirds Set is path-connected, and even locally path-connected. It is therefore globally and locally connected. Yet, as shown in §3.1.2, between any two points in The Set lie points outside of it. The *points* are dispersed, but the *sets* are connected.

Chapter 6

Manifolds

As a warm-up, let's invent the circle.

The trick upon which this chapter is based is to declare two points A and B to be topologically equivalent, in the sense that if point A is in any given open set, then point B is also in that set, and vice versa.

Start with the line segment $[0,1]$ and the Usual Topology restricted to that subset; i.e., take all the u-open sets, and intersect them with $[0,1]$, meaning that sets like $(.9,1]$ and $[0,.1)$ are open in this topology. Now, paste: declare zero and one to be equivalent, meaning that if a set contains 0 then it also contains 1, and vice versa. Now you've got a circle, herein the one-dimensional sphere $\mathbb{S}^1$.

- Because $(.9,1]$ and $[0,.1)$ are open in this topology, their union is also an open set.
- This was the case before, but with the pasting, that open set is also connected and path connected. You can draw a line from 0.9 to 0.1 without picking up your pencil, because your pencil teleports from 1 to 0 at the appropriate moment.

There's a clear difference between $(0,1)$ by itself and $\mathbb{S}^1$. For example, if you remove a single point from $(0,1)$, it becomes (globally and path) disconnected, whereas you could still draw a path from any point to any point after removing one point from $\mathbb{S}^1$.

Having warmed up in 1-D, let's step up a dimension. Find yourself a rectangular sheet of paper, and roll it into a cylinder by touching the left edge against the right. This is the same pasting operation, making points on the left edge topologically equivalent to the points they touched on the right.

Setting aside the thickness of the paper, this is a two-dimensional surface, turned into a shape we need three dimensions to present. Your topologists would describe this as *an embedding* of the 2-D space into a 3-D space.

The surface of the Earth looks flat enough that we treat it as such in our day-to-day affairs, but it eventually curves enough that the globe is expressible only in three dimensions. Perhaps 3-D space itself curves enough that it is best expressed in a higher-dimensional space. A parabola (the swoop drawn out by $y = x^2$) is an embedding of a 1-D line into two dimensions. If you lived on that line (authors will refer here either to Edwin A Abbot's *Flatland* [Abbott, 1884] or ants with limited vision walking the line), you would see a line like any other, but an outside observer can see that it curves. Topologically, the parabola and a straight line are equivalent: you can draw a one-to-one mapping between open sets on the parabola and open sets on the plain number line.

Are there 1-D embeddings in 2-D space that are not topologically equivalent to a line, and can we get a full catalog of them? What about 2-D embeddings in 3-D?

The discussion in this chapter is sometimes affectionately referred to as *Rubber-sheet Topology*, in which we picture deforming an infinitely stretchy sheet. It will feel very different from the rest of the book, because most of this book is about understanding infinite sets and how they turn into smooth lines and surfaces, and here we assume from the start a well-behaved, path-connected topology over $\mathbb{R}^2$. It is the gateway to *Algebraic Topology*, which pushes hard on the idea that we can map concepts from the Abstract Algebra section of the library onto shapes made from rubber sheets.

6.1 Objects and an operation

So far we have two topologically distinct classes: things which are circles, possibly curved in whatever rubber-band shape you would like to bend them into, and lines, possibly noodled into arbitrary shapes. Is there anything else we can do? For example, a figure eight has a crossing point in the center which seems topologically distinctive.

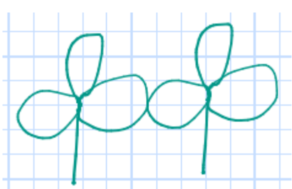


Figure 6.1: Two cloverleafs, or $(\mathbb{L}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1) \vee (\mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{L}^1)$.

Given what we have, we can construct a figure eight by pasting together two circles

at a single point; technically, selecting two
single points, one in each, and declaring that any open set which contains one contains the other. You might even have a cloverleaf as in the left half of Figure 6.1 where several loops all meet at the same central point, so you're picking one point from each of your clover leaves and pasting those together.

That's it, we've explored the space.

- We have two objects, the line segment $(0, 1)$, which we'll write as $\mathbb{L}^1$ for now, and $\mathbb{S}^1$.
- Write the pasting operation as $\vee$, the *wedge sum*. If we put several in parens, we're joining all the elements at a single point. If we chain several together not in parens, each pasting will be at a different point.
- Then a three-leaf clover is $(\mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1)$. If you want to include the clover's stem, you're also pasting a line to the central point, $(\mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{L}^1)$.
- If you want to draw two adjacent clovers without picking up your pen, you would be pasting together two of those objects: $(\mathbb{L}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1) \vee (\mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{S}^1 \vee \mathbb{L}^1)$, where we take as understood that the pasting across objects happens between the last subpart in the first and the first sub-part of the second.

It's already starting to look like an algebra, with a fixed number of core objects (the line $\mathbb{L}^1$ and circle $\mathbb{S}^1$) and an operator ($\vee$) with which an infinitude of other objects can be constructed.

- The typeface used for the reals $\mathbb{R}$ and rationals $\mathbb{Q}$ is called *blackboard bold*, and it is generally reserved for spaces (almost always with the Usual Topology). So the notation $\mathbb{S}^1$ tells us that we should be thinking of the circle (the one-dimensional sphere) and other objects in blackboard bold below as a space in which mathematical objects will live.

6.2 M-band and friends

If you've gotten this far in life as a mathematically curious person without making an M-band (Möbius), you owe it to yourself to experience it. Here is the only construction in this book you could do with construction paper.

1. Cut a thin rectangular strip from a piece of junk mail, maybe 3cm wide by 15cm long.

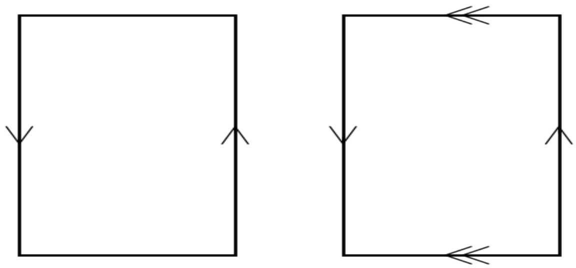


Figure 6.2: Left (easy): make an M-band by twisting a paper as pictured until the arrows line up.
Right (physically impossible): make a K-bottle by lining up the double-arrows on top to the double-arrows on the bottom to form a tube, then align the left-edge and right-edge arrows.

2. As with the cylinder you imagined making at the head of this chapter, take one end of the strip and tape it to the other. But before taping them together, give one end a half-flip, so the bottom of one side is taped to the top of the other side, and vice-versa.
3. Get a pen and draw a line down the middle of the strip, and keep going until you make a full round-trip where the end of your line meets the beginning. You'll find that the line covers both sides of the paper—all sides of the paper are now path-connected.

You used tape in step two, but we called it pasting above, the process of making two points topologically equivalent.

The left side of Figure 6.2 is a schematic of what you did to make an M-band. The arrows are edges that you want to match, and indicate that you have to give the page a twist to get the arrows to align.

The K-bottle (Klein) adds only a metaphorical twist: the right schematic of Figure 6.2 also links the top and bottom elements, with no actual twist, just touching the edges together as with a cylinder. The cylinder of course has two circles on either end. To keep track of the orientation of the circles, picture arrows pointing outward from the cylinder. The next pasting step is to put the now-circular edges together, in a manner such that the arrows are pointing in the same direction. This is not actually possible with an actual 3-D rubber sheet—we'll use paths to explain why below.

People who draw K-bottles usually draw one end of the tube passing through the wall of the other end so that the circles touch. Marar [2022], who covers these cuttings and foldings in great detail, points out that the K-bottle is easily folded without overlaps by moving through a fourth spatial dimension; this is nice to know but will not help you to make one.

- The K-bottle differs from the M-band in that it is closed, meaning here that there is no open edge like the top and bottom of the strip and an imaginary ant on the surface can always go in any direction. Note that this usage of closed has nothing to do with the complement-of-open definition used through the rest of this book, and does not imply that you could fill the K-bottle or the projective plane (see below) with water.

But we don't need to make these things to envision what the paths on them look like: just draw lines on the diagrams, following them around the edges as necessary.

- If you have a path that starts in the *upper* center of either square and goes straight to the right, when it reaches the edge it will wrap around to the *lower* left. Keep going straight, and the path arrives at the lower center. Symmetrically, if you keep drawing the line to the right, it will exit at right toward the bottom, and enter at left toward the top, until finally reaching the starting point.
- If you had a path going straight to the right from the *center* and extended it through the wrap-around, it would return to the starting point. Note that 2-D space has no concept of front or back, though on the M-band you made from your junk mail, this one wrap-around brings you to the back of the paper. You'll have to imagine a paper one atom thick, or a transparent sheet where you don't care if you are marking the front or the back, or use a marker that bleeds through the page.
- Topologists like to imagine ants walking on spaces the ants don't know are weird. Maybe google M.C. Escher's 1963 *Möbius Strip II*, which is a woodcut of ants on an M-band. Here, our animated ant will be a soldier, carrying a spear to protect the queen. In the opening image of our sequence, we draw an ant standing in the center of the M-band page, at the start of the path from the second bullet point above, with the tip of his spear at the start of the upper path from the first bullet point.

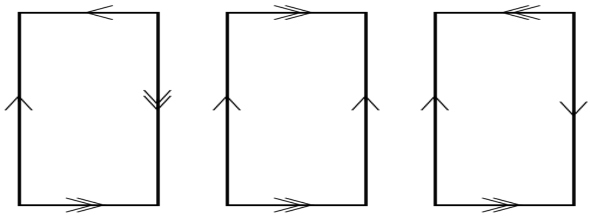


Figure 6.3: Pasting diagrams. Match single arrows to single arrows, and double arrows to double arrows. Left: A sphere. Center: A torus. Right: A projective plane.

As our animation continues, the ant walks the path, taking the middle road, and wraps around once to find himself at the same spot. His spear wrapped around to find itself at the bottom of the page. The ant's spear flipped.

- But on the cylinder, where you just rolled your paper so the edges touch, the ant's tour is unremarkable: there is an *up* and the ant's spear points to it no matter how often the ant walks around the edges.

A surface where *up* is well-defined is called *orientable*.

Including the K-bottle from Figure 6.2, Figure 6.3 shows all possible closed surfaces you could make with matching the edges of a square.

- The diagram for a sphere ($\mathbb{S}^2$) at the left of Figure 6.3 is also how you make a calzone: take a square and fold along the diagonal so left and top match, right and bottom match, and there is room in the center for tomato sauce.
- A torus ($\mathbb{T}$), the doughnut that is the inanimate, edible mascot of Topology, involves pasting the top and bottom edges of a cylinder together, which is hard to do with paper unless you flatten the cylinder, though you could perhaps picture a glassblower wilting a glass cylinder so far its top can be sewn to its bottom, or a pipemaking machine bending a long pipe until the ends can be welded together. The pasting diagram, however, is easy; see middle of Figure 6.3.
- The third diagram in Figure 6.3 is called the projective plane ($\mathbb{P}$), though descriptions of it as a projection sometimes fall flat. But let's build a spinner: put a long stick held down by a thumb tack in the center of the space, and let the two

opposite ends of the spinner extend to the edges of the space. Identify the two edge points under the spinner as topologically indistinguishable. You could do this with the planar representation of the M-band, but only when the spinner is pointing to the East-West edges. The projective plane expands to the full 360° range.

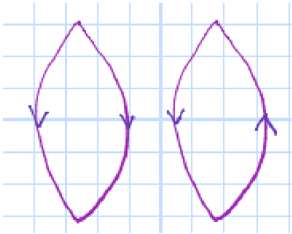
- You can take my word for it that the four closed solids (K-bottle, sphere, torus, projective plane) are the only way to draw a square with two pairs of matching edges, or you can check for yourself. If the North side is a reference arrow always facing right, there are only so many ways to assign to the other three sides one arrow matching the top and two matching each other. If the East side points down and is supposed to be pasted with the North side, there is no way to handle the Northeast corner, so options with those sorts of collisions are out. Other options are reflections or otherwise re-drawings of our four closed solids.

This also brings us back to a thread we had back when we were drawing shamrocks with a single line. There, we found a complete list of 1-D manifolds embeddable in 2-D space: circles, lines, and the wedge sum formed by pasting them together. In 3-D, we have:

Theorem 6.1. The Classification Theorem
Any closed shape can be constructed by pasting together spheres, tori, and projective planes.

You'll notice the K-bottle is in our diagrams but isn't on that list—it can be constructed from projective planes. To make that happen, there need to be a few operations by which the sort of diagrams in Figure 6.3 can be manipulated.

- For the sphere, note that the East and North edges have the same pattern as the West and South edges. You could merge East and North into one long edge without a corner, and do the same with West and South,



  and now have only one edge to connect to one other edge, more a dumpling than a calzone. See doodle.
- The same could be done with the projective plane. This is the space where you pinned a two-arrow spinner in the center, and identified the points the spinner pointed to as equal. You don't need four edges to make that work. Just reverse the arrow on one side of the dumpling you used to represent the sphere.
- Mirror-flipping a diagram so left is on right and vice versa, or top swaps with bottom, doesn't change anything. Rotating the diagram does nothing.
- If you draw a diagonal line down the middle of one of the diagrams, you could cut the square into two triangles. That creates two edges that could be re-pasted together later, just like we've been doing with the box edges. In both cases, we are specifying topological equivalencies in open sets that straddle the edge.

Figure 6.4 uses these various transformations beginning with a K-bottle, exactly as in Figure 6.3 but with letters labelling the edges which are to be pasted together.

- The custom is to read counterclockwise, and if the arrow is instead going clockwise, to write that edge as an inverse of sorts, like a^{-1}. For example, in the two-sided doodles above, $\mathbb{S}^2 = a^{-1}a$ and $\mathbb{P} = aa$
- Figure 6.4 starts with the K-bottle, $bab^{-1}a$, then splits the square into two triangles by naming the diagonal c, giving $ac^{-1}b^{-1}$ and bac.
- It's OK to reflect the first of those, generating bca^{-1}, and note that you can start anywhere in the loop, so you could rewrite the second triangle bac as acb.
- In the diagram, I repositioned the lower triangle to the top so you can see what parts are about to get matched together. In

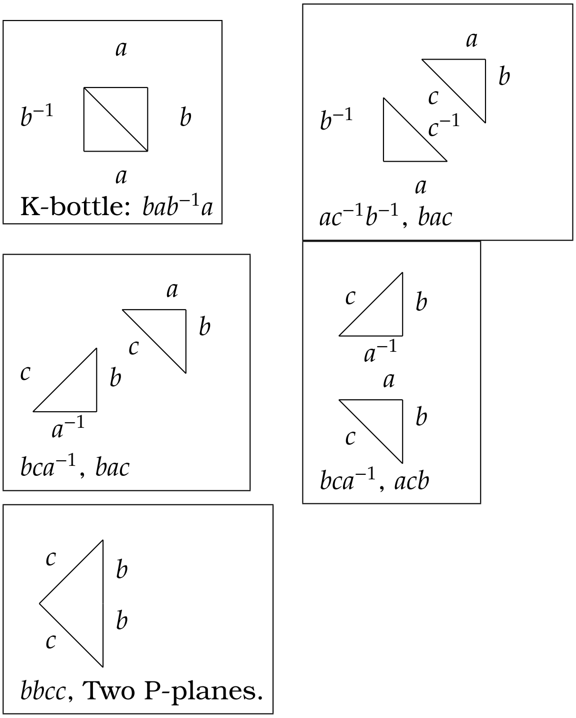


Figure 6.4: K-bottle to projective planes: (1) start with the K-bottle, cut along the diagonal. (2) Split into two triangles. (3) Reflect one. (4) Reposition the triangles to clarify the next step. (5) Merge.

text, we have bca^{-1} and acb, and when we paste them together, the a and a^{-1} cancel out, giving $bccb$ or, rotating where we start reading the cycle, $bbcc$.

- The shape of the projective plane is simply bb or cc, so this is two projective planes merged onto each other.

We're on our way to having an algebra of shapes, where we start with a sphere ($a^{-1}a$), torus ($abab$), and projective plane (aa), and have a list of expansions, reflections, and other transformations that allow us to connect-sum those atoms into whatever shape we may wish.

Why did we only work with squares and two-sided dumplings? What about hexagons and other odd shapes? All of them can be split into triangles, and then those triangles reconstructed using the shapes we've already covered, or by pasting together such triangulated shapes using the sort of wedge operator we used with $\mathbb{L}^1$ and $\mathbb{S}^1$ above. By setting up a few basic units and a few simple operations on those units, all the other possibilities can be constructed, much like we drew a field of cloverleafs with lines, circles, and the wedge operator.

6.3 Invariants

Drawing paths on our planar diagrams tells us things. The sphere and the torus are orientable: pick an *up* direction, send your ant on any path you draw, and the ant will never turn over as he did on the march around the M-band. The projective plane and K-bottle are non-orientable. By no mere coincidence, the orientable closed spaces can be represented—embedded—in 3-D, while the non-orientable closed spaces can't be.

Drawing paths also allows us to distinguish one diagram from the other. There is a space of paths, call it $\mathcal{P}$, and we can go meta and build paths on the space of paths, a function $f : [0, 1] \to \mathcal{P}$ describing a continuous movie where the first frame is one path, then each subsequent frame is another path until we get to the last. If one path can be continuously transformed into another, they are *homotopic* (where *homo-*=same, *topy*=shape). To be less pedantic, you could picture a loop, a path where the initial point is the end point, as a rubber band, and you can continuously transform the loop by stretching it to different spots.

If you had a loop on the surface of a sphere, you could deform it anywhere, over the poles, around the equator, compressed to circle any spot you want. Back to the planar diagram, you could translate a loop anywhere along the simple plane of Figure 6.2, even around the edges or corners.

For the M-band, you could draw a loop that goes all the way around, meaning the ant walks off the right edge and back around the left to find his original spot, or you could draw a loop where the ant's spear tip goes off the upper right edge, back in at lower left, off the lower right, back in at upper left, and then finally reaching the start. Or you could just draw a little circle as on the sphere. There's no way to rubber-band deform these loops into one another. Our little loop winds around zero times, and the other loops wind around once for the ant and twice for the tip of his spear. It would be no problem to construct a loop that winds around three, four, or infinite times.

The torus has three distinct classes of loop: the ones that are simple circles around a point, the ones that wrap around the top/bottom edges, and the ones that wrap around the left/right edges. On an actual doughnut, you can draw little circles on the icing, you could draw a loop through the hole, or you could draw a loop around the equator of the doughnut. Any of these could wind around once, twice, or a million times.

We can use the classes of loops to classify the spaces that host them, and this is where we finally get to the thing about how a coffee mug and doughnut are in the same class. The paths on the coffee mug also fall into three homotopy classes, depending on whether your path goes around the handle, through the handle, or along the more monotone surface of the rest of the mug.

