3 Limits of Sequences and Filters

The Axiom of Choice is obviously true, the well-ordering theorem is obviously false; and who can tell about Zorn's Lemma?
—Jerry Bona (Schechter, 1996)

Introduction. Chapter 2 featured various properties of topological spaces and explored their interactions with a few categorical constructions. In this chapter we'll again discuss some topological properties, this time with an eye toward more fine-grained ideas. As introduced early in a study of analysis, properties of nice topological spaces X can be detected by sequences of points in X. We'll be interested in some of these properties and the extent to which sequences suffice to detect them. But take note of the adjective "nice" here. What if X is any topological space, not just a nice one? Unfortunately, sequences are not well suited for characterizing properties in arbitrary spaces. But all is not lost. A sequence can be replaced with a more general construction—*a filter*—which is much better suited for the task. In this chapter we introduce filters and highlight some of their strengths.

Our goal is to spend a little time inside of spaces to discuss ideas that may be familiar from analysis. For this reason, this chapter contains less category theory than others. On the other hand, we'll see in section 3.3 that filters are a bit like functors and hence like generalizations of points. This perspective thus gives us a coarse-grained approach to investigating fine-grained ideas. We'll go through some of these basic ideas—closure, limit points, sequences, and more—rather quickly in sections 3.1 and 3.2. Later in section 3.2 we'll see exactly why sequences don't suffice to detect certain properties in all spaces. We'll also discover those "nice" spaces for which they do. Section 3.3 introduces filters with some examples and results about them. These results include the claim that filters, unlike sequences, do suffice to characterize certain properties. Finally, in section 3.4 we'll use filters to share a delightfully short proof of Tychonoff's theorem.

3.1 Closure and Interior

Here are a few basic definitions, which may be familiar from analysis. Given any subset B of a space X, two topological constructions suggest themselves. There is the *closure* $\overline{B}$ which is the smallest closed set containing B, and there is the *interior* B° which is the largest open set contained in B. When $\overline{B} = X$, we say B is *dense* in X. If $\left(\overline{B}\right)^\circ = \varnothing$, we say B is *nowhere dense*.

Notice that the definition of a topology guarantees that the interior and closure exist. For example, because a topology is closed under arbitrary unions, the interior B° is precisely the union of all open subsets of B. Contrast this with the ideas of a "largest closed set" contained in B and a "smallest open set" containing B, which might not exist.

A point x is called a *limit point* of a set B if every open set around x contains a point of $B \smallsetminus \{x\}$. The closure $\overline{B}$ consists of B together with all of its limit points. A point x is called a *boundary point* of B if every open set containing x contains both a point in B and a point in the complement of B.

Limit points help to understand closures and interiors. So, taking a cue from analysis, let's turn to a study of *sequences* in an attempt to characterize limit points.

3.2 Sequences

Definition 3.1 Let X be a topological space. A *sequence* in X is a function $x\colon \mathbb{N} \to X$. We usually write x_n for $x(n)$ and may denote the sequence by $\{x_n\}$. A sequence $\{x_n\}$ *converges* to $z \in X$ if and only if for every open set U containing z, there exists an $N \in \mathbb{N}$ so that if $n \geq N$ then $x_n \in U$. When $\{x_n\}$ converges to $z \in X$ we'll write $\{x_n\} \to z$. A *subsequence* of a sequence x is the composition xk where $k\colon \mathbb{N} \to \mathbb{N}$ is an increasing injection. We'll write x_{k_i} for $xk(i)$ and denote the subsequence by $\{x_{k_i}\}$.

Here are a few examples.

Example 3.1 Let $A = \{1, 2, 3\}$ with the topology $\mathcal{T} = \{\varnothing, \{1\}, \{1, 2\}, A\}$. The constant sequence $1, 1, 1, 1, \ldots$ converges to 1. It also converges to 2 and to 3.

Example 3.2 Consider $\mathbb{Z}$ with the cofinite topology. For any $m \in \mathbb{Z}$, the constant sequence $m, m, m, \ldots$ converges to m and only to m. Indeed if $l \neq m$, then the set $\mathbb{R} \smallsetminus m$ is an open set around l containing no elements of the sequence.

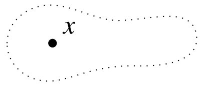
However, the sequence $\{n\} = 1, 2, 3, 4, \ldots$ converges to m for every $m \in \mathbb{Z}$. To see this, let m be any integer, and let U be a neighborhood of m. Since $\mathbb{Z} \smallsetminus U$ is finite, there can only be finitely many natural numbers in $\mathbb{Z} \smallsetminus U$. Let N be larger than the greatest natural number in $\mathbb{Z} \smallsetminus U$. Then $n \in U$ whenever $n \geq N$, proving that $\{n\} \to m$.

Example 3.3 Consider $\mathbb{R}$ with the usual topology. If $\{x_n\} \to x$, then $\{x_n\}$ does not converge to any number $y \neq x$. To prove it, note we can always find disjoint open sets U and V with $x \in U$ and $y \in V$. (We can be explicit if necessary: $U = (x - c, x + c)$ and $V = (y - c, y + c)$ where $c = \frac{1}{2}|x - y|$.) Then there is a number N so that $x_n \in U$ for all $n \geq N$. Since $U \cap V = \varnothing$, the set V cannot contain any x_n for $n \geq N$, and hence $\{x_n\}$ does not converge to y.