6.4 Characteristic numbers

This chestnut has been known for centuries (Euler): take a cube, a pyramid, or any other solid polyhedron where each face is a simple polygon, count the vertices V, faces F, and edges E, and you will find that $V + F - E = 2$.

- For a cube, your standard die from a board game,
  - there are four vertices on top and four on bottom for a total of $V = 8$;
  - six faces ($F = 6$); and
  - four edges on top, four on sides, four on bottom for a total of $E = 12$,

  and $8 + 6 - 12 = 2$.
- For a pyramid where all four faces are triangles, in the style of the Dungeons & Dragons 4d:
  - there are three vertices on the bottom triangle and one more on top ($V = 4$)
  - the aforementioned four faces ($F = 4$),
  - and three edges on bottom and three reaching to the top vertex ($E = 6$).

  Then $V + F - E = 4 + 4 - 6 = 2$.
- Let's make the bottom face of the pyramid a square, in the style of the Egyptian pyramids. There are now five faces, eight edges, and five vertices, and $5 + 5 - 8 = 2$.

The way to understand why this works is to flatten the shape to a plane. Put your wireframe polyhedron on a sheet of paper (turn your pyramids point-down) and hang a light bulb over the top face, and trace where all the shadows of the edges land on the page. This will have the same $V + F - E$ tally, as long as we remember to count the space outside the lines as representing the top face.

- Set aside the wire frame shadow you just drew for now and instead let's start simpler: a triangle. Three vertices, three edges, two faces—remember to count the exterior as a face—and $V + F - E = 3 + 2 - 3 = 2$.
- Adjoin a triangle by adding a vertex and connecting to two vertices in the original triangle, giving what will look like a box with a slash through it. One more vertex, one more face, two more edges, so the change in $V+F-E$ is zero (we now have $4 + 3 - 5 = 2$).
- This is reversible: if you take this box with a slash down the diagonal and remove the slash, you've joined two faces into one by removing the edge, and have the same vertex count. That's another zero-change, and the square is $V + F - E = 4 + 2 - 4 = 2$.
- That wire frame projection of your polyhedron can be grown using the above steps, adding triangles as needed, then removing extraneous edges to turn adjoined triangles into squares, hexagons, or whatever your polyhedron faces may be.

Then your polyhedron, which has the same $V + F - E$ tally as the projection, will always have a characteristic value of two—as long as it can be projected onto the plane with a simple light bulb, which means it is a solid polyhedron, which is rubber-sheet deformable into a sphere by softening all the edges.

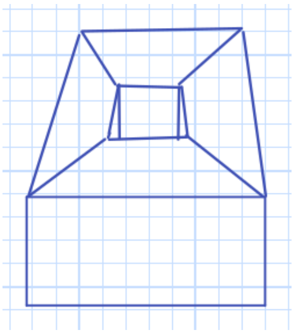


Figure 6.5: The mecha-doughnut.

What happens with a torus? In the doodle of Figure 6.5, I present to you the mecha-doughnut, the toroidal equivalent to the cube, formed by boring a hole through the center of a cube, and updating the top and bottom faces to be four simple trapezoids instead of a ring, because this characteristic number game works only with simple polygons.

- Boring the cube-shaped hole through the center means adding exactly as many edges and vertices as the original cube ($E =$

12, $V = 8$). But it is open-faced, with only the four sides but no top or bottom. That is, with the same V and E as a cube but two F fewer, $V + F - E$ for the hole in the mecha-doughnut is zero, and adding it didn't change the characteristic $V+F-E$ tally.

- You can see that the top face (and infer that the bottom face) is split into four polygons, because our $V+F-E$ game requires simple polygons. No new vertices, a change from one face on top to four, and four new edges, gives a total $V + F - E$ change, including top plus bottom, of -2.

Then the mecha-doughnut's characteristic number is zero, not two. Projecting the mecha-doughnut to the torus means picking a point on one of the faces as the four corners in the plane of Figure 6.3, then cutting the edges. There is no exterior face anymore, though the face that got cut to pieces is now a space around four corners. But from here, the same triangulation game can be played to split faces with extra edges without changing the characteristic number, or remove edges to form differently shaped faces, and the characteristic number won't change from zero.

When doughnuts come out of the oven or the doughnut shop display, they are sometimes twinned. Picture putting one mecha-doughnut by another, producing a super-mecha-doughnut with a figure-eight top and bottom, and where the two side faces put face-to-face to merge the two doughnuts are now part of the interior. That is, F fell by two and V and E stayed the same, giving a characteristic number of -2. This generalizes: the connect-sum of a torus, projective plane, or sphere with another solid shape is the sum of the characteristic numbers minus two for the faces that got eaten. The characteristic number of a polyhedron in the projective plane is 1, and we know that the K-bottle is the sum of two projective planes, so we conclude that the characteristic number of a polyhedron on the K-bottle (a mecha-K-bottle) must be $1 + 1 - 2 = 0$.

We can fact-check by adding something homotopic to a sphere, like a cube or pyramid. If you glom two cubes together you get. . . a long cube, which if made of rubber could be squished down to the shape of the original proper cube. The characteristic number of a cube is two, and when we glom objects together we add the characteristic numbers and subtract two, so we come back to where we started: $2 + 2 - 2 = 2$.

For any C, of course, $C + 2 - 2 = C$, so connecting something homotopic to a sphere cannot change the characteristic number of a shape. If we weld a cube onto the mecha-doughnut, no new holes are formed, and the characteristic number will reflect that.

This is another real victory for the topologists. We have a comprehensive catalog of every 3-D shape, a means of expressing even the ones that can only theoretically be built in real-world 3-D via easily drawn flat surfaces, and we have rules for manipulating those flat representations via simple transformations like rotation, mirroring, and pasting. For tori, we can categorize the shapes we've built by simply counting the number of holes, but for other surfaces, we can rely on the invariant characteristics to understand what sort of surface we're dealing with.

As a reminder and a segue to the next chapter, despite all the talk of mecha-doughnuts and coffee cups, the great majority of this discussion of manifolds assumed nothing but path-connected open sets. When you gather data about some real-world situation, you are gathering a simple grid of numbers, not a beautiful high-dimensional surface, and as we've been doing right from the start, it's up to you to decide how to define the groupings—the open sets—within the data. We don't know what sort of 3-D surface those open sets might cohere to, and there's certainly no reason they have to form a sphere or a doughnut, but we have enough tools to check for the invariants for a given set of open sets and determine the corresponding model.

Though, bear in mind that the key trick of this chapter, making a set of points topologically indistinguishable from each other, was with the aim of forming a cycle. The t-shirt sales in the offhand example from §5.2 included color, which is often described via a color wheel. I've often heard comments that the far-left on the political spectrum talk like the far-right; along with the color wheel, this is an assertion that the slogan $\times$ color space is a torus.

Chapter 7

Separation

The last chapter built a structured world, showing how any closed manifold can be built from a few simple shapes ($\mathbb{S}^2$, $\mathbb{T}$, $\mathbb{P}$) and simple transformations and adjoinings, which is an amazing feat. If you thought that was interesting, this chapter will be even more impressive, because it won't restrict itself to manifolds. The definition of a topology (Definition 2.1) was unbelievably broad, allowing whatever list of open sets you want (plus finite intersections and arbitrary unions), and yet there is decidedly a structure to the boundless space of topologies.

This exercise in classification is interesting, by my vague definition of *interesting* in the intro chapter, in three ways.

First, these conditions nest together. You'll see that they are numbered, from T_0 to T_6, though we have some half-numbers, so we'll have a hierarchy of nine types of topological space. And indeed, it is a hierarchy: if $n > m$, then every topology with the T_n property also has the T_m property. In some cases this will be somewhat obvious, but in others, there is something of a surprise in how one implies the previous.

Second, for these to be different concepts, there has to be something that is, for example, T_2 but is not $T_{2\frac{1}{2}}$. With nine types, there are eight topologies we'd have to construct that fit in one step of the hierarchy, but not the next step up (though I'll skip a few). Some of these will be straightforward, and some will be, to put it lightly, eccentric.

Third, there is a remarkable amount of structure here, given that all we have are open sets and their complements, the closed sets. We gave the topologists dirt to play with, and when we came back they had built infinitely layered sand castles. This chapter is a journey following the construction, from the most basic concept that points are topologically distinguishable to worlds where infinite sequences of sets wrap each other in distinctive patterns. Figure 7.1 shows most of the stops along the way in one column, which may not make sense to you right now, but check back in a few pages and this roadmap will be increasingly legible to you.

All these separation axioms, as they are called, were explored because the only way to understand how points relate, from the perspective of topology, is to look at which open sets contain one but not the other. It's like a surveyor who sees only households and not the residents. Each person has their own name and life, but if all you have is a list of households, if you don't have the equipment to see inside each home and interrogate the individuals, everybody in the household is indistinguishable.

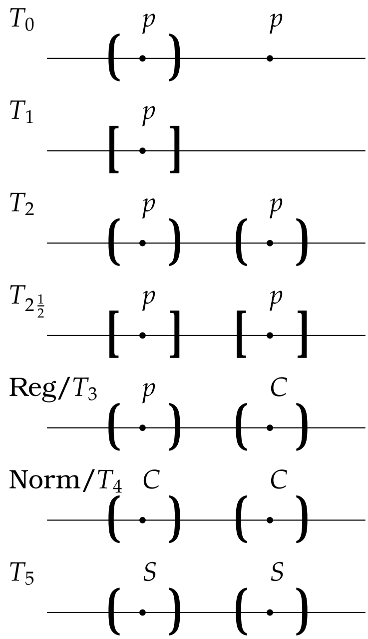


Figure 7.1: Most of the separations, in one column. Separation by open sets are marked by round parens; separation by closed sets marked by square braces. Not shown: $T_{3\frac{1}{2}}$ (separation by functions) and T_6 (normal⟨(C)(C)⟩ plus closed sets are a countable intersection), and some T_1 requirements.

At the extreme on this end, we have the Indiscrete Topology, whose only open sets are $\emptyset$ and the entire space. Topology has collapsed, making every point indistinguishable from every other, and in fact making every space ($\mathbb{R}$, Nicaraguan newspapers, integers evenly divisible be seven) indistinguishable.

At the extreme on the other end of the hierarchy, we'll have the Discrete topology, which assigns to every point in your space its own open set. Topology has collapsed, such as how every one-point set is clopen, so you can disconnect the space into any segments you want (as per Definition 5.1).

Let's see what happens in between these two extremes.

Some logistics about notation before we embark on the journey: The nomenclature in this chapter reads like a sketch comedy caricature of math. One person says *A* $T_{3\frac{1}{2}}$ *topology, being completely regular, is therefore normal and* T_0, and the other replies *Yes, but we still don't know if it's completely normal*, and they both laugh.

If much of math is the invention of new objects and new adjectives to describe them, much of the process of studying math is fluency with vocabulary. I made vocab flash cards when I was reading Category Theory, and I don't know how I could have done it without. But this is a recreational math book, and you will not be tested on this, so I will give you reminders via a tiny text diagram in angle brackets after any non-descriptively named terms, miniature versions of Figure 7.1. In these, p=point, C=closed set, S=arbitrary separated set (by definition 7.12 below), and one or both of the elements will be wrapped in open or closed sets, like a point wrapped in an open set would be drawn as $\langle (p) \rangle$, or two closed sets each wrapped in separate closed sets might be $\langle [C][C] \rangle$. Read $\langle C - f - p \rangle$ as C *is separated from* p *by a function*, which will make sense to you by the time you encounter it.

- Much of this was developed in German, where the word for *separation* is *Trennung*, and thus all the Ts.

7.1 Simple separations: T_0 to T_3

Our journey starts with separations so basic they are almost given.

7.1.1 T_0 (Kolmogorov): asymmetric pairwise

Definition 7.1. T_0 **separation**
A topology is T_0 $\langle (p)p \rangle$ iff any two distinct points in X are topologically distinguishable, meaning there exists some open set which contains one but not the other.

7.1.2 The smallest nontrivial topology

For example, consider a topology whose space is two points, p_1 and p_2, and whose open sets are $\emptyset, \{p_1, p_2\}$, and $\{p1\}$ ("the Sierpinski Space"). The box is exactly checked: there are exactly two distinct points, and there exists exactly one open set containing one but not the other.

- I'd described this as "almost given", which it is in the sense that for any topology, there exists a closely similar topology for which T_0 ⟨(p)p⟩ holds. Say that p_1 and p_2 are topologically indistinguishable: any set that contains one contains the other. Then replace them with a point bundle p_{12}. Repeat, replacing any and all such groups of topologically indistinguishable points with a single point bundle, and you end up with a topology much like the original but satisfying T_0. At the extreme, the Indiscrete Topology reduces to a single point, but that correctly expresses just how blunt that topology is.

7.1.3 T_1 (Fréchet): closed single-point sets

Here's an easy way to distinguish every point with open or closed sets: declare that every singleton set is closed.

Definition 7.2. T_1 **separation**
A topology is T_1 ⟨[p]⟩ iff, for every point $p \in X$, $\{p\}$ is a closed set.

- The example space for T_0 fails to be T_1, because the set $\{p_2\}$ is closed, but the set $\{p_1\}$ is not—its complement isn't on the list of open sets.
- But every topology which is T_1 ⟨[p]⟩ must be T_0 ⟨(p)p⟩, because the complement to the closed single-point set is an open set which distinguishes all other points from that point.

I was surprised by how low down on this hierarchy we find the concept of every single point being closed. Many topological workers insist on working only with T_2 (up next), and because T_2 implies T_1 they take as given that single-point sets are closed.

This is in one sense the end of the line for those toy sets we have over a small set of points. Say the space is $X = \{x_1, \ldots, x_n\}$. The finite union of closed sets is closed, so the union of every point but x_1, $\bigcup_{i=2}^{n}\{x_i\}$, is a closed set, and its complement, $\{x_1\}$ by itself, is open. Same for every other point, and we have a space where every singleton set is clopen. That is:

Theorem 7.1. T_1 **+ finite ⇒ Discrete**
If a topology is over a finite number of points and is T_1 ⟨[p]⟩, then it is the Discrete Topology over that set of points.

This isn't the end of the line in the sense of the journey from T_0 to T_6, as the Discrete Topology satisfies every single one of these properties, but it does so in a sometimes vacuous and dull manner, and so won't be discussed much below.

The topology over graphs in §2.3.1 doesn't even get this far, because the positive integers have a lowest value of one.

- The topology effectively organizes graphs by their number of edges, with a base graph of n elements and the graphs one can form by adding or subtracting an edge, which would have $n+1$ and $n-1$ edges.
- But it all starts with the trivial graph of a single node with zero edges (herein g_0). The only thing you can do with g_0 is add an edge, meaning it is an element only of two sets: the one based on itself, and the one based on the one-edge graph (herein g_1).
- The union of every set but the set consisting of $\{g_0, g_1\}$ covers every graph but g_0, so $\{g_0\}$ is closed.
- But what about $\{g_1\}$? There is no open set containing g_0 but not g_1, so there is no open set containing everything but g_1, so $\{g_1\}$ is not closed.

7.1.4 T_2 (Hausdorff): separated by open sets

To define the next step in the chain, we need a side definition:

Definition 7.3. Separated by open sets
Two points or sets x_1, x_2 are separated by open sets iff $\exists$ two open sets S_1, S_2 where $x_1 \in S_1$ and $x_2 \in S_2$, such that S_1 and S_2 are disjoint (i.e., $S_1 \cap S_2 = \emptyset$).

Definition 7.4. T_2 separation
Any two distinct points are separated by open sets.

As promised, every T_2 $\langle(p)(p)\rangle$ topology is T_1 $\langle[p]\rangle$, meaning that, for any point p, the entire space minus p is always an open set:

- For T_2 to work, we need a topology where every point is in at least one open set—no stray points not tracked by our list of sets.
- For every point that isn't p, T_2 says there is some open set that includes that point but excludes p.
- The union of all of those open sets is an open set covering the entire space minus p. That is, $\widetilde{\{p\}}$ is open, meaning $\{p\}$ itself is closed—that's the definition of T_1 $\langle[p]\rangle$.

7.1.5 Finite-complement topology

The standard example of a set which is T_1 but not T_2 is the finite-complement topology: start with any infinite set of points, $\mathbb{R}$ will do, and define our open sets as any set whose complement is finite. That is, all finite sets are closed—this is sometimes called the *finite-closed topology*.

- Say that the open set S_1 covers everything but the finite list of points P_1, and S_2 covers everything but the finite list of points P_2.
- Then the infinite set of points not in $P_1 \cup P_2$ is in both S_1 and S_2. That is, any two nonempty open sets intersect.
- Our definition of T_2 $\langle(p)(p)\rangle$ requires two open sets which are disjoint, and we don't have *any* such pair of sets.
- Nonetheless, a point consisting of a single point is finite, so our finite-closed topology meets the definition of T_1 $\langle[p]\rangle$ separation.

7.1.6 $T_{2\frac{1}{2}}$: points separated by closed neighborhoods

Definition 7.5. $T_{2\frac{1}{2}}$ **separation**
Any two distinct points are separated by closed neighborhoods $\langle[p][p]\rangle$.

The setup here is that point p_1 is in a closed set C_1 which excludes p_2, while p_2 is in a closed set C_2 which excludes p_1.

Then $\widetilde{C_1}$ includes p_2 but not p_1, and $\widetilde{C_2}$ includes p_1 but not p_2, so those complementary open sets are the open sets you would need for T_2 $\langle(p)(p)\rangle$ separation. But given two open sets separating two points in a T_2 space, we have no guarantee that the closure of one doesn't impinge into the other (or its closure). That's the hook for constructing a topology which is T_2 but not $T_{2\frac{1}{2}}$.

7.1.7 The Negative Origin topology

This is just the real plane, $\mathbb{R}^2$, but with an extra point we'll call the negative origin.

- Outside of the origins, both at $(0,0)$, the open sets are the Usual Topology.
- But the origin's open sets are only the half-ϵ-ball greater than zero—take the usual ϵ-ball around zero and restrict it to strictly above the horizontal axis, but include the origin itself.
- The negative origin has similar half-ϵ-balls, this time strictly below the horizontal axis, plus this negative origin point.

The result is T_2 $\langle(p)(p)\rangle$. There is no controversy that any points besides the origin and negative origin are separable by small enough ϵ-balls. And open sets for those two points contending for being the origin are always disjoint: one always lives above the horizon and one always below.

But if a closed set contains the negative origin, then it is the complement of those positive half-ϵ-balls which do not touch the horizon, meaning the complement includes the horizontal axis. Similarly, the complement to the origin, with its above-the-horizon half-ϵ-balls, must also touch the horizon. The closed sets aren't disjoint, so we don't have $T_{2\frac{1}{2}}$ $\langle[p][p]\rangle$.

7.1.8 T_3: closed sets separated from points

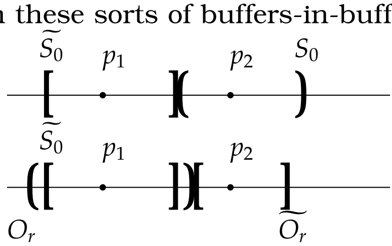
Part of T_3 has its own (meaningless) name:

Definition 7.6. Regular
Given any closed set C and point $p \notin C$, p and C are separated by neighborhoods.

By itself, regular $\langle(p)(C)\rangle$ gets us almost nowhere. It does not guarantee even T_0 $\langle(p)p\rangle$, that any two points are topologically distinguishable. Say you have the C and p separated as in the definition, and p_2 is topologically indistinguishable from p, or in a little text diagram where parens are surrounding open sets, we have $(C)(p_1p_2)$. Then p_2 is also separated from C by the same neighborhood, even though we don't have the point-separating set we need for T_0 $\langle(p)p\rangle$. But together with T_0, we start to go places.

Definition 7.7. T_3 separation
A topology is both regular $\langle(p)(C)\rangle$ and T_0 $\langle(p)p\rangle$.

Here, sets are starting to get layered. If we find a point in a closed set, then regular $\langle(p)(C)\rangle$ will demand that the closed set itself be wrapped in an open set. With these sorts of buffers-in-buffers, $T_{2\frac{1}{2}}$ $\langle[p][p]\rangle$ will follow.



I've made some doodles to help you follow the logic of the bullet points below. Mathematicians get twitchy when proofs are done via diagrams, because there are always implicit assumptions made by diagrams, like how this one is along $\mathbb{R}$, and the closed and open sets all have finite extent. But I'm trusting we can all be adults about this and recognize that these little line diagrams are intended to keep track of what sets are open or closed or contain each other, and nothing more.

- To save you flipping back, T_0 means that for any two points, one (call it p_2) is in an open set the other (p_1) is not in. Call it S_0 because it's guaranteed by T_0; see the first doodle.
- The complement $\widetilde{S_0}$ is a closed set containing p_1 but not p_2.
- Regularity $\langle (p)(C) \rangle$ tells us that that $\widetilde{S_0}$ and p_2 are separated by neighborhoods, meaning there are disjoint open sets O_r and O_{other}, where O_r contains $\widetilde{S_0}$ (and so $p_1 \in \widetilde{S_0}$) and O_{other} contains p_2 (though we'll only need O_r and I left O_{other} off the second doodle).
- The complement $\widetilde{O_r}$ is a closed set which doesn't intersect $\widetilde{S_0}$, and which contains p_2.
- We've arrived at $T_{2\frac{1}{2}}$ $\langle [p][p] \rangle$: our two distinct points are separated by the closed sets $\widetilde{S_0}$ and $\widetilde{O_r}$.

7.1.9 The mushroom-head topology

I owe you an example of a $T_{2\frac{1}{2}}$ $\langle [p][p] \rangle$ topology which is not T_3 $\langle (p)(C) + (p)p \rangle$, so here is one which borrows some ideas from the Negative Origin Topology you just met. The space of the half-disc topology is the horizontal axis of $\mathbb{R}^2$ and everything above it. Picture a mushroom-like open set centered on a point on the horizontal axis, call it x, which is only the intersection of an ϵ-ball around x with the half of $\mathbb{R}^2$ strictly above the horizontal axis. We define the set to include x itself, but the half-ball does not otherwise touch the horizontal axis; see Figure 7.2. Imagine mushroom heads with a stem of exactly one point. The balls freely floating in the upper half of $\mathbb{R}^2$ plus the half-balls along the x-axis are the basis of our topology, then add all the unions and intersections to complete it.

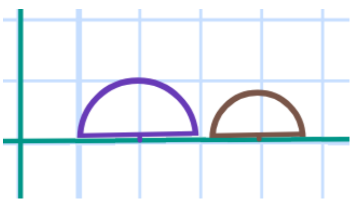


Figure 7.2: The open sets along the x-axis are half-circles that lie just above the axis, but include the single center point.

Are any two distinct points separated by closed neighborhoods, i.e., is this space $T_{2\frac{1}{2}}$ $\langle [p][p] \rangle$? Sure, just find the distance between the two points, and put small enough ϵ-balls around them such that there is some space between the balls, and use the closure of those balls. Having $T_{2\frac{1}{2}}$ of course gives us T_0 $\langle(p)p\rangle$.

But the oddity that a half-ball open set around x includes x but not the rest of the axis breaks the regular $\langle(p)(C)\rangle$ part of the definition of T_3.

- The complement of any given half-ball around x is closed (call the closed complement C), and it includes the axis under the half-ball mushroom head minus x itself. Now we need to invent some sort of open set wrapping x and another wrapping C such that those open sets are disjoint.
- It can't be done. Say you decide that the p-enveloping side of your separation will use a half-ϵ-ball mushroom above x of size ϵ, named M. If you want an open set around C, it has to include nearby points like $x + \frac{\epsilon}{4}$ on the axis. An open set including that point, though, is a mushroom head which has to intersect M.

7.2 Interlude: normal and completely regular

We could keep wrapping closed sets in open sets in closed sets *ad infinitum*, *ad astra*, *ad nauseum*, but the standard hierarchy takes an entirely different approach on its way to $T_{3\frac{1}{2}}$. After this detour, the sets will start wrapping themselves. We know $[0, 1]$ with the Usual Topology is a very well-behaved space (spoiler: it will be all the way at the end of this journey at T_6); the new trick is to find a function linking our unknown, arbitrary topology to $[0, 1]$.

Definition 7.8. Completely regular
Given any closed set C and point $p \notin C$, $\exists$ a continuous function $f : X \to [0, 1] \ni f(C) = 0$ and $f(p) = 1$.

- This looks like a path as per §4.4.1, but it's backward: a path in an arbitrary space X is $f : [0, 1] \to X$, and here we require a function $f : X \to [0, 1]$. The fundamental asymmetry of all functions is that each point in the source maps to exactly one point in the destination, but points in the destination may be the target of any number of source points. Here, the definition of completely regular requires that every point in a set C points to the same destination, zero. I'll refer to these as anti-paths below.

This is the *separated by a function* concept I'd alluded to in the notation section. But *separated* doesn't feel like quite the right word to me; the arrangement is more like a short animation; see Figure 7.3.

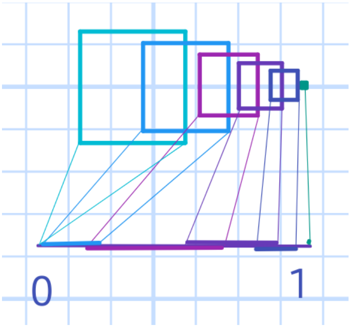


Figure 7.3: A little filmstrip: as time moves from zero to one, the preimage moves from the closed set at left to the point at right.

- Continuity implies a smooth transition from zero to one, and the continuous function implies a smooth sequence which transforms from something including C at $f^{-1}(0) = C$, to some unknown sequence of open sets as the preimage of $f^{-1}([0,0.1))$, $f^{-1}((0.1,0.2))$, and so on until we arrive at open sets around $f^{-1}(1) = p$.
- We're calling this *separation*, and we do have that, as whatever the preimage $f^{-1}([0,0.1))$ may be, it definitely includes $C = f^{-1}([0])$ and excludes p, and whatever the preimage $f^{-1}((0.9,1])$ may be, it includes $p = f^{-1}([1])$ and excludes C.

Thusly, C and p are separated by a function that distinguishes between them, but whose inverse provides something of a path between them.

The real line itself is of course eminently completely regular $\langle C - f - p\rangle$. If we want to find a continuous function for $C = [0,.4]$ and $p = 1$, we could set $f([0,.5)) = 0$, and $f(x) = 2(x - \frac{1}{2})$ for $x \in [\frac{1}{2},1]$. A simple linear function is continuous, $f(C) = 0$, and $f(p) = 1$, and we've checked all the boxes. I know it's arbitrary and artificial, but the rules of the definition let us make up a different function for every pair of C and p, so why not invent whatever function makes it easiest. This is why I'd spent so much of Chapter 4 harping on the perspective of functions as bridges between spaces which we can use as a tool to learn about the spaces. The important part is not the function/bridge, but that such a function is possible at all.

- If you have a metric, define the distance from a point to a closed set as the smallest distance between that point and any point in the set. For a point in the set itself, define the distance to be zero. If we need a function to demonstrate complete regularity for a pair C and p, scale the distance so

$d(C, p) = 1$, and you have your function: $d(C, C) = 0$, $d(C, p) = 1$, smooth continuity in between.

- Nothing in the rules says that the function has to be surjective, (meaning, as per Definition 4.3, that every point in $[0, 1]$ has a defined preimage in X). If $f(1) = 0.2$ and $f(0) = 0.3$, but $f^{-1}(x) = \emptyset, \forall x \notin \{0, 1\}$, then $f^{-1}([.1, .4])$ is well-defined as $\{0, 1\}$—just ignore the dead-air points with a preimage of $\emptyset$.

Here's another example that's one step from trivial: the space is the numbers zero through ten, and the open sets are $\{0, 1\}$ and all the singletons $\{2\}$ through $\{10\}$, plus all the intersections and unions. That is, this is the Discrete Topology over a set of a few points, except that 0 and 1 are topologically indistinguishable. We want to separate the point 10 from the set $\{0, 1\}$. You can verify that $f(x) = x/10$, except $f(1) = 0$, would be a function which meets the requirements.

But even the simple divide-by-ten function is trying too hard, because the space is disconnected, in the sense from Chapter 5 that there are two open sets which cover the space.

- That disconnect can include C on one side and p on the other. For example, $S_1 \equiv \{0, 0.1\}$ is clopen, because $S_2 \equiv \{0.2, 0.3, \dots, 0.9, 1\}$ is clopen.
- Given a split, our function-writing problem is trivial: put the divide-by-ten function out of your head and let $f(x) = 0$ for all points in S_1 and $f(x) = 1$ for all points in S_2. This function is decidedly not surjective, as $f^{-1}(y)$ for any $y \notin \{0, 1\}$ is $\emptyset$.
- Continuity is immediate: for a set in the image space $[0, 1]$,
  - if it includes neither 0 nor 1 its preimage is the open set $\emptyset$,
  - if it includes both its preimage is the open set of the entire space,
  - if it includes 0 but not 1 its preimage is the open set S_1,
  - if it includes 1 but not 0 its preimage is the open set S_2.

Let's wrap that up in a concise statement:

Lemma 7.1. *If a space is globally disconnected (Definition/Haiku 5.1) into two clopen components S_1 and S_2, and $C \in S_1$ and $p \in S_2$, then the function with $f(x) = 0, \forall x \in S_1$ and $f(x) = 1, \forall x \in S_2$ satisfies the requirements for separating C and p by a function.*

- Recall that the Discrete Topology is completely disconnectable, in the sense that for any two points or disjoint sets a and b,

$\exists S_1, S_2 \ni S_1$ and S_2 disconnect the space and $a \in S_1, b \in s_2$. So whatever C and p may be, we know there exists a separating function.

Now that you've met these anti-paths $f : X \to [0, 1]$, we are going to link them to another definition, which will eventually appear in our hierarchy as a part of T_4:

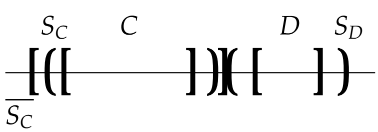
Definition 7.9. Normal
Any two disjoint closed subsets are separated by neighborhoods ⟨(C)(C)⟩.

- The *regular* properties you'll see here are about separation of a point from a set; the *normal* properties separate two sets. The names are meaningless, but at least there's some consistency.

Here's the link, a theorem which I put in this section entitled *interlude*, but whose discussion is really the denouement of the chapter. The sets-within-sets start to generate themselves, like how an electric field creates a magnetic field, which creates an electric field, which creates a magnetic field.

Theorem 7.2. Anti-path iff Normal (Urysohn)
X is normal ⟨(C)(C)⟩ iff, for any two disjoint closed sets C and D in X, $\exists$ a continuous function $f : X \to [0, 1]$ such that $f(C) = 0$ and $f(D) = 1$

- As belabored in Chapter 4, continuity means that open sets in the target have a preimage that is an open set in the source. Then $f^{-1}\left(\left[0, \frac{1}{2}\right)\right)$ is the preimage of an open set in the $[0, 1]$ subspace and so an open set in the X-space containing C (which is the preimage $f^{-1}(0)$ by itself), while $f^{-1}\left(\left(\frac{1}{2}, 1\right]\right)$ is an open set in X-space containing $D = f^{-1}(1)$.
- That checks off the definition of normal⟨(C)(C)⟩: given the function, it hands us two neighborhoods which separate the closed subsets.



But to show that normality implies the existence of this function, we have to find the function. Gillman and Jerison [1960] are jazzed about this theorem and the side quest we'll be taking over the next few pages: "[Theorem 7.2] stands alone as a theorem whose conclusion asserts the existence of a continuous function, but whose hypothesis provides no functions to work from:

in the proof, a function is constructed from 'nothing.' " Before we get to the functional part, let's survey what open and closed sets are available to us. To put it lightly, there are several. This is a long construction, but it builds something incredible.

1. Say you have some clopen set S which includes C and excludes D; then, as per Lemma 7.1 we'd be able to use $f(S) = 0$ and $f(\widetilde{S}) = 1$ as our continuous function and be done with it. So the clopen-set situation is an open-and-shut case. I'm going to assume in all of the below that none of the sets discussed are clopen, because that's the case that isn't yet solved.
2. Normality says that the closed set C is separated from D by some neighborhood (and vice versa), meaning there is some open set S_C such that $C \subseteq S_C$, but $D \cap S_C = \emptyset$; see doodle above. There is a corresponding open set around D, S_D. By the *clopen sets are already handled* rule, we only have to worry about the case where $C \subset S_C$ but $C \neq S_C$.
3. The closure of S_C doesn't intersect D. To be in the closure of S_C, a point $p_d \in D \subseteq S_D$ has to be a limit point, meaning every open set containing p_d intersects S_C. But by normality⟨(C)(C)⟩, we know $S_D \cap S_C = \emptyset$, so S_D is the set which proves that p_d is not a limit point of S_C and therefore not in $\overline{S_C}$.
4. If $\overline{S_C}$ and D are disjoint closed sets, then we are back where we started, when we assumed two disjoint closed sets, and we can apply the normal⟨(C)(C)⟩ condition all over again. Then $\overline{S_C}$ is a strict subset of an open set S'_C, where by the same logic above $\overline{S'_C} \cap D = \emptyset$. Again, I'm using $S_C \subset \overline{S'_C}$ instead of $S_C \subseteq \overline{S'_C}$ by the principle that continuity is handed to us if S'_C were clopen.
5. And with $\overline{S'_C}$ disjoint from D, the logic applies to this pair too, and there is some S''_C wrapping $\overline{S'_C}$.
6. The cycle continues, with an outward-radiating sequence of C, S_C, $\overline{S_C}$, S'_C, $\overline{S'_C}$, S''_C, . . . , where each is a strict subset of the previous.

Let's pause to admire the view: Two disjoint closed sets separated by open sets plus connectedness gave use a countably infinite number of nested open sets—an immense amount of structure to our topology. I didn't even mention that radiating out from D there is *another* countably infinite sequence of nested open sets, then add the unions and intersections of all of those. We dropped two closed-set stones into a still fountain, and waves rippled and reflected everywhere.

Munkres [1999, p 210]: "All this is easy. The only hard part is to show that f is continuous." Like me and everybody else at my college, you may have learned how to show continuity via ϵ-δ proofs, and we've seen a lot of examples of showing that a function is continuous via nice features like disconnectedness, but here we're going to see how topologists would show continuity in a more general case, by constructing all possible open sets in the post-image space and showing that they all have preimages that are open.

1. The open sets radiating from C are countably infinite, the rational numbers $\mathbb{Q}$ are countably infinite, so there is a one-to-one correspondence between the two. For every rational in $[0, 1]$, we can assign, in order, a set in the infinite radiance of open sets from C. How would you do this? I'm not entirely sure, maybe start from $\frac{1}{2}$ and work your way out, but we take it as axiomatic that such an ordering exists (a variant of *the Axiom of Choice*; see 8.3.1). To zero, we will assign S_0 and to one we will assign D, and for any p and q in $[0, 1]$ where $p < q$, the set assigned to p, herein S_p, will be a subset of the set assigned to q, herein S_q.
2. Then, that neat ordering can map to $[0, 1]$ by defining $f(S_p) \equiv [0, p)$. This will work, because if $p < q$, then $[0, p) \subset [0, q)$, and $S_p \subset S_q$. As you radiate outward from the core set C to larger and larger sets, they map to larger and larger intervals in $[0, 1]$.
3. We've got our first step to continuity: for every rational p, an open set $[0, p)$ with an open preimage S_p.
4. What about the irrationals? For any irrational x, there is an increasing infinite series of rationals $p_{x1}, p_{x2}, \ldots$ that converges to it, and that gives us an infinite series of preimages $S_{x1}, S_{x2}, \ldots$. And what do you get when you take the infinite union of open sets? An open set, which we will define as $f^{-1}([0, x))$.
5. Our radiance of nested sets from C alternated closed-open-closed-open, so between the preimage of any pair x and q, there was a closed set nested between S_x and S_q, call it $\overline{S_{xq}}$. What do you get when you take the infinite intersection of all closed sets $\overline{S_{xq}}$ over all $q > x$? A closed set, which fits our hierarchy as $f^{-1}([0, x])$. Then $f^{-1}((x, 1])$ is open.

How are you feeling right now? Because we made it: every open set of the form $[0, a)$ has an open preimage, every open set of the form

$(b, 1]$ has an open preimage, and every open set of any other form like (b, a) is the intersection of open sets like $[0, a)$ and $(b, 1]$, meaning the preimage is the intersection of the preimage of those open sets. We have a continuous function.

I know it's the mecha-doughnuts that get all the attention, but to me, this result is more astonishing. If you had told me that normality⟨(C)(C)⟩ is all you need to produce a continuous function from any arbitrary space with that property to the pleasantly well-behaved open intervals in $[0, 1]$, I would not have believed you. Especially when all I knew about continuity was something about ϵs and δs, the normality assumption seems too lite, not enough information. It's already an impressive feat that (with connectedness) it's sufficient to produce the infinite regress of sets-within-sets, but those are sufficiently well-ordered that they map neatly to the sets-within-sets of the unit interval.

7.3 $T_{3\frac{1}{2}}$ to T_6

The interlude was a real turning point, in that from here on out there is a great deal of structure to topologies conforming to these axioms. I'm going to pick up the pace on this journey a little now that the methods will be familiar, but we are only halfway to the final goal of T_6, where we'll finally arrive at the Usual Topology.

7.3.1 $T_{3\frac{1}{2}}$: Separated by a function

You may have noticed that the above ten-points example of a discrete but completely regular topology above included two topologically indistinguishable points $\{0, 1\}$. Like regular ⟨(p)(C)⟩, completely regular does not give us T_0 ⟨(p)p⟩, but explicitly including it puts us back on the hierarchy:

Definition 7.10. $T_{3\frac{1}{2}}$ **separation (Tychonoff)**
A topology is both completely regular ⟨$C - f - p$⟩ and T_0 ⟨(p)p⟩.

T_3 ⟨(p)(C)+(p)p⟩ was based on regular ⟨(p)(C)⟩, which was just another separation of points and closed sets via open sets enveloping each, while this next step in the hierarchy uses completely regular and its concept of separation via a function. Given the function, we can almost read the sets needed for plain regular directly: if $f(C) = 0$ and $f(p) = 1$, then $f^{-1}([0, .1))$ and $f^{-1}((.9, 1])$ are separated open sets enveloping C and p. So we checked off the two open sets needed for regular ⟨(p)(C)⟩, and added potentially infinitely more.

7.3.2 T_4: T_1 plus neighborhood-separated closed sets.

Theorem 7.2 linked normal $\langle (C)(C) \rangle$ to an anti-path comparable to but not quite the ones used to define regular $\langle (p)(C) \rangle$. So there's still some breathing room between $T_{3\frac{1}{2}}$ and a normal topology, which appears in position T_4:

Definition 7.11. T_4
A topology is T_1 $\langle [p] \rangle$ and normal $\langle (C)(C) \rangle$.

In the next chapter on compactness, Theorem 8.4 will tell you that every compact T_2 space $\langle (p)(p) \rangle$ is normal $\langle (C)(C) \rangle$. Because T_2 is not an especially demanding condition, and compactness will prove to be extremely useful, many commonly used topologies will fall here at T_4 or above.

7.3.3 T_5: separated sets are separated

Here is—I am not making this up—another concept named *separation*, from which is defined another concept named *normal*:

Definition 7.12. Separated
Two (not necessarily open) sets A and B are separated iff the closure $\overline{A}$ does not intersect B, and vice-versa. I.e, $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$.

Definition 7.13. Completely normal
If two sets A and B are separated, there are two disjoint open sets A_o and B_o such that $A \subseteq A_o$ and $B \subseteq B_o$ and $A_o \cap B_o = \emptyset$.

Normal$\langle (C)(C) \rangle$ worked in terms of closed sets, which are a very specific group within the set of all possible sets, while completely normal $\langle (S)(S) \rangle$ works over any two separated sets, no matter how or whether they relate to the list of open sets. Because closed sets are a type of set, normal$\langle (C)(C) \rangle$ is a special case of completely normal $\langle (S)(S) \rangle$.

Definition 7.14. T_5
A topology is T_1 $\langle [p] \rangle$ and completely normal $\langle (S)(S) \rangle$.

Surreal numbers, The plane plus infinity
Surreal numbers, where *sur-* is a prefix meaning *above* or *beyond*, is a term referring to the usual real numbers plus a single additional number assumed to be larger than all others—assume a point called ω (lower-case Ω, read as *omega*), but

which the rest of us might write as ∞. Inverting it, you get something smaller than any real number, as $1/\omega$ becomes analogous to ϵ from the Calculus books. From there, the usual analytic approach to continuity can be applied, and expressions like $2\omega+3$ are treated as valid, and their position in the order of things can be readily established. You can get far by assuming one more point.

Or, imagine drawing a map by setting a transparent globe on top of a paper and shining a light from the North Pole, as you did with the polyhedra whose characteristic number you were checking in §6.4. Pick any other point on the globe, and the North Pole plus that point form a line, which traverses from the North Pole, through that point you picked, and eventually to the page; mark that as the projection of the given point. Points near the South Pole are almost on the page to begin with and the light from the North Pole will project through them directly to the page below; for points near the Equator the line from the North Pole crosses through at 45°; for points close to the North Pole, lines will be just shy of parallel to the plane the globe is resting on, and will eventually reach the plane after some distance. But we can't use this method for the North Pole itself, which would project to somewhere further out than all other points on the globe. So, add one additional point at infinity. Surreal numbers distinguish between +∞ and -∞, but for the projection, the North Pole point is just beyond all directions, positive longitudes and negative.

We wind up with something that behaves much like a circle, (aka $\mathbb{S}^1$), which is a line segment with the ends tied together at some point. The complement to an ϵ-ball around the tie-together point of the circle is a closed interval covering the rest of the circle, and we can use that when considering the topology for ω or the North Pole. Take the complement to a closed interval in $\mathbb{R}$ like $[-100, 100]$, and you have all numbers with absolute value greater than a hundred, including ω. Given that intervals around a point at infinity will necessarily have infinite size, this is a sensible way to define balls around ω.

Finite-complement plus one: Let's modify the finite-complement topology (§7.1.5), in which a set S in a space with an infinite number of points is open iff its complement $\widetilde{S}$ is finite, as per the box. Pick a point NP (North Pole) and we'll add S to the list of open sets if $NP \in \widetilde{S}$ (i.e., $NP \notin S$). So

- any set not involving the North Pole is open,
- and any set with the North Pole but a finite complement is still open.

This is going to be T_4 ⟨[p] + (C)(C)⟩ but not T_5 ⟨[p] + (S)(S)⟩.

- For both the plain finite-complement topology and the version with the North Pole, we get T_1 ⟨[p]⟩ almost for free: a single point is certainly a finite set, so its complement is open, so the point itself is closed. This holds even for the North Pole, because $\widetilde{\{NP\}}$ is both a set with a finite complement and a set which excludes NP.
- The plain finite-complement topology, without the added North Pole glitch, doesn't have any disjoint open sets. Some finite part of set one may be missing from set two, and vice versa, but that leaves an infinite number of points in common (see §7.1.5). Without disjoint open sets, we didn't even get past T_2 ⟨(p)(p)⟩.

Including the North Pole, we have a few possibilities for two disjoint sets S_1 and S_2 to check for separability: one or both can be finite, and one or the other can contain the North Pole. In all cases, we have to find disjoint open sets around S_1 and S_2 (possibly themselves). So let's take a deep breath and check all the cases:

- For T_4 ⟨[p] + (C)(C)⟩ we only have to worry about the case where S_1 and S_2 are both closed, meaning both are either finite, or one is finite and the other contains the North Pole.
- If a set is finite and excludes the North Pole, then it is clopen: open because it excludes NP, closed because it is finite.
- If both S_1 and S_2 are closed because they are finite and exclude the North Pole, then they are their own open sets separating each other.
- If S_1 contains NP and S_2 does not (and so is finite and clopen), $\widetilde{S_2}$ is an open set for which (by separability) $S_1 \subseteq \widetilde{S_2}$ and (by definition) $\widetilde{S_2} \cap S_2 = \emptyset$. That is, S_2 and $\widetilde{S_2}$ are open sets, which form a separation of S_1 and S_2.

That covers all the ways you can have closed sets, so we can check off T_4 ⟨[p] + (C)(C)⟩.

For T_5 ⟨[p]+(S)(S)⟩, we need *any* sets. For example, say that $S_1 = \mathbb{Q}$, and $S_2 = \{NP\} \cup \{x + \sqrt{2} | x \in \mathbb{Q}\}$. They are disjoint, so we need open sets to separate them. Then S_1 is open (because it is missing the North Pole), but any open set containing S_2 has to exclude only a finite number of points because of the finite-complement rule—but S_1 will intersect any set which excludes only a finite number of points.

It just takes one counterexample (though there are infinitely more), and we've eliminated the possibility of T_5.

7.3.4 T_6: perfectly normal.

We top off our hierarchy with two definitions:

Definition 7.15. Perfectly normal
For every closed set C, there is a continuous function such that $f(C) = 0$, and $f(x) \neq 0, \forall x \notin C$.

The definition of *completely regular* (Definition 7.8) involved functions where $f(C) = 0$ for all closed sets, but here we add precision, a sharp edge, that for at least one function, the open set that is the complement to C *always* has a nonzero value. Or, here is another definition of the same term:

Definition 7.16. Perfectly normal
The topology is normal⟨(C)(C)⟩ and every closed set is a countable intersection of open sets.

These two definitions look very different, but the key to seeing why these are equivalent is back in the interlude/denouement of §7.2 in Theorem 7.2, that normality⟨(C)(C)⟩ as in the second definition of perfectly normal implies that for any two disjoint closed sets C and D in your space X, $\exists$ a continuous function $f : X \to [0, 1] \ni f(C) = 0$ and $f(D) = 1$. Now that both definitions imply a continuous function, how does the countable intersection condition give us the sharp-edged version?

- Given normality⟨(C)(C)⟩ and that C is a countable intersection of open sets $S_1, S_2, \dots$, (meaning $C \cap \widetilde{S_1} = \emptyset$, $C \cap \widetilde{S_2} = \emptyset$, $\dots$). Applying Theorem 7.2 here tells us that, $\forall i, \exists f_i : X \to [0, 1] \ni f(\widetilde{S_i}) = 1$ and $f(C) = 0$.
- Sum those f_is together to produce a function which is zero at C and nonzero everywhere else. If you're worried about the sum of a countably infinite number of functions spinning to infinity, take a weighted sum like $\Sigma_i \frac{f_i}{2^i}$.

For any point not in C, one of those functions is nonzero, so we got the precision we want: the composite function is nonzero for anything outside of C.

Definition 7.17. T_6
A topology is T_2 $\langle(p)(p)\rangle$ and perfectly normal.

All the way up here at T_6, we have the Usual Topology. Things are very structured and very—*regular* and *normal* are taken, so we'll say typical and expected.

- The T_2 $\langle(p)(p)\rangle$ condition requires that we find ϵ-balls around any two points p_1 and p_2, but in the Usual Topology ϵ-balls are abundant. We know the distance $d(p_1, p_2)$ between the two, so let $\epsilon = d(p_1, p_2)/4$, and put ϵ-balls of that radius around each to separate them.
- Given a metric, the distance from any point to a closed set (zero for points within the set, shortest distance between the point and the set outside of it) is just what we need for the first definition of perfectly normal.
- The thing about any closed set being a countable intersection of open sets is also new. We just met a few topologies which were an infinite space like $\mathbb{R}$ plus one extra point at infinity, ω. Let the open sets be the Usual open sets, excluding ω (and that's it, no other classes of open sets). Being a single point in a T_1 $\langle[p]\rangle$ space, $\{\omega\}$ is closed, and is the complement to an infinite *union* of open spaces, $\bigcup_{i=1}^{\infty}(-i, i)$. But the intersection of two sets is never larger than the original sets, so there is no intersection of Usual open sets that gives us the closed set $\{\omega\}$—Definition 7.16 isn't met and so we don't have T_6.

We've reached the ninth stage of our voyage. I want to call it the summit, but being the home of the Usual Topology, this is where most people live, and most people never consider that the other stages exist. I had to present to you the story beginning at T_0 and ending at T_6 because that's the direction in which the concepts build upon each other, but it's interesting to think about the story in the other direction: start at T_6, a heavily structured world, and at each stage lift some restrictions and allow more topologies to flourish—less a hike to the summit and more a cyberpunk story where the surface-dwellers are all predictable rule-followers and the further down you go the more unexpected and varied everything and everybody becomes.

Before I knew that such a hierarchy existed, I hadn't imagined that there would even be a hierarchy; now that I know about it, my first question about any topology is where it falls on the T-scale. But don't forget that there are other characteristics of a space and its topology which aren't captured by what sets are contained in what other sets. We can find examples up and down the T-hierarchy of sets which are connected or disconnected, by any of our three definitions, or dense or not. The Outer-thirds Topology is totally disconnected, in the sense that between any two points in the set there are points outside of it (i.e., there is a number with a trinary representation with a one between any two numbers in The Set). It has measure zero and, as we'll see in §10.2, dimension 0.631, but is also T_6. Once you know it is T_6, you immediately know all of the other facts on the hierarchy apply, from T_1 $\langle [p] \rangle$ through T_5 $\langle [p] + (S)(S) \rangle$.

Chapter 8

Compact

Imagine finding yourself in an ill-equipped cabin in the woods somewhere, and you find that you have nothing but a bunch of hand towels for bedding. You carefully arrange them so that every part of the surface of your body is covered by one towel or another, each towel only slightly overlapping those adjacent to it. Or for a less chilly image, when you were a child, one big blanket would be sufficient to cover you completely. Though they are not warm-blooded, we can do the same with sets, covering one set with a collection of other sets—possibly an infinite number of them.

Definition 8.1. Open covering of X
A set of open sets C such that X is a subset of the union of the sets in C. If a subset of C is sufficient to cover X, we call the subset a subcovering.

The open covering may have an infinite number of sets, which engenders the key definition for which this chapter was named:

Definition 8.2. Compact
A set is compact iff, for every open covering, there exists a finite open subcovering.

You want an example to clarify how this works, but first I'll give you some motivating context. Compactness does fit in with the main theme of the book, as convergent infinite sequences and compact sets go hand-in-hand. Given an infinite sequence of real numbers, the creative topologist will respond with an infinite sequence of open sets capturing the meaning and intent of that sequence, and compactness is a key test of whether that sequence represents

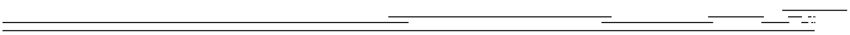


Figure 8.1: Zeno's covering: the second and third layers are a sequence of open sets, each just over half as large as the previous, but slightly overlapping. Together, they cover the baseline $(0, 1)$ interval, but not 1 itself. To cover that, we need one more interval, at top right, but any interval at 1 makes an infinite number of tiny sets redundant.

something convergent. It is a staple of basic Topology for probably more pedestrian reasons: everything is easier when things are compact, and I'll show you how compactness allows some leaps of logic that would otherwise be impossible. But if you'd rather talk about infinities of infinities, recall the infinite-dimensional spaces from §5.2. Can an infinite covering in an infinite-dimensional space have a finite subcovering? The answer is yes—under certain conditions, which give us an excuse to look at the Axiom of Choice.

On to the motivating example. The core intuition is about constructing a covering of the canonical open set $(0, 1)$ plus 1:

1. The set we want to cover, the base line in Figure 8.1, is $(0, 1]$ in the Usual Topology.
2. Define $\wr_x$ as a small offset, $\frac{0.1}{x}$.
3. With the offset, $(0, \frac{1}{2})$ and $(\frac{1}{2} - \wr_4, \frac{3}{4})$ just overlap, as we shifted the right endpoint of the first set by 10% to set the left endpoint of the second. So these two are a valid covering of $(0, \frac{3}{4})$.
4. The covering can be extended to embody Zeno's paradox, by a sequence of open sets where each is, apart from the overlap, half as big as the prior:
$$\left(0, \frac{1}{2}\right), \left(\frac{1}{2} - \wr_4, \frac{3}{4}\right), \left(\frac{3}{4} - \wr_8, \frac{7}{8}\right), \left(\frac{7}{8} - \wr_{16}, \frac{15}{16}\right), \ldots.$$
But the union of this infinite sequence of sets is $(0, 1)$, just missing the $(0, 1]$ we are trying to cover.
5. So we are going to need one more open set of which 1 is an element. To complete the open covering, we need one such set, an interval of the form $(1 - \epsilon, x)$, where x indicates that we don't care what the upper bound is.
6. Adding that last set to Zeno's covering gives us an infinite set of open sets which fully covers $(0, 1]$.
7. Figure 8.1 displays twenty-three overlapping intervals out of the infinite sequence that would cover $(0, 1)$, and one interval

to cover $\{1\}$ itself (depending on the resolution of your document viewer). That one interval hangs over most of the smaller intervals at the end of the sequence of open sets.

Our open covering is constructed, but has redundancies. Whatever ϵ may be, there is some value in the sequence $\frac{1}{2} - \wr_4$, $\frac{3}{4} - \wr_8$, $\frac{7}{8} - \wr_{16}, \ldots$ which is greater than $1-\epsilon$. Call that point b. All subsequent values in the sequence are also greater than $1-\epsilon$. The infinite list of open intervals going forward from b can be removed from our infinite open covering and the set $(0, 1]$ would still be covered.

That is, Zeno's covering plus that one last open interval $(1-\epsilon, x)$ is an infinite open cover, but if you retain only the finite set of open sets up to b and the set $(1-\epsilon, x)$, you still have an open cover, even though you kept only a finite number of open elements.

You couldn't do this with the open interval $(0, 1)$. Zeno's covering is an infinite open cover, and if you remove even a single element of the covering, you'll leave uncovered some part of $(0, 1)$. This finite subcovering game has revealed another distinction between open and closed sets.

- Recall back in §3.1.1 that a point which is in no open sets is a limit point to every set, meaning it is in every closed set. If any given closed set includes a point lying in no open sets, it has no open covering. That makes all closed sets compact in a vacuous feels-like-cheating way: all open covers have a finite subcover because there aren't any open covers.

Zeno's covering took a closed set and showed how it is compact, but if we include one condition with which we are already familiar, we can go in the other direction to show that compactness implies closedness:

Theorem 8.1. Compact + T_2 ⇒ closed
If a set in a T_2 $\langle(p)(p)\rangle$ space is compact, then it is closed.

This is a variant of the story that a closed set contains all its limit points; let me show you how. Rephrasing the definition of *limit point* (Definition 3.2) in the context here, a limit point of a compact set CS is a point such that all open sets around the point intersect CS (minus the point itself). For an ostensible limit point x outside of CS, can we construct an open set around it which does not intersect CS?

1. If a point x is outside a compact set CS, then for any point $y_i \in CS$, T_2 $\langle(p)(p)\rangle$ tells us there is a pair of open sets, S_x^i and S_y^i which separate x and y_i.

2. Pick another point $y_j \in CS$, and there is another pair of open sets, S_x^j and S_y^j separating x and this new y_j. We have to use superscripts because for different points in the compact set, the pair of separating sets may be different.
3. But compactness: even if there are an infinite number of points $y_i \in CS$, we cover CS with a finite number of sets in the S_y^i series. The others—and their corresponding S_x^is—we can throw out.
4. Because the remaining list of S_x^is is a finite set of open sets, its intersection is an open set. It is the portion of the S_x series that doesn't intersect any of the sets that cover CS, meaning this finite-intersection set is disjoint from CS.

We've constructed an open set containing x and disjoint from CS. So whatever x is, if it isn't in CS, it can't be a limit point of CS. If CS contains all its limit points, that makes it a closed set. That segues nicely to the discussion of infinite sequences and that sort of limit point.

But not yet, because one more problem intercedes: $\mathbb{R}$ is closed in the Usual Topology (it's $\widetilde{\emptyset}$, and $\emptyset$ is always open). But picture open balls around each integer, just big enough that one open ball overlaps the next. That gives you an infinite open covering, and removing any one of them leaves a gap in the covering, so $\mathbb{R}$ is a closed set that is not compact. That is, compactness reveals a distinction between open sets on the one hand and closed *and bounded* sets on the other.

Generalizing the concept of closed to any arbitrary topology was the first thing we did back in Chapter 3. Boundedness is a little harder to generalize to oddball spaces; compactness is as close as we'll get.

Theorem 8.2. Compactness on Reals (Heine-Borel)
For a subset A of $\mathbb{R}^N$ (for any finite dimension N) in a metric-induced topology (§2.2), A is compact iff it is closed and bounded.

All of $\mathbb{R}^N$ is unbounded and not compact, but this too can be mitigated:

- In §7.3.3, there was a box about adding ω or a North Pole, a single point which represents infinity, to an open set which otherwise trails off to infinity. There's a term for this: *the one-point compactification.*

- Recall that the complement to a ball around ω is a finite closed interval; a ball around ω is intended to embody the concept of all points above and below some cutoff.
- That does indeed compactify: an open covering of the space including ω has one set covering everything but a finite closed set, plus some number of open sets covering that finite closed set, plus who knows how many redundant sets. As per the theorem, the finite closed set is compact and so has a finite subcover, then add the set covering ω, and you have a covering of the space with a finite number of open sets.

8.1 Limits in compact spaces

On the first page of the introduction, I'd brought up Zeno's paradox, which in numeric terms asks whether the infinite sequence $\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots$ converges, and here toward the end of the book, we have a response to the question, not via careful algebraic manipulation, but by looking at the space in which the sequence lives. In this case, for example, the sequence lives in the Usual Topology inside of $[0, 1]$, which Theorem 8.2 states is compact.

For now, we'll stick to metric spaces, where we have a sense of distance and what it means for the turtle to cross the finish line; we'll easily generalize after the discussion of the main theorem.

Theorem 8.3. Convergence in compactness
Given a space S in a topology generated by a metric (as per §2.2.1, or which could be generated by one, as per Chapter 9 below). Then S is compact iff every infinite subset of S has a limit point in S, and iff every sequence of points in S has a convergent subsequence.

This is the sort of comfort a mathematician seeks: if you have compactness, you know your sequences are going to be convergent, and that you have as many convergent sequences as you'd like.

- The intuition works with the canonical open set, $(0, 1)$ and its closure $[0, 1]$. The sequence $\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots$ converges to one, even though one never quite appears in the sequence. The endpoint 1 is in the compact set $[0, 1]$, but with the limit of 1 missing from $(0, 1)$, the open interval is disqualified from compactness.
- We did something similar in the lead-in to Definition 3.4, which said that a closed set contains all its limit points, and if you have all your limit points you're closed. For one point p to

be a limit point of a set S, all sets containing p also contain some portion of S (Definition 3.2). You have one fixed point and a possibly infinite list of sets containing it. Convergence in compactness goes in the other direction: we have an infinite sequence of points, and we are checking whether they live inside one set or not.

The theorem mentions that every sequence of points inside a compact set has a convergent subsequence. This is, conceptually, a big deal, because we've found a way to quickly understand an infinite number of points which could otherwise be incomprehensible.

Imagine a sequence of points in the Outer-thirds Set (§2.3.3), arbitrarily jumping around—and it has to do some jumping, because every point is isolated from every other. Does this sequence have a convergent subsequence in it? We have a quick and immediate answer: the Outer-thirds Set is closed and bounded, and so compact. You don't need to tell me anything else about the sequence for me to say that it must have a convergent subsequence.

The theorem is about a subsequence, not the entire sequence, because sequences don't have to be in any way monotone, and may harbor complexities. Consider the sequence $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{3}{4}$, $\frac{1}{5}$, $\frac{4}{5}$, $\frac{1}{6}$, $\frac{5}{6}$, …, where the even entries are converging neatly to zero and the odd entries converging neatly to one. Maybe you wouldn't call this a convergent sequence, but because it has two subsequences that converge to definite limits, we'd consider it to be very well-behaved and understandable.

- Found a convergent subsequence in your sequence? Remove it, and you either have a finite number of points (which you can slip into the infinite sequence), or there is a remaining infinite sequence. If you still have infinite points, re-apply the theorem to find a convergent subsequence of the remaining elements. In the above example, maybe you've discovered the sequence $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$, $\frac{1}{6}$, …, and can note that it converges to 1 and remove it to leave $\frac{2}{3}$, $\frac{3}{4}$, $\frac{4}{5}$, $\frac{5}{6}$, … .
- Remove that new convergent subsequence, and re-apply as often as needed. You'll wind up with a list of convergence points, a nice characterization of the sequence.
- The list could be infinite, and an infinite number of convergence points sounds like a hassle, but the theorem tells us that those convergence points are in the same compact set as the original sequence, so re-apply the theorem to the infinite

list of convergence points: strung into their own sequence, those convergence points have a convergent subsequence.

- Repeat on that list of points if you need to. The structure may be many-layered, but once you have all limit points in a compact space, there is a structure to be had.

We can't draw the Outer-thirds Set, only the initial steps toward it like Figure 2.4, let alone any infinite sequences of points on the set. Yet we know that any sequence is decomposable into subsequences that each converge to a single point, and if we have an infinite number of such subsequences, that those convergence points can themselves be sorted into convergent subsequences. We may not be able to draw the sequence, but we can say much about its structure.

8.2 Proving things with compactness

Topologists love compactness because assuming whatever object of interest has a finite subcover makes possible all sorts tricks which can only be executed in a finite number of steps.

As per the box at the head of Chapter 3, the fundamental asymmetry of closed versus open sets is that the infinite union of open sets is infinite, and the infinite intersection of closed sets is closed. But if we have compactness, then the distinction softens, because anything associated with an open covering, including the set of closed complements of all of its elements, is finite.

Here's an example of how we can use compactness to go places. In this case, we can jump the T-hierarchy of Chapter 7 from T_2 $\langle(p)(p)\rangle$ to T_4 $\langle[p] + (C)(C)\rangle$ via compactness.

Above, you saw how compactness plus T_2 was used to generate pairs of separations and then pare those down to finite lists; here the construction of a separation will do that two times. For it to work, we need every nontrivial closed set to be compact, at which point we can call the topology itself compact.

Theorem 8.4. Compact $T_2 \Rightarrow T_4$
Every compact T_2 $\langle(p)(p)\rangle$ space is normal$\langle(C)(C)\rangle$.

1. Given two closed sets C_1 and C_2 in a compact T_2 space, we want to show that they are separated by open sets. Pick a point in the first, p_α, and by T_2, for any point p_{2a} in the second closed set, there is a pair of open sets U_{1a} and U_{2a} where U_{1a} contains p_α and U_{2a} contains p_{2a}.

2. Repeat for every point in the second closed set, producing a separation between our fixed p_α and that other point in the second closed set. Produce a separate separation for every point $p_{2a}, p_{2b}, \ldots$ and you produce a sequence of p_α-excluding sets $U_{2a}, U_{2b}, \ldots$. Because there's a set for every point in C_2, the union of all those sets covers C_2.
3. So by compactness, a finite subset of those covers C_2. Maybe U_{2b} and U_{2d} are redundant sets, and we have some set of open sets like $U_{2a}, U_{2c}, U_{2e}, \ldots$. Call the union of these $U_{2\alpha}$.
4. Removing the partners of the subsets we removed from that sequence, we have a corresponding finite sequence of $U_{1a}, U_{1c}, U_{1e}, \ldots$, and the intersection of a finite sequence of open sets is an open set, and by construction that intersection includes p_1. This intersection, call it $U_{1\alpha}$, will be disjoint from C_2.
5. At this point, we have $U_{1\alpha}$ and $U_{2\alpha}$ which don't intersect, but $U_{1\alpha}$ includes a point in C_1, and all of C_2 is a subset of $U_{2\alpha}$.
6. Now pick the next point in C_1, named p_β. Repeat the whole game to produce $U_{1\beta}$ and $U_{2\beta}$, where $p_\beta \in U_{1\beta}$ and $C_2 \subseteq U_{2\beta}$.
7. Repeat for every point in C_1, so now you have two infinite sets of sets, $U_{1\alpha}, U_{1\beta}, U_{1\gamma}, \ldots$, and $U_{2\alpha}, U_{2\beta}, U_{2\gamma}, \ldots$. Compactness strikes again: there is some subset of the U_1-series whose union covers C_1, call that U_{t1}, and you can throw out all the others—and their corresponding elements in the U_2-series, leaving a finite U_2-series.
8. Once again, the U_2-series is finite, so its intersection is an open set. Because all elements in that series contained C_2, their intersection is an open set which contains C_2; call it U_{t2}.
9. We've hit our target: U_{t1} is an open set containing all of C_1; U_{t2} is an open set containing all of C_2; we constructed them so they don't intersect. Then any closed sets C_1 and C_2 can be separated by open sets, as per the definition of normality$\langle(C)(C)\rangle$.

By T_4 $\langle[p] + (C)(C)\rangle$ we had gotten to situations where there were infinite ripples of closed-in-open-in-closed-in-open sets, which we did not have with T_2 $\langle(p)(p)\rangle$, so compactness is giving us a lot of structure at once. I leave to you to decide where you want to rank compactness on the interestingness scale, but on the utility scale it certainly ranks highly. It's like discovering that a formidable set is really just a finite number of smaller sets in a trench coat, and suddenly it isn't so formidable anymore.

8.3 Choice

This section is motivated by failure. In the next topology, we're going to start with the simplest possible topology, clearly compact, and multiply it an infinite number of times to get something that is clearly not compact. Can we work around this and generate infinite products which are compact? The answer is yes, but to get there we're going to need some things which are not Topology *per se* but which do fit the definition of interesting from the intro chapter, with interesting names like *posets* and *the Axiom of Choice*.

8.3.1 The Infinite Cube Topology

Start with a two-point Discrete Topology, just 0, 1, and the corresponding singleton sets (and everything and $\emptyset$). This is most definitely a compact topology: every infinite cover of any set in the topology has a one- or two-set subcover using the nontrivial sets $\{0\}$ or $\{1\}$. It helps that this is a discrete topology.

- The product of this topology with itself would form a square, the space of points $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ and all the possible sets you could form from those four points.
- The product with itself again gives us the eight vertices of a cube. Keep going, and you can get the vertices of a 4-D cube from $(0,0,0,0)$ winding out to $(1,1,1,1)$. Keep going, and you can generate an infinite-dimensional cube, where each vertex has an infinite coördinate like $(0,1,0,0,1,\dots)$.
- The sets in this product space are the product of sets in the 1-D space, so they too are list-like, like $\{1\}\times\{0\}\times\{0,1\}\times\emptyset\times\{1\}\times\dots$. Within all those sets, there are some where every dimension picks either a zero or a one, corresponding to a single point in the space.
- Instead of writing such a set as a cross product or an ordered list like $(0,1,0,0,1,\dots)$, you could put a decimal point in front and write it as a single binary decimal, like $.01001\dots$. That is, there is an injective mapping from $[0,1]$ to the Infinite Cube Topology, as we can read the point in $[0,1]$ as a binary decimal giving the coördinates of a point in the Infinite Cube, which is a singleton open set in the topology. (Question: is the function continuous?)
- If every closed set in a topology is compact, we call the topology itself compact, but this topology is not compact. Let's pull an infinite number of points from $[0,1]$; for example, take

the decimal parts only of π, 2π, 3π, ... and you have an infinite number of points, distributed evenly around the interval, which map to an infinite number of corners of the infinite cube, which correspond to an infinite number of singleton sets. In a Discrete Topology, the union of singletons is clopen: each singleton in the collection is open so the union is open; each singleton outside the collection, and their union, is open so the collection itself closed.

- That infinite set of singleton sets has no finite subcover of that infinite list of open points. We had what was most definitely a compact topology in one dimension, took its product with itself an infinite number of times, and wound up with a definitely not compact topology.

A few things to unpack from this example:

- To answer the parenthetical question above: no, the mapping from $[0,1]$ to the Infinite Cube Topology is not continuous. A singleton set like $\{(0,1,0,1,\dots)\}$ is an open set in the Infinite Cube, but the preimage, the singleton set $\{0.0101\dots\}$, is not open in the Usual Topology. We took the $[0,1]$ interval we're so used to and gave each point its own space in an infinite-dimensional grid.
- A discrete topology over a finite number of points will always be compact. But the infinite product of compact topologies is not necessarily compact.
- In Chaper 2 the topologies were usually built by starting with some natural-seeming group of sets, like how τ_{hc} started with open sets of the form $[a,b)$ and then derived intersections and unions therefrom. That's not what we did with the box-of-pairs topology, where we made new infinite-dimensional sets not by union or intersection, but by cross-product.

Now you see the problem: an infinite product of compact spaces is not compact. To address it, we embark on a side-quest to the Axiom of Choice. The Axiom of Choice has a dramatic name, but is usually stated in a very pedestrian way. Don't overthink the key definition, which is just a function that pulls one element from an infinite number of sets:

Definition 8.3. Choice function
For an infinite set of sets, $\mathcal{S} \equiv S_1, S_2, S_3, \dots$, where each set has elements in the space X, a choice function is a function $f : \mathcal{S} \to X$ selecting one element per set.

The Axiom of Choice states/assumes that such a choice function always exists. You might be able to construct a choice function for some infinite set of sets, in which case there is no need to rely on an axiom, but ever since it was pointed out in the early 1900s (attributed to Zermelo) that proofs often assume a choice function over infinite sets of sets, mathematicians are more careful to state when they are making this assumption. This bit of pedanticness carves out a part of math, the set of things you need the Axiom of Choice to prove, where the constructions have a distinctive feel.

And yes, it is an assumption, not something that can be derived from other first principles. You are likely familiar with *The Parallel Postulate*, the fifth axiom in Euclid's *Elements*, that parallel lines never meet, just stay the same distance apart to infinity. With this axiom added to all the others, you can derive all the hundreds of theorems of plane geometry. It can't be proven via the other axioms, and in fact, if you don't assume it you derive different results which are consistent amongst themselves. On a globe, the up-down lines of longitude are parallel at the Equator, but all meet at the North and South Poles; assuming parallel lines eventually meet can give you theorems about shapes on a sphere (or other $\mathbb{S}^2$-like surfaces). Assuming that parallel lines diverge gives you *hyperbolic geometry*.

Though, if you assume the opposite of the Axiom of Choice, that for some infinite sets of sets there does *not* exist a choice function, you don't discover a wonderful new world, there's just less you can do.

Partial order

This is a generalization of the familiar $\leq$ comparison. A partial order is reflexive ($a \leq a$), antisymmetric ($a \leq b \Rightarrow b \nleq a$), and transitive ($a \leq b$ and $b \leq c \Rightarrow a \leq c$).

Clothing manufacturers will often sell the same product in men's small, medium, and large, and women's small, medium, and large. A women's medium is definitely larger than a women's small, but how does a men's small compare to a women's medium? It depends on the manufacturer, and the comparison might not even make sense, like if women's medium jackets have a wider chest but narrower waist than men's small. We have a partial order, where we are guaranteed that the $\leq$ relation is clearly defined between some pairs of elements, but between other pairs all we can do is shrug.

- Sets with an associated partial order are called *posets*, for the obvious savings in syllables, and because it's a fun word.
- Of course, $\mathbb{R}$ is *totally ordered*, meaning there is a relation such that for any two numbers, either $a \le b$ or $b \le a$. To be pedantic, a totally ordered set is still a poset.
- Inclusion, the $\subseteq$ comparison, makes a nice partial order. For example, $\{1,2\} \subseteq \{1,2,3\}$, but in this context there is no way to compare $\{1,2\}$ to $\{1,3\}$.
- Partial orders generate chains, which may interweave to lattices, like how $\{1,2\} \subseteq \{1,2,3\} \subseteq \{1,2,3,4\} \subseteq \ldots$ and $\{1,3\} \subseteq \{1,2,3\} \subseteq \{1,2,3,4\} \subseteq \ldots$. Those are two chains which merge. Or, chains can diverge, as $\{1\} \subseteq \{1,2\} \subseteq \{1,2,3\}$ and $\{1\} \subseteq \{1,2\} \subseteq \{1,2,4\}$. Other chains may remain incomparable throughout, strands in their own separate spaces.

Even the set of theorems of *The Elements* are a poset, using the *is proved using* relation: The first few theorems are proved using only the axioms, then those theorems are used to prove others, and so on through thirteen books, forming a series of forking paths, or a tree; you can easily find such maps online. The map has a Parallel Postulate region consisting of everything subsequent to the Parallel Postulate on the *is proved using* paths, where every theorem has some plane-geometry character to it. Like cities and forests, there is a geography to mathematical theorems, and different neighborhoods have a different character or feel to them.

One interesting thing about theorems that rely on the Axiom of Choice (i.e., theorems in the Axiom of Choice region of the poset of theorems) is that they often don't just rely on the Axiom, but are *equivalent* to it. These theorems have a form like *given an infinite set of sets in some context, and assuming the Axiom of Choice, you can do this useful trick with them*, and then somebody comes along later and shows that if you take that useful trick as assumed and given, then you can use it to construct a choice function by recasting any given set of sets to the context of the theorem.

Theorem 8.5. Maximal element if Axiom of Choice (Zorn)
Given a nonempty poset where every chain has an upper bound, then the poset has a maximal element.

As per the header, the proof requires the Axiom of Choice. But for any given set of sets, you can invent some kind of ordering which creates a poset—at the extreme, if nothing is comparable to anything else then you have a poset where every element is its own one-element isolated chain. The details are well beyond the scope of this Topology book, but if you take Theorem 8.5 as given, then you can derive the Axiom of Choice.

Theorem 8.6. Maximal element iff Axiom of Choice
Given Theorem 8.5, the Axiom of Choice follows.

So there's one example of how the Axiom of Choice can be put into a certain context, posets, and then the context is general enough that we can turn around and use the theorem as an axiom from which choice functions follow.

8.3.2 Infinite cross products

If we bring this back to Topology, the Axiom of Choice will tell us something about compact sets, and that something will be equivalent to the Axiom itself.

Theorem 8.7. Compactness in Product Topology iff Axiom of Choice (Tychonoff)
Given The Axiom of Choice, the infinite cross product of compact sets is always compact in the Product Topology; and vice versa.

I mentioned above (and when I first mentioned The Box Topology in §5.2) that a product of $\{0, 1\}$ against $\{A, B\}$ is neither a union nor an intersection, but a new type of operation. Maybe things are easier if we go traditional and start with a basis and build the topology with intersections and unions. We'll need a definition of *basis* more precise than *a natural-seeming group of sets*; here's one:

Definition 8.4. Subbasis for a topology
A set of sets where those sets plus their finite intersections form a basis for the topology.

Those intersections aren't in the subbasis, and there will be a lot of them. If you have four sets, there are eleven intersections (six pairs, four triplets, and the intersection of all four, many of which may be $\emptyset$). Of course, if you have an infinite subbasis, there are an infinite number of intersections, and so potentially infinite ways to cover a set. The next theorem gives us a little shortcut: instead of checking all possible open coverings, we only have to check those formed using elements of a subbasis.

Theorem 8.8. Compactness via Subbasis (Alexander)
Given a topology (X, τ_x) and a subbasis $\mathcal{S}$ (and the Axiom of Choice), if every open covering of a set via elements of $\mathcal{S}$ has a finite subcover, then X is compact.

Its dependence on the Axiom of Choice is not direct. What's going to happen is that the Axiom of Choice is going to put a structure on the collection of infinite coverings with no finite subcovering—so much structure that it will be impossible for that collection to exist.

DONUT: OK, so we have some set C that we want to cover, and I'm given that some subbasis exists such that if there is an open covering of C using just the elements of the subbasis, it has a finite subcover.

SOCRATES: Yes, and from this the theorem says that every infinite open covering of a set via *any* set of sets, not just those in the subbasis, will have a finite subcover.

DONUT: I don't believe you.

SOCRATES: What?

DONUT: I don't believe you. I think there is a whole space of infinite covers of C with no finite covers. I'm calling it $\mathcal{IC}$.

SOCRATES: OK, I'll humor you. You think not just that the theorem is wrong, but there is a large set of infinite covers with no subcovering?

DONUT: Yeah, once you get one, you can add and subtract elements.

SOCRATES: But adding elements to an infinite cover can induce a finite subcover. Remember when we built Zeno's cover, then added a single set, which obviated the need for an infinite number of other sets?

DONUT: Then that was a mistake. There are lots of sets you could have added that *wouldn't* have induced a finite subcover. Just add something like $(0.2, 0.3)$ to Zeno's covering and that won't cause a finite subcovering to suddenly appear.

SOCRATES: It is an additional set redundant to the open covering.

DONUT: It makes the covering *thicker* and *more lush*. There's probably an infinite number of sets you could add that would have not induced a finite subcover.

SOCRATES: When you start with one set of sets C_1 and add more sets to it to form C_2, then $C_1 \subseteq C_2$.

DONUT: Yeah, you get me.

SOCRATES: We were just talking about how this inclusion operator $\subseteq$ creates a poset.

DONUT: Those were fun. I like how they turn a bunch of elements into a beautiful lattice. Now $\mathcal{IC}$ isn't just a bunch of covers, it has a structure to it, like a collection of interweaving towers of coverings and thicker coverings. I am liking $\mathcal{IC}$ more and more.

SOCRATES: Don't get too attached. We know that the subbasis has a finite covering of C, so if you add the elements of the subbasis to C_2, then your Jenga tower of coverings will collapse to a finite covering.

DONUT: I ran with your blanket metaphor of nice, warm coverings that get thicker and plusher, and you turned it into a precarious Jenga tower?

SOCRATES: Regardless of your metaphor, there is some upper bound to how thick you can make your covering, some upper bound to the chain of inclusion inside of $\mathcal{IC}$.

DONUT: OK, I guess so, there's an upper bound, because if you include all the elements of the subbasis—

SOCRATES: Possibly even fewer elements from the subbasis.

DONUT: —yeah, I dunno how many elements from the subbasis, you would no longer be in $\mathcal{IC}$ because you'd be an infinite cover with a finite subcover. But, that means those aren't in $\mathcal{IC}$.

SOCRATES: Right, an upper bound of a chain in $\mathcal{IC}$ doesn't have to be in $\mathcal{IC}$ itself to be a bound.

DONUT: I guess the point is just that none of our chains can go forever?

SOCRATES: Yes, and this brings us to Theorem 8.5: "Given a nonempty poset where every chain has an upper bound, then the poset has a maximal element." Is $\mathcal{IC}$ nonempty?

DONUT: Yes! Yes it is! I just know it.

SOCRATES: We'll see about that. It's a poset-

DONUT: Yeah, because so many infinite coverings include other infinite coverings, so we get those chains of inclusion.

SOCRATES: And every chain has an upper bound, because adding enough of the subbasis would certainly create a finite cover.

DONUT: Well if you want to turn mathematics into filling in forms, I've filled in your form: nonempty, poset, upper bounds. All the boxes are checked, so it looks like $\mathcal{IC}$ has a maximal element.

SOCRATES: In fact, there could be many, but there is definitely at least one.

DONUT: Should we name it? I'm calling the maximal covering MC. Wait, is $MC \in \mathcal{IC}$?

SOCRATES: Yes, an upper bound might be outside the set, but a maximal element is in the set.

DONUT: Oooh, that theorem just got a lot more interesting.

SOCRATES: I'm interested to hear how MC relates to the subbasis. Let's pick one set M out of the maximal covering MC. By the definition of the subbasis $\mathcal{S}$, there is some finite list $B_1, B_2, \cdots \in \mathcal{S}$ such that $B_1 \cap B_2 \cap \cdots = M$.

DONUT: That's fine. It doesn't mean any of those Bs are in MC, which means you still might not have a finite subcover. There have to be sets $M \in MC$ such that for their subbasis breakdown, $B_1 \cap B_2 \cap \cdots = M$, some Bs are not in MC.

SOCRATES: True, let's focus on one such set and think through the breakdown. Donut, would you please get a bucket?

DONUT: Like a big one?

SOCRATES: Nothing extravagant, just big enough to hold a finite number of sets.

While the talking doughnut is getting a bucket, we can pause to review the construction so far. Our interlocutors found a covering named MC with the property that it is a maximal infinite covering with no finite subcover, and because they relied on Theorem 8.5, they relied on the Axiom of Choice. Being maximal, adding but one more set to MC would allow us to get a finite subcover from it.

SOCRATES: Thank you. We have some element of MC, named M, which is the intersection of a finite number of elements of the subbasis, $B_1 \cap B_2 \cap \cdots = M$.

DONUT: Yeah, that's the point of a subbasis.

SOCRATES: And some of those B sets will not be in MC.

DONUT: There have to be some Ms for which that's true, because otherwise you'd have enough elements from the subbasis that you could get a finite cover. But some of those Bs might be in MC.

SOCRATES: Possibly so. If they are, please put these Bs which are in MC in the bucket.

DONUT: Sure. Here's one, B_3, which isn't in MC.

SOCRATES: We've agreed that MC is a maximal element of $\mathcal{IC}$, meaning that if we add even one more set to it, that larger covering is no longer in MC, because it will have a finite subcover.

DONUT: Yeah, we said that if you add too many subbasis elements you'd get a finite cover, but if MC is maximal, all it takes is one more.

SOCRATES: What does that finite cover look like?

DONUT: It would be a bunch of sets V_1, V_2, ... all in M, plus B_3.

SOCRATES: Set B_3 aside, but please put that finite group of sets V_1, V_2, ... into the bucket.

DONUT: Sure. It looks like B_5 is also not in MC.

SOCRATES: Adding that to MC would produce another finite cover, then?

DONUT: Yeah, with $W_1, W_2, \cdots \in MC$.

SOCRATES: Please put those Ws in the bucket too.

DONUT: Do you want me to do this with all the Bs not in MC?

SOCRATES: Please do. There are only a finite number of them, so it won't take long.

DONUT: [*Donut sings songs as open sets fall into the bucket.*] Done!

SOCRATES: Great, now what do we have in the bucket?

DONUT: A bunch of covers, almost, because you told me not to put B_3, B_5, and all the other elements not in MC into the bucket. The covers all have a little hole where they don't cover C entirely.

SOCRATES: We picked those Bs so that their intersection forms M, an element of MC. Let's throw M into the bucket too.

DONUT: Nice, because when I take the union of all the coverings, the hole is the intersection of all the holes in each covering, and that intersection has to be smaller than M. Swell, M covers the hole! I have a finite covering of C by taking the union of everything I have in the bucket plus M. All those are in MC, so MC has a finite covering after all!

SOCRATES: Wonderful. So MC—

DONUT: Why am I happy about this? If $\mathcal{IC}$ is the set of infinite covers with no finite subcover, and we found a fininte subcover within MC, then that means MC isn't actually in $\mathcal{IC}$! How can a set have a maximal element that isn't in the set? Something went wrong.

SOCRATES: "Given a nonempty poset where every chain has an upper bound, then the poset has a maximal element." We definitely have a poset, and we agreed that every chain has an upper bound. That leaves only one other box on the form.

DONUT: You mean $\mathcal{IC}$ is not nonempty? In other words, empty! I spent all that time building the set of infinite covers with no finite cover, and it turns out there aren't any?

SOCRATES: That's the only conclusion.

DONUT: Let's go play Jenga.

SOCRATES: Eh, my hands aren't so steady, now that I'm approaching 2,500 years old.

DONUT: Well I'm a talking doughnut. No excuses, let's go.

Poor Donut: after all that effort to find infinite coverings with no finite subcovering, Theorem 8.8 stood, and having finite subcoverings for any covering with the subbasis was sufficient to guarantee finite subcoverings for any covering at all. They used posets and the Axiom of Choice to get to that conclusion, and as noted by, e.g., Gamelin and Greene [1999], there's no way around it: any proof of Theorem 8.8 will involve The Axiom somehow.

After that long walk, it is a short hop to Theorem 8.7, that the Product Topology is compact. The subbasis theorem told us that if you can get compactness using only the elements of a subbasis, then you have compactness overall, and the Product Topology has a subbasis. In the box of §5.2, these were named π_x^{-1}, π_y^{-1}, the inverse projections from individual dimensions, and the Product Topology was the finite intersection of these sets which are each restricted on one dimension. But without the formality the topology is easy enough to understand as the cross-product of an infinite number of dimensions, with the caveat that only a finite number of them are restricted at a time.

In the discussion above, an infinite number of trips to the bucket would fill the bucket with an infinite number of open sets, but as it turns out, a set in the Product Topology is a finite intersection of (sub)basis elements.

As it is, an open set in the Product Topology could be written as a list of open sets, bearing in mind that an infinite number of them will be $(-\infty,\infty)$, maybe $\{(0,1),(-\infty,\infty),(1,2),(2,3),(-\infty,\infty),\dots\}$. If you know that there is a finite subcover for every infinite cover on each dimension, you can pick out an overall finite subcovering. This is the Axiom of Choice all over again: given an infinite collection of potentially infinite sets, there is some means of selecting a finite number of elements from each set in the collection.

For many proofs of *if and only if* theorems, the proof goes in one direction, and then the author basically rewrites the steps in reverse order to prove the other direction. To get to the compactness of the Product Topology over compact spaces, we used the Axiom of Choice not to construct some sort of choice function, but to show that a certain set of things ($\mathcal{IC}$) must be empty. Once we got to this side of the Axiom of Choice dividing line, we found something where the return walk is straightforward.

Chapter 9

Metrizable

We got far without talking about distances. Of course, the Usual Topology wasn't really invented by topologists, but was based on the standard Euclidian distance function, from which ϵ-balls can be measured, though topologists used those ϵ-balls without serious regard for how they encode a concept of distance.

Other topologies like the Finite-complement Topology have no real ordering to them and its open sets are in no specific place. But metrics are relatable, and there's a wealth of literature based on the concept of things which are closer together or further apart. Given a topology of sets, can we work backward to find a metric that could have generated that topology? Perhaps we can hang the points in just the right configuraton in space so the balls around each happen to exactly match the list of sets in our topology. Maybe our open sets were meant to embody some odd and unique concept, but if they can be expressed via a metric, the result is still just another subspace of the Usual.

Regarding those toy examples interspersed throughout this book based on a small collection of N points with some arbitrary-but-complete set of open sets we made up, yes, you can hang those points in $\mathbb{R}^{2N}$ (i.e., $2N$-dimensional reals) such that your open sets match ϵ-balls around the points. If you have $N = 20$ points, $2N = 40$ dimensions is a lot of space to move around in, so although I'm not going to go through the construction, it feels plausible at first glance that you'd be able to find some sort of configuration given all that space.

What about topologies where there are an infinite number of points? Are they still just another version of N-dimensional reals?

This is relevant because metric-induced topologies are very well-behaved. They are T_6, which implies T_0 through T_5. Theorem 8.2 said that on $\mathbb{R}^N$ with a metric-induced topology, a set is compact iff it is closed and bounded. Of course, the theorem doesn't care whether you built your topology around a metric, or you could have done so but chose to use some other method that wound up with the same topology.

9.1 Second-countable

Convergence, continuity, and other manifestations of Zeno's paradox are often reduced to being about some ordered sequence of points or numbers, $x_1, x_2, \ldots$ and whether that sequence converges to a single value or not. On the $\aleph$-ladder (§2.1.7), those sequences are always $\aleph_0$: countable, with a neat integer index attached to each element of the sequence. In this section, we look at situations where a topology's list of open sets has a countable basis.

The hierarchy for countability has nomenclature as bad as the T_0–T_6 hierarchy but it mercifully stops at two, and to my ear has a pleasant tone, reminiscent of hobbits having second breakfast. From it, we'll build a structure, the infinite-dimensional hedgehog, which will be theoretically useful for the question of when a topology is a metric in disguise.

First-countable isn't necessary for the thread of this chapter, but if I jumped straight to second-countable you'd be wondering, so let's get one useful thing out of it and move on.

Definition 9.1. Countable point-basis
A point has a countable point-basis iff there exists a countable set of open sets B such that every neighborhood including the point contains at least one of the members of B.

Definition 9.2. First-countable space
Every point in the space has a countable point-basis.

The Usual Topology is indeed first-countable. There are an uncountable number of ϵ-balls around a given point, but you can express them using a countable basis. Around each point, there is a rational-radius ϵ-ball of radius q for all $q \in \mathbb{Q}$, and there's one small enough that every neighborhood will have one.

In the section on convergent sequences (§3.3), we defined such a sequence $x_1, x_2, \ldots$ as one where, for every open set U around the ostensible point of convergence, there is an n such that $x_m \in U, \forall m > n$. If you have a countable basis around x, then you are guaranteed that such a convergent sequence exists, basically by pulling one point from each set. Around zero, you have $(-10, 10)$, $(-1, 1)$, $(-0.1, 0.1)$, $(-0.01, 0.01)$, $\ldots$, and you could pull the points just inside the upper ends, to form a sequence $9, 0.9, 0.09, 0.009, \ldots$. Stated as a theorem:

Theorem 9.1. First-countable ⇒ convergent sequence
If a point has a countable point-basis, ∃ a sequence which converges to the point.

Now that you've met first-countable and the point-basis and they have shown you a little of what they can do, we can move on to the entirely different concept of second-countable.

Definition 9.3. Second-countable space
The space has a countable basis.

Second-countable implies first-countable, because if a space has a countable basis then you can't have individual points with an uncountable point-basis. So Theorem 9.1 directly applies to tell us that in a second-countable space, there is a sequence which converges to any given point. But in the other direction, a countable point-basis around each of an uncountable number of points may or may not boil down to a countable number of open sets overall.

- Yes, $\mathbb{R}$ under The Usual Topology is second-countable, because the rationals are dense in the reals (§3.4). Any ϵ-ball around an irrational number can be constructed by a countably infinite union of ϵ-balls around rational numbers. Every irrational-radius ϵ-ball is the infinite union of rational-radius balls.
- The half-closed topology τ_{hc} (§2.1.7) is not second-countable, because if we do the same rational-to-irrational construction, we get the same result: the Usual Topology, not the τ_{hc} topology. If you want to express the τ_{hc}-open interval $[-\pi, \pi)$, you might want to use the union of $[-3, 3) \cup [-3.1, 3.1) \cup [-3.14, 3.14) \ldots$, but you are going to wind up with $(-\pi, \pi)$, which is one point short of the half-closed interval you wanted.
- The Lexicographic Topology (§2.3.4) can't be second-countable. Recall the fundamental weirdness of this topology was that there were an infinite number of points between any two points on different vertical lines, like the line segments $(3.14, y)$ and (π, y), $\forall y \in [0, 1]$. There is no way to compose the basis for nearby rational verticals into the basis for irrational verticals.

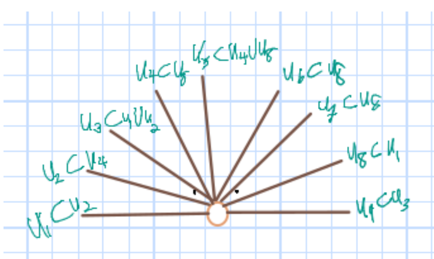
Like the finite cover we get from compactness, having a concept of *not too many* in the basis for a topology will pay off.

9.2 Metrizing second-countable

In this section, I'll show you an interesting infinite-dimensional construction, then a trick to produce metrics on infinite dimensions. After those, there will be a few sentences at the end showing that the two together allow us to produce a one-to-one mapping between ϵ-balls in the infinite-dimensional metric and sets in topologies which meet certain requirements (i.e., that those topologies are metrizable). We touched on infinite-dimensional topologies in discussing the Box versus Product Topologies (§5.2), but here we're going to spend more time getting a sense of what life in infinite dimensions is like.

9.2.1 The infinite-dimensional hedgehog

To build an infinite-dimensional hedgehog, we'll need the topology we want to metrize to be T_0 $\langle (p)p \rangle$ and completely regular $\langle C - f - p \rangle$. To save you flipping back to Definition 7.8, that means that on a topology (X, τ), for every point p and closed set C, $\exists$ a continuous function $f : X \to [0,1] \ni f(x) = 0, \forall x \in C$, and $f(p) = 1$. As a nomenclature reminder, T_0 + completely regular = $T_{3\frac{1}{2}}$.



1. If a point p_1 lives in an open set U_1, then its complement $\widetilde{U_1}$ is a closed set, and complete regularity guarantees a function f_1 that is one at p_1 and zero throughout $\widetilde{U_1}$. Within U_1, other points could readily have nonzero values.
2. Second-countability dictates that there is a countable number of open sets, $U_1, U_2, \ldots$, from which all open sets can be constructed via unions. Do the same for all of them: pick a point $p_i \in U_i$, and you are guaranteed a function where $f_i(p_i) = 1$ and $f_i(\widetilde{U_i}) = 0$.
3. It helps to bundle some of those together. If you have a pair of basis elements such that $U_1 \subset U_2$, then all the point-wise functions from the last bullet point for points in U_1 can be

joined, rolled up to a single function which is one in U_1 and zero outside of U_2. For every basis element U_i, you can do such a roll-up (with an arbitrary wrapping set), after which there is exactly one dimension i where $f_i(U_i) = 1$ and f_i outside of U_i is not.

4. If f_1 maps to the line segment $[0, 1]$ and f_2 maps to the line segment $[0, 1]$, then $f_1 \times f_2$ maps to the square $[0, 1] \times [0, 1]$. Repeat, and you can assign one function to each dimension of an infinite-dimensional $[0, 1] \times [0, 1] \times \cdots$ cube. This is where second-countable paid off: with a countable number of basis elements, we get a countable number of functions, and we can assign one dimension to each.
5. You might want to draw the two axes as orthogonal, but this space isn't like your usual X-Y-Z axes, just a lot of $[0, 1]$ lines, which we'll join at 0. As the number of dimensions gets large, people get loose and, as per the doodle above, start drawing this as a bunch of spikes all emanating from a single point. Each spike is associated with one basis element and some arbitrary open superset. This is why more whimsical topologists, like Starbird and Su [2019], call this bunch of spikes connected only at the origin *the hedgehog*.

Consider a point in the Usual Topology, like 1.2, which exists in such open sets as $(1.1, 1.3)$ and $(1.19, 7.8)$ and $(-7.8, 1.2001)$. Each of those is associated with certain spikes in the hedgehog, meaning 1.2 maps to 1 only on that specific list of spikes. You could think of that list of spikes i where $f_i(1.2) = 1$ as a signature or address in the post-mapping space. Here's a little sub-theorem about the signature:

- T_0 $\langle (p)p \rangle$ tells us that for any two points x_1 and x_2 there is at least one open set which includes one but not the other.
- Given that every open set is a union of basis elements, that bullet point can only be true if there is at least one *basis* element which includes x_1 but not x_2, or vice versa.
- That's enough to guarantee that the signature for x_1 differs from the signature for x_2. Somewhere in all those dimensions, there is always at least one dimension where $f_i(x_1) = 1$ and $f_i(x_2) \neq 1$.

Denoting the aggregate function $f_1 \times f_2 \times \cdots$ as $F(\cdot)$, the sub-theorem means the aggregate function $F(\cdot)$ is injective, meaning that for every pair $x_1 \neq x_2$, $F(x_1) \neq F(x_2)$. There's a signature mismatch on some spike of the hedgehog or another.

Another more formal way to put this is to be a little more careful about what *a point* means on the hedgehog. In your usual 3-D, the product of $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$, a value of 7 on the first dimension is not yet a point. A point is a combination of some value on all three axes, like $(7, 8, 9)$. Similarly with the hedgehog, but in a trippier manner, as the values on each axis form the signature of the point as described above, indicating which basis elements the point is or is not in. Some points may be incoherent relative to the original topology, like if $U_1 \subset U_2$ then for any point where $f_1(p) = 1$ it must be that $f_2(p) = 1$ so a point in the cube of the form $(1, 0, \dots)$ could not have come from anywhere. Define $\mathcal{H}$ as the subset of the hedgehog where the F^{-1} pre-map is a point in the original topology.

By definition and construction, F as a function from the original space to $\mathcal{H}$ is surjective, meaning every point in the post-function space has a corresponding point in the pre-image space (Definition 4.3). Then we've built an F from our initial space to $\mathcal{H}$ that is injective + surjective = bijective.

And continuity?

- Let the basis of the Hedgehog Topology be the Usual open sets on any spike. We make up the rules, so let's also include $\{1\}$ and $\{0\}$ by themselves. Then include any unions and intersections. We'll consider only the subspace topology formed by the intersection of those open sets with $\mathcal{H}$.
- One spike was generated by some continuous function mapping some basis element to the spike, so the preimage of any open set on that spike is an open set in the original space.
- By continuity and bijectivity on individual dimensions, if a set S in the original space is not open, there must be some spike where it maps to a non-open set. Rephrasing that, if we have an open set in the post-space $\mathcal{H}$, which is composed of open sets on every spike, it cannot be the image of a non-open set. Rephrasing that, the pre-images of open sets in $\mathcal{H}$ are open sets. That's continuity.

Vocab: a *homeomorphism* is a continuous, bijective function between two topological spaces (*homeo* = same, *morphism* $\approx$ shape). That's what we have here, between the original space and the $\mathcal{H}$ subset of the hedgehog.

9.2.2 The infinite-dimensional cube

As mentioned in step 4, the infinite-dimensional hedgehog (and $\mathcal{H}$) is a subset of the infinite-dimensional cube (Hilbert), which has a customary metric attached.

- We can't just use the Manhattan metric, where, indexing each dimension as i, the distance between two points is the sum of the one-dimensional distances on each dimension, $d(x_1, x_2) = \Sigma_i |x_1^i - x_2^i|$, because with infinite dimensions, that sum will typically be infinite.
- Instead, our metric will be $d(x_1, x_2) = \Sigma_i |x_1^i - x_2^i|/2^i$. That is, we're ordering the dimensions and declaring each to have half as much weight in the distance calculation as the previous. Because the max distance on any one dimension is 1, with the modification you are guaranteed a sum less than $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1$. I find it æsthetically unpleasant that dimensions aren't treated symmetrically, but it's not at all against the rules.
- For every ϵ and point x in an open set U in the infinite cube, is there an ϵ-ball in the metric around x and inside U? Easily: on each individual dimension, there is a one-dimensional range which is a 1-D Manhattan-metric ball surrounding x on that dimension and inside of U, then once you've drawn all those at the per-dimension level, shrink them all until the weighted sum of their widths is less than ϵ.
- In the topology built by taking the product of sets in every one-dimensional Usual-topology spike, our ϵ-balls are all valid open sets.
- You may remember the discussion of Theorem 2.1, where every point x in a Euclidian ϵ-ball had a Manhattan-metric ϵ-ball around it, and vice versa, meaning you could construct any Euclidian open set as the infinite union of Manhattanite open sets, and vice versa, meaning the set of Euclidian open sets is a subset of Manhattanite open sets, and vice versa, and if $E \subseteq M$ and $M \subseteq E$, then $M = E$—the two sets of open sets are equivalent. The same logic tells us that the sets of ϵ-balls from our modified Manhattan and the set of Usual-product Topology open sets are identical.

The short way to say this is: the infinite cube is itself metrizable.

9.2.3 Adding it up

Let's bring it all home. There is a homeomorphism, a one-to-one correspondence, between open sets in the hedgehog subspace $\mathcal{H}$ and open sets in the original space. Any open set in the infinite-dimensional cube can be constructed using ϵ-balls in the metric, and $\mathcal{H}$ is a subset of the infinite-dimensional cube. Then the chain is complete: any open set in the original space maps to ϵ-balls; any ϵ-ball generating something in $\mathcal{H}$ pre-maps to open sets.

Theorem 9.2. second-countable + completely regular ⇒ metrizable
If a topology has a countable basis covering all points (i.e., is second-countable and T_0 $\langle(p)p\rangle$) and is completely regular $\langle C - f - p\rangle$, then it is metrizable.

- In the notes above on second-countable we established that τ_{hc} isn't second-countable, so we already have an example of where this theorem doesn't apply and metrizability won't work. Whatever arrangement you invent for the open intervals in τ_{hc} is going to turn into an arrangement for the Usual Topology.
- The long line formed by the lexicographic ordering also failed to be second-countable, and so fails to be metrizable. It is fully ordered (for any two points, we know definitively which is greater than the other), but that isn't enough to build a distance function.

Topological properties

A *homeomorphism* is a continuous, bijective (one-to-one and onto) function between two topological spaces. Such a mapping may reshape the open sets one way or another, but it doesn't change their number or many of their interrelations.

A *topological property* is a property that does not change under homeomorphism. Distance is clearly not a topological property, as a function like $f(x) = 2x$ is continuous and bijective, but by most of our metrics the distance on the post-space is twice the distance on the pre-space. But if you want to talk about the properties of the open sets themselves, the great majority of what this book has covered is the same both before and after the double-stretching:

- T_0–T_6 and related characteristics like regular $\langle(p)(C)\rangle$, completely regular $\langle C - f - p\rangle$, normal$\langle(C)(C)\rangle$, and completely normal $\langle(S)(S)\rangle$.
- Global, local, and path connectedness.
- Compactness.
- First- and second-countability.

- And now: metrizability.

9.2.4 All the sets of $\mathbb{Q}$

Here's a new topology: all possible combinations of the rationals, $\mathbb{Q}$. This is not a subspace of $\mathbb{R}$, but more a combinatorial game: pick any bundle of rational numbers from anywhere up and down the line, and call that an open set. This is an intimidating beast of a topology. In the same infinite box of §2.1.7 where we saw that the rationals are countable, $\aleph_0$, we saw that the set of all sets of rationals is uncountable, the continuum, $\mathfrak{c}$. There's no obvious way to order all those sets: is $\{\frac{1}{2}, 1\}$ greater than or less than $\{\frac{2}{3}, \frac{3}{4}\}$? Can this topology be reduced to a metric?

- As noted, the basis of this topology, singleton sets $\{q\}$ where $q \in \mathbb{Q}$, is countable, and having a countable basis is the definition of second-countable.
- This is a Discrete Topology, in that every point $q \in \mathbb{Q}$ is an open set. As with all Discrete Topologies, every set is clopen, because if you have one set of rationals Q, then its complement $\widetilde{Q}$ is a well-defined set of other rationals.
- If you give me a set of rationals C, and another singleton set $\{p\}$, then those are two clopen sets, and as per Lemma 7.1, two clopen sets can always be separated by a function as per the requirements of complete regularity.

The form has been completely filled in: we have second-countable and complete regularity, so there exists a metric over the topology. If you have a metric, you have T_6, and therefore every property from T_0 to T_5 holds without any further proof-writing effort on our part. This combinatorial beast of a metric is very well-behaved.

Is this just because the set of all rationals is a Discrete Topology? The reals are uncountable, $\mathfrak{c}$, so all sets of reals are $2^{\mathfrak{c}}$. You still have a Discrete Topology where every point is an open set, but now it's too much: we don't have a countable basis and so can't build a hedgehog with a countable number of dimensions.

Chapter 10

Dimension

I was around in the 1990s when fractals were A Thing. You could get fractal-of-the-month wall calendars and video artists would use them to enweirden their videos. I thought they were a new concept, but it turns out that the only real progress was that computing power made it possible to do much better images, and to generate fractals procedurally, on the complex plane.

No, fractals date back to 1890, when one of them re-broke mathematics, after it was broken the first time by the Outer-thirds Set (§2.3.3), which generated a set of points in one-to-one correspondence with the points in $[0,1]$, but with zero measure. The re-break was when somebody (Peano) drew a one-dimensional line which fills two dimensions.

We'll start with something one step simpler, building one third of a Snowflake (Koch). Start with a line, and replace its middle third with two line segments, the top two parts of an equilateral triangle. By replacing, I mean that one third of the line is erased leaving the outer two lengths of $\frac{1}{3}$ each, but two pieces of length $\frac{1}{3}$ (the top two legs of the triangle) are put over the gap. With this erase-and-replace operation, we have extended the length of the path to $\frac{4}{3}$. See the first step of Figure 10.1.

Now repeat: we have four line segments, and do the same thing to each of them, replacing the middle sub-segment of each with two sub-segments, thus multiplying the length of each of the four line segments by $\frac{4}{3}$. At each step, the curve is rougher, a little less like a simple line segment.

Every time we do this operation of doubling the middle bit, we multiply the length of the path by $\frac{4}{3}$. Do this 16 times and you have

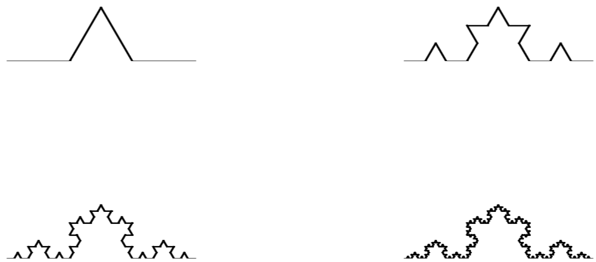


Figure 10.1: Do you want to build a snowflake? Replace the middle third of a line with two lines of the same length (the top two legs of an equilateral triangle). Now you have four line segments; repeat the same procedure with each of those four. Now you have sixteen line segments; repeat the same procedure with each of those.

just short of a 100× lengthening ($\frac{4}{3}^{16} \approx 99.77$). If you started with a one centimeter long line segment, you've now got a one meter long path. When you popped an equilateral triangle onto the 1cm line segment, you got a shape that fits into a 1cm × 0.87cm box; that's the thumbnail-sized box you fit a meter-long path into. If you do the shrink-and-duplicate operation forty times, you've fit just under a 1 kilometer long line into that box (994.4 meters, to be more exact). Shrink-and-duplicate an infinite number of times, and you will have a path of infinite length.

You could do this procedure of replacing a line segment with multiple line segments, then repeat on each new line segment, in myriad manners. Maybe replace a segment with a pentagon, or an asymmetric, steeper triangle—we'll take that to its logical extreme to fill a 2-D space below. Each scheme you invent will create a different but equally attractive fractal, and as long as you replace a line segment with a path of longer length, and do so an infinite number of times, you will evolve a path of infinite length.

10.1 Space filling

Squeezing a path of infinite length into the confines of a one-by-one box on $\mathbb{R}^2$ is not especially novel. The Topologist's Sine curve

(§3.2.1), whose bulk is the curve $y = \sin(1/x)$, is also of infinite length within the x-range $(0,1)$ and y-range $[-1,1]$.

We'll have to get even more extreme, with a one-dimensional path so long it becomes two-dimensional. Figure 10.2 starts with a path that goes forward, up, down along the same route it went up, then forward again. Imagine doing the snowflake construction with a steep triangle with very tall sides and short base instead of a flat-ish equilateral; as the base shrinks to zero the triangle becomes the spike we have here. That the spike retraces along one segment makes it harder to visualize walking along the path, but the first panel of Figure 10.2 is otherwise much like the first panel of the snowflake in Figure 10.1, with four line segments. The up- and down-segments are of height one half, so the total path length is two. Then we can play the same game of replacing all four line segments with inverted Ts of the same type. Along the vertical up-path, the inverted T will point West, and along the vertical down-path, the inverted T will point East.

To help with tracing the path, Figure 10.2 puts each T in its own triangle with one long base, and puts a number in the corner of the triangle where the path enters. So the path in the second panel starts at the Westernmost end, goes East, North, doubles back South, East, North, West, doubles back East, North, doubles back South, and so on until it walks all the way to the easternmost end. I encourage you to trace out the path walked by the third panel, after every line segment in the second panel was again replaced with an inverted T. Again, the triangles are numbered in sequence to help guide you from West to East.

The linear path is starting to look like a grid covering the space, where the grid in the third panel is twice as fine as the grid in the second panel. This is going to keep happening: The grid constructed by this single path will be twice as fine with each step.

- A little binary digression (picking up from back in §2.3.3): If you want to find a point along the x-axis, picture a binary search: split the $[0,1]$ range in half, and you either hit it, or you missed and have to go left or right. If you have to go left, write down zero; if you have to go right, write down one.
- Maybe you have to go left, then split the left half of the line segment into half. After the digit you wrote down in the first step, either write down a zero if you need to go left, or a one if you have to go right.
- Repeat with the half-of-a-half you are now focusing on, writing down the direction you have to step in and then checking that

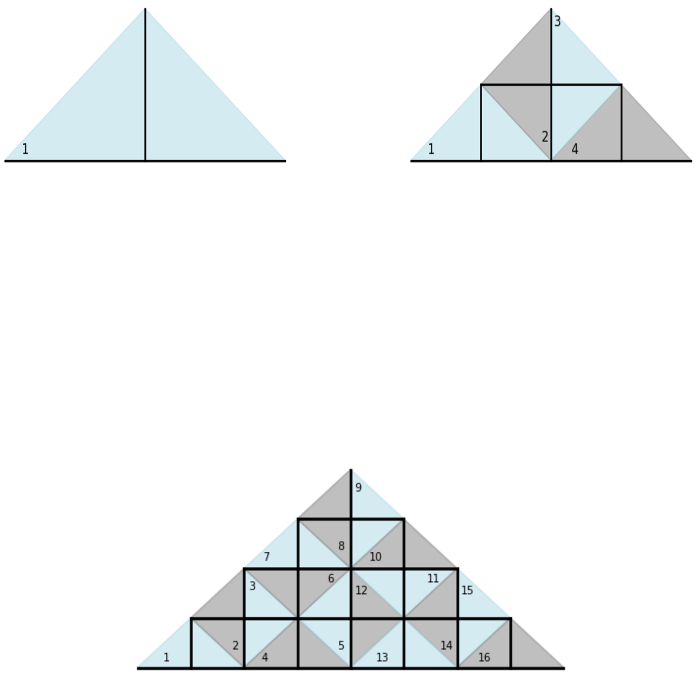


Figure 10.2: The path goes forward, up, down (doubling back), then forward again. Repeat the same operation as with Figure 10.1: replace each line segment with a replica of this shape. A grid of finer and finer scale evolves. To facilitate identification after the replication steps, the order in which the triangles are traversed is labelled at the entry corner.

eighth of the range.

- When you do a split such that your midpoint is exactly on the point you seek, write down a final 1, and you will have written down the binary decimal representation of the point.

A search like this will find any point in $(0, 1)$. The grid our fractal is constructing, which gets twice as fine at every step, is effectively executing this binary search for you, and just as a binary search will eventually hit any point (possibly in an infinite number of cleavages), this grid will eventually tighten enough to touch every point. We have covered the entirety of a two-dimensional space with a one-dimensional path.

I like this example because of the binary grid, but if the backtracking up and down annoys you, you can do the same with lines that don't overlap—look up the Peano curve. Perhaps it's even more remarkable that the curve we drew here is hopelessly inefficient, retracing the same points, and it still fills a space.

10.2 Fractional dimension via covering

And remember the Outer-thirds Set (§2.3.3), where we took a line of length one and deleted segments with a total length of one. It seems inappropriate, though, to call it zero-dimensional. Which brings us to fractional dimensions.

The concept of finding a dimension is awkward because from the start you're comparing different units: if we were here in real life, comparing 1-D versus 2-D measurements is comparing meters versus square meters. If you want to measure a 2-D shape on a paper with 1-D lines, you'd need an infinite number of them, and if you tried measuring its volume, you'd get zero every time. Dimension is the Goldilocks of the children's story, who rejects one dimension as too small, the next as too large, but she is especially exaggerated in her opinions and rates everything she doesn't like as either zero or infinity. This right-or-extremely-wrong rating is a fundamental weirdness about dimension, which makes the process of searching for a dimension a different game from the usual *solve for x* game in an already-specified space.

There are several ways to construct fractional dimensions—more than I'll cover here—but they mostly involve coverings. There are so many partly because there aren't many constraints on what a dimension has to do. Lines, squares, and cubes definitely have to be 1-, 2-, and 3-dimensional, but having established that a curve falls between one and two dimensions, there is no Correct Answer as to whether it is 1.4-dimensional or 1.6-dimensional.

- An easily skipped aside: I couldn't find many hooks for where certain non-integer dimensions have solid interpretations. If you know what the Γ (Gamma) function is, then I could tell you that the volume of a sphere of radius r in n dimensions is $\frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}r^n$. That's not a strictly increasing function, and as $n \to \infty$, the percent of a $1\times1\times1\times\ldots$ cube taken up by the sphere of radius one that would exactly fit into that high-dimensional cube actually goes to zero. The optimum (which you can find by the usual *solve for* x process of taking the derivative in n and finding where the derivative is zero) is at 5.269471. What does that number mean? How can we use it as a hook to define partial dimensions? I honestly have no idea.
- As somebody who deals with data, I usually think in terms of information captured, like putting t-shirt slogans on one axis and t-shirt color on another independent axis. That's hard to work with here in Topology-land. Recall the infinite box in §2.1.7, that showed that $\mathbb{R}$ is in one-to-one correspondence with $\mathbb{R}^2$. If you wanted to put all the t-shirt slogans in all the colors in a single linear list, it's just a question of reordering. We know the Outer-thirds set is in one-to-one correspondence with $[0,1]$—as per §2.3.3, if you take your binary zeros and ones representing a point in $[0,1]$ and reread it as trinary zeros and twos you get a point on the Outer-thirds Set. It's not about information content, it's about how we distribute it in space.

We're going to use a simpler hook: how many cubes do you need to make a bigger cube?

- Start with a 1×1 square, and cover it with $\frac{1}{2}\times\frac{1}{2}$ squares. You need four of them, or 2^2.
- If you had squares of size $\frac{1}{3}\times\frac{1}{3}$, you would need nine squares to cover the 1×1 square with your tic-tac-toe grid, which we can write as 3^2 sub-squares. In general, you need n^2 squares of size $\frac{1}{n}\times\frac{1}{n}$ to cover the unit square. It's right there in the notation: $\frac{1}{n}\times\frac{1}{n}\times n^2 = 1$
- This works in three dimensions too. You can fill a unit cube with n^3 cubes of size $\frac{1}{n}$. Your diagram for this is a Rubik's Cube: if it were entirely made of cubes, it would be a $3\times3\times3$

block of cubes with sides $\frac{1}{3}$ the length of the whole. At museum gift shops, you may have seen "advanced" Cubes which, if made only of cubes, would be a $4 \times 4 \times 4$ grid of cubes with sides $\frac{1}{4}$ the length of the whole.

Here, we're taking advantage of one of the favorite features of a fractal: self-similarity. If you look at the left leg of the snowflake (after infinite iterations) and expand it to full size, you get the same snowflake. Same with a cube, as above: take a sub-cube, zoom in, and you have the original cube again. (Does that make a cube a fractal? Almost, but we'll see below that it doesn't quite fit the name.)

Summarizing the process, we have a shrink-and-clone procedure: Start with a line, square, or cube of length, volume or area (let's just call it *measure*) one, then shrink it down to $\frac{1}{n}^{\text{th}}$ the size on each dimension, so in d dimensions shrink it down by $\frac{1}{n}^{d}$. If you need m of these shrunk-down objects to recover the original measure of one, then the equation is $m\frac{1}{n}^{d} = 1$.

- We almost made it to the end of the book without solving any equations, but at least this is an easy one. Taking the log of everything to turn multiplication into addition and exponentiation into multiplication, we go from $m\frac{1}{n}^{d} = 1$ to $\ln(m) - d\ln(n) = 0$.
- Solving for d, we get $d = \ln(m)/\ln(n)$. That is, the dimension is the log of the shrink factor divided by the log count of sub-cubes. I didn't list remembering how logs work as a requisite for this book, so if you've forgotten, thanks for your perseverance here. If you do remember, note that the base is irrelevant because we're only looking at a ratio: $\ln(m)/\ln(n) = \log_{10}(m)/\log_{10}(n), \forall m, n > 0$.
- For example, our 2-D tic-tac-toe grid had a shrink factor of $n = 3$, and $m = 9$ cubes, and yes, $\ln(9)/\ln(3)$ is exactly two, because $\ln(3^2) = 2\ln(3)$.

For the snowflake, if you look at what happened to the leftmost third of the original line, you get a shrunk-down version of the original, at a shrink factor of $n = 3$. You need $m = 4$ of those to replicate the original, though, because after the first step there were four line segments of length equal to the leftmost third of the original line. And not just after the first step: at any point along all the shrinks-and-duplicates, the part grown from the left leg is a one-third shrinking of the whole. Applying the formula, we get a dimension of $\ln(4)/\ln(3) \approx 1.262$.

- This is a fractional dimension, and yes, that is where we get the word *fractal* from. This is why the cube, with its whole, non-fractional dimension, doesn't qualify as a fractal, despite such perfect self-similarity.

For the Outer-thirds Set, we shrink by a factor of three, and make two copies: $\ln(2)/\ln(3) \approx 0.631$, exactly one half the dimension of the snowflake (because $\ln(4) = \ln(2^2) = 2\ln(2)$).

For the space-filling curve, there were also four equal segments after the first step (left, up, down, right), so we need $m = 4$ replications of the miniaturized second-step version, and that miniaturized version was shrunk down by a factor of $n = 2$. Applying the formula, the dimension is $\ln(4)/\ln(2) = \ln(2^2)/\ln(2) = 2$. Yes, the space-filling curve has dimension two.

Just as we can cross $[0, 1]$ with itself to form a square, what if we cross the Outer-thirds set with itself, with one Outer-thirds Set along the x-axis and one along the y? (Because the linear version is called the Cantor Set, you can look up Cantor Dust to read more about the square version.) In the first step of the generation, we would take $[0, 1] \times [0, 1]$ and remove a middle-third vertical band and a middle-third horizontal band, leaving four squares in the corners, which like the cells of the tic-tac-toe grid are a $\frac{1}{3}$ shrink of the original. Replace each block with its four corners; repeat infinitely. Doing the shrink-and-duplicate math, $\ln(4)/\ln(3) \approx 1.262$. Yes, this is exactly the calculation we did for the snowflake; the square-ified Outer-thirds Set and the snowflake have exactly the same dimension.

10.3 Dimension via covering

We have a concept of fractional dimension which seems coherent for the usual cubes and for fractals generated via shrink-and-duplicate. That isn't the universe of shapes. Ever since the 1967 article "How long is the coast of Britain? Statistical self-similarity and fractional dimension" [Mandelbrot, 1967], the standard example has been measuring the coastline of an island. At the level of the typical road map, the coastline is a sequence of straight lines which add up to some length, but a topographical map might go into more detail about the curves along the coast adding up to a longer length, and if you had photos you would see the rocks forming the coastline, and if you had a magnifying glass you would see the jagged edges of the rocks extending the total length still more. It's not an especially well-formed question, as the length of the coastline expands and contracts with the tides, but one could tell the same story of more solid objects, say the surface of an irregular crystal (and to a chemist, everything that doesn't move is a crystal). Next time you make almond bread cookies (German: *Mandelbrot*), take a look at the bottom of your cookie, which will be a mostly flat rough pattern of points where margarine melted away or sugar crystals congealed. For the coastline of the island or the crystal, there is no shrink-and-duplicate procedure which will give us a measurement, but we can generate one using coverings. Though I've worked diligently in this book to avoid anything with real-world application, the procedure below is something which could actually be implemented to find the dimension of real-world complex surfaces.

- In its introduction, Mandelbrot [1967] directly cites Steinhaus [1954] and his discussion of "the paradox of length", where the coastline that has different lengths on different maps is the left bank of the Vistula River in Poland. Steinhaus: "The same difficulty arises when measuring such objects as contours of leaves or perimeters of plane sections of trees: the result depends appreciably on the precision of the instruments employed."
- Steinhaus [1954] readily traces a thread of authors observing some variant of the paradox of length back to 1812, though "in this chain no author is thoroughly conscious of his debt towards his immediate predecessor."

We'll warm up with two simple definitions.

Definition 10.1. ***Diameter of a set*** S
The length of the longest line you can draw inside the set. Write as $diam(S)$

For a circle, this is the diameter as usual; for a square, the diagonal; for more exotic shapes it may take a moment to find the longest interior line segment, but it is always there and well-defined. Note also that it is always a linear, 1-D distance, so the same math applies whether our covering sets below are 0-D or 2-D or 50-D. From the definition of *diameter* we can build another definition:

Definition 10.2. δ-covering of S
A covering of the set S where the largest diameter of any of the elements of the covering is δ.

Now that you've warmed up, it's time to confront our covering measure of a set (Hausdorff), a multi-layered mess to be explicated below:

$$\mathcal{H}^d(S) = \lim_{\delta \to 0} \min_{\text{coverings}} \{\Sigma_{s \in \delta-\text{covering}} diam(s)^d\}.$$

This expression is best understood from the inside out.

- At the core, we have $diam(s)^d$, the diameter of an element of the covering to the d^{th} power, where d will be the dimension.
- Sum that over all sets in the δ-covering.
- The *min* indicates that, given a fixed δ, we need to search all possible δ-coverings and find the one with the smallest value of the sum. Don't use just any covering, find the most efficient one.
- Now that you did this for one value of δ, repeat that process for an infinite sequence of δs approaching zero. As $\delta \to 0$, the number of elements in the covering will go to infinity (unless you are trying to cover only a finite number of points).

We took d as given, but if we're trying to determine what the dimension of a set is, then we need to try all possible ds. How are we supposed to pick amongst them? It turns out—and this is going to be one of those remarkable little pieces of magic—that one will be picked for us.

- You at least don't have to rearrange coverings every time: If you found a sum-minimizing covering for d, then it will still be a sum-minimizing covering for any other positive dimension candidate, say c or e where $c < d < e$.
- Say that for d, $\mathcal{H}^d(S)$ is finite and nonzero. Apply the standard algebraic form $x^e = x^d x^{(e-d)}$ from the basic Algebra textbooks, then use the fact that the maximum diameter in any δ-covering is δ:

$$diam(s)^e = diam(s)^d \cdot diam(s)^{(e-d)} < diam(s)^d \delta^{(e-d)}.$$

- We generated a factor that is common to every element in the core of the calculation of $\mathcal{H}^e(S)$, so that factors out of the sum:

$$\Sigma_s diam(s)^e < \Sigma_s \delta^{(e-d)} diam(s)^d = \delta^{(e-d)} \Sigma_s diam(s)^d.$$

- As $\delta \to 0$, that gives us an upper bound of zero for the thing we are minimizing, meaning as $\delta \to 0$, $\mathcal{H}^e(S) \to 0$.
- Similarly, if $c < d$ and $\mathcal{H}^d(S)$ is nonzero but finite, then all the same math happens, but with $c - d$ negative, the $\delta^{(c-d)}$ factor is better written as $\left(\frac{1}{\delta}\right)^{(d-c)}$. Then we now have a lower bound,

$$diam(s)^c = diam(s)^d \cdot diam(s)^{(c-d)} > diam(s)^d / \delta^{(d-c)},$$

  and as $\delta \to 0$, $\frac{1}{\delta} \to \infty$, so our lower bound will give us $\mathcal{H}^c(S) = \infty$.

It doesn't get any finickier than that: we have exactly one value of d where $\mathcal{H}^d(S)$ is greater than zero but less than infinite. This awkward construction, then, captures the fundamental weirdness of dimension-finding I'd mentioned above, the Goldilocks complaint that covering a 2-D square with 1-D lines takes an infinite number of them, and covering a square with a 3-D cube requires only a zero-measure slice. We're going to call that unique value of d which produces a nonzero-but-finite value of $\mathcal{H}^d(S)$ the dimension of S.

There are two straightforward properties we can say about $\mathcal{H}^d(S)$.

- If you have two non-overlapping sets S_1 and S_2, then $\mathcal{H}^d(S_1 \cup S_2) = \mathcal{H}^d(S_1) + \mathcal{H}^d(S_2)$, because for any small enough δ, the δ-coverings for S_1 and for S_2 don't interact, and the sum in the definition of $\mathcal{H}^d(S_1 \cup S_2)$ simply becomes a sum over both coverings.
- If you found the δ-covering for a set S, then create a new set S_n by shrinking S by $\frac{1}{n}$, then your sum $\Sigma_s diam(s)^d$, shrinks by $\frac{1}{n}^d$. That holds for any value of δ, so the entire calculation of $\mathcal{H}^d(S_n)$ is identical to the calculation of $\mathcal{H}^d(S)$, except multiplied by $\frac{1}{n}^d$ everywhere.

Combining them, we can do another shrink-and-duplicate. If you can replicate your original set S by shrinking it by a factor n, then making m copies of the miniature version, then $\frac{m}{n^d}\mathcal{H}^d(S_n) = \mathcal{H}^d(S)$. If $\mathcal{H}^d(S_n)$ is zero or infinity, this equation collapses to $0 = 0$ or $\infty = \infty$. But at that sweet spot where that value is finite and nonzero, it cancels from both sides of the equation and we get $\frac{m}{n^d} = 1$. If you flip back to the subsection just before the covering measure was introduced, you'll find exactly the same relationship between shrink factor, duplication, and dimension.

The process of finding the unique value of d that makes the covering dimension calculation finite via $\mathcal{H}^d$ required finding an infinite sequence of covers, then finding the exact value of d for which the above calculation is nonzero and finite. We don't have to do any of that, though, if we can do a shrink-and-duplicate on the set we are measuring, because whatever value of d works must satisfy the very same equation $\frac{m}{n^d} = 1$, or $d = \ln(m)/\ln(n)$. It's a weird multi-layer object, but $\mathcal{H}^d$ always picks out exactly one number out of all of $\mathbb{R}$ to assign as the dimension of a given space, and that one number matches all of our expectations to this point.

But because the covering method matches the shrink-and-duplicate method but doesn't rely on it, the covering method can be applied to any set—we may have come upon a practical application. The story tracks with the story of using a ruler on a series of coastline maps of finer and finer resolution: we have a sequence of easily measured coverings, one at each scale, and those will approach some well-defined limit with well-defined properties.

In fact, measuring coastlines has become something of a pastime; see, e.g., Husain et al. [2021] for a long list of references. It is usually done via box counting, which is the computer-friendlier version of the $\mathcal{H}$-measure above, where we restrict our open sets to be a grid of squares. With that, there is some range of smoothnesses, from weather-beaten South Africa with dimension 1.02, to continental China=1.09, to Australia=1.13, to the fjords of Norway=1.52.

Chapter 11

Conclusion

Thank you for reading. I know there's a lot you could be doing with your time, and I hope the hours you spent getting to know point-set Topology were worthwhile.

The intro promised a close analysis of Zeno's paradox-like questions of what it means for something to be smooth, for two distinct points to join into a larger whole. We started with Definition 2.2: an open set is whatever set of points we want, plus finite intersections and arbitrary unions (and X and $\emptyset$), which decidedly does not seem like enough to say much of anything. But using only that definition and occasional mappings from open sets in one space to open sets in another, the topologists over-delivered on conceptualizations of connection, with the concept of limit points, a concept of cross-topology smoothness (i.e., continuity), three types of connectivity, nine types of set-within-set T-axioms, compactness to guarantee convergence, and a side-mention of density. And you've seen a wealth of examples showing that these are all different concepts. And none of these methods relied on the traditional method of somehow defining infinitely small distances. It's not closeness that matters, but connection.

What's next?

There's an incredible online catalog of topologies at `topology.pi-base.org` [pi Base Community], a grid of hundreds of topologies versus dozens of properties each may or may not have. If you need still more counterexamples or still more concepts of how things connect, there's your checklist. And if you're looking for trouble, mixed in with the oft-told stories I retold here, you'll find an abundance of question marks in the grid, properties which just haven't been definitively checked for a given topology. Maybe you can be the one to check one more off the list.

As of this writing, Wikipedia does a remarkably good job with Topology, and if you've read this far, you can understand the language, vocabulary, and general mindset of the pages on Topology there. If you're the sort of person who goes down Wikipedia rabbit holes, the ones about tori are now yours to fall into.

By contrast, generative AI chatbots, as of this writing, are terrible at Topology and in my occasional experimentation I have yet to get a fully correct answer from a chatbot about any of the subjects covered in this book (though a chatbot did give me first-draft code for producing some of the figures). Such systems work by putting together sentences that are most probable, and Topology is too full of improbable outcomes and subtle differences. After everybody else has lost their job to a robot, Topologists will be the last workers standing.

The chapter on manifolds (§6) was a taking-off point to Algebraic Topology, which links together topics from Abstract Algebra—mostly groups, not algebras!—and links them to paths on the odd shapes we constructed. Because it is concerned with paths, it focuses on path-connected spaces (§5.5), typically on the Usual Topology, as folded and twisted as per our pastings in §6. If you know group theory, seeing that every group maps one-to-one to a weird curve on a weird surface is a blast, but I'm assuming you're a reader who hasn't (yet) read any group theory. Now's your chance, though, and it may be just as fun going from curves on weird surfaces to abstract groups.

Regardless, I hope you'll keep reading. If you have no previous experience reading math, well, now you do, and the language and means of thinking is remarkably standard across math books. There's a linguistic disagreement between Brits, who refer to algebra, arithmetic, &c. as *maths*, and Yanks, who call the whole of it *math*. I'm decidedly on the U.S. side of this one: there is one unified mathematics, and you can immerse yourself in the same questions regardless of the point from which you first approach them.

Bibliography

Edwin Abbott Abbott. *Flatland: A Romance of Many Dimensions*. Seeley & Co., London, 1884.

David Auburn. *Proof*. Faber & Faber, London, England, July 2001.

Jordan Ellenberg. *Shape*. Penguin, New York, NY, May 2022.

Harry Furstenberg. On the infinitude of primes. 1955. doi: 10.2307/2307043.

T W Gamelin and Robert E Greene. *Introduction to Topology*. Dover Publications, second edition, 1999.

Martin Gardner. *The Unexpected Hanging and Other Mathematical Diversions*. University of Chicago Press, Chicago, IL, September 1991.

Leonard Gillman and Meyer Jerison. *Rings of Continuous Functions*. 1960. doi: 10.1007/978-1-4615-7819-2.

Douglas R Hofstadter. *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books, London, England, 1979.

Akhlaq Husain, Jaideep Reddy, Deepika Bisht, and Mohammad Sajid. Fractal dimension of coastline of Australia. 2021. doi: 10.1038/s41598-021-85405-0.

Elizabeth Landau and Ben Klemens. Overlooked no more: James sakoda, whose wartime internment inspired a social science tool. May 2023. URL `https://t.co/X4RS0wVdl4`.

Benoit Mandelbrot. How long is the coast of Britain? Statistical self-similarity and fractional dimension. 1967. doi: 10.1126/science.156.3775.636.

Maynard J Mansfield. *Introduction to Topology*. D Van Nostrand, 1963.

Ton Marar. *A Ludic Journey into Geometric Topology*. Springer, Cham, Switzerland, September 2022.

James R Munkres. *Topology*. Pearson, Upper Saddle River, NJ, second edition, December 1999.

The pi Base Community. π-base, a community database of topological counterexamples. URL `https://topology.pi-base.org/`.

Siobhan Roberts. Conway's cat. 2021. doi: 10.1007/s00283-021-10064-y.

Walter Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Professional, New York, NY, February 1976.

A B Sossinsky. *Knots*. Harvard University Press, London, England, December 2002.

Michael Starbird and Francis Su. *Topology Through Inquiry*. AMS/MAA Textbooks. American Mathematical Society, Providence, RI, December 2019.

L A Steen and J A Seebach. *Counterexamples in Topology*. Springer, New York, NY, second edition, September 1978.

Hugo Steinhaus. Length, shape and area. *Colloquium Mathematicum*, 3(1):1–13, 1954. URL `http://eudml.org/doc/210004`.

William P. Thurston. On proof and progress in mathematics. 1994. doi: 10.48550/ARXIV.MATH/9404236.

David Foster Wallace. *Everything and More: A Compact History of Infinity*. Great Discoveries. WW Norton, New York, NY, October 2003.