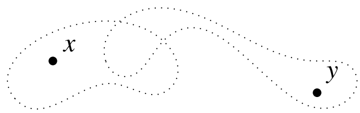
As we'll see below, sequences can be used to detect certain properties of spaces, subsets of spaces, and functions between spaces. But before continuing, it will be helpful to first introduce a couple more topological properties—two of the so-called "separation" axioms.

Definition 3.2 We say

(i) A topological space X is T_0 if and only if for every pair of points $x, y \in X$ there exists an open set containing one, but not both, of them.



(ii) A topological space X is T_1 if and only if for every pair of points $x, y \in X$ there exist open sets U and V with $x \in U$, $y \in V$ with $x \notin V$ and $y \notin U$.



We could have added a third property to the list. A space X with the property that for every pair of points $x, y \in X$ there exist open sets U and V with $x \in U$, $y \in V$ with $U \cap V = \varnothing$ is sometimes called T_2, but we've already named the property *Hausdorff*, after Felix Hausdorff who originally used the axiom in his definition of "neighborhood spaces" (Hausdorff and Aumann, 1914). Note that T_0, T_1, and T_2 all define topological properties.

Here are a few theorems about sequences that might evoke familiar results from analysis. Some of the proofs are left as exercises. You'll want to keep the examples above in mind.

Theorem 3.1 A space X is T_1 if and only if for any $x \in X$, the constant sequence $x, x, x, \ldots$ converges to x and only to x.

Proof. Suppose X is T_1 and $x \in X$. It's clear that $x, x, x, \ldots \to x$. Let $y \neq x$. Then there exists an open set U with $y \in U$ and $x \notin U$. Therefore, $x, x, x, \ldots$ cannot converge to y.

For the converse, suppose X is not T_1. Then there exist two distinct points x and y for which every open set around y contains x. Thus $x, x, x, \ldots \to y$. □

Theorem 3.2 If X is Hausdorff, then sequences in X have at most one limit.

Proof. Let X be Hausdorff and suppose $\{x_n\}$ is a sequence such that $\{x_n\} \to x$ and $\{x_n\} \to y$. If $x \neq y$, then there are disjoint open sets U and V with $x \in U$ and $y \in V$. Since $\{x_n\} \to x$ there is a number N so that $x_n \in U$ for all $n \geq N$. Since $\{x_n\} \to y$ there is a number K so that $x_n \in U$ for all $n \geq K$. Let $M = \max\{N, K\}$. Since $M \geq N$ and $M \geq K$ we have $x_M \in U$ and $x_M \in V$, contradicting the fact that U and V are disjoint. □

Theorem 3.3 If $\{x_n\}$ is a sequence in A that converges to x, then $x \in \overline{A}$.

Proof. Exercise. □

Theorem 3.4 If $f\colon X \to Y$ is continuous, then for all sequences $\{x_n\}$ that converge to x in X, the sequence $\{fx_n\}$ converges to fx in Y.

Proof. Exercise. □

You'll notice that theorem 3.1 is an if-and-only-if theorem that characterizes the T_1 property with a statement about sequences. So you might wonder if sequences are also enough to characterize Hausdorff spaces, closed sets, and continuous functions. That is, do theorems 3.2, 3.3, and 3.4 have if-and-only-if-versions, too? The answer is no.

Example 3.4 Sequences don't suffice to detect Hausdorff spaces. Consider $\mathbb{R}$ with the cocountable topology. This space is not Hausdorff, and yet convergent sequences have unique limits.

Example 3.5 Sequences don't suffice to detect closed sets. Let

$$X = [0,1]^{[0,1]} := \{\text{functions } f\colon [0,1] \to [0,1]\}$$

with the product topology, and let A be the subset of X consisting of functions whose graphs are "sawtooths" with vertices on the x axis at a finite number of points $\{0, r_1, \ldots, r_n, 1\}$ and spikes of height 1, as in figure 3.1. The zero function is in $\overline{A}$, but there is no sequence $\{f_n\}$ in A converging to it.

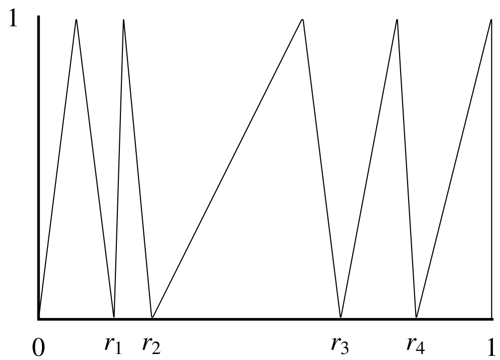


Figure 3.1 A sawtooth function

Example 3.6 Sequences don't suffice to detect continuous functions. Let $X = [0,1]^{[0,1]} :=$ $\{\text{functions } f\colon [0,1] \to [0,1]\}$ with the product topology, and let Y be the subspace of X consisting of integrable functions. The function $I\colon Y \to \mathbb{R}$ defined by $If = \int_0^1 f$ is not a continuous function but $\{If_n\} \to If$ whenever $\{f_n\} \to f$.

Sequences don't suffice to detect the Hausdorff property, closure, and continuity in these examples because the spaces in question have too many open sets around each point for their topological properties to be adequately probed by sequences. But for spaces without too many open sets around each point, sequences do suffice to characterize their properties. These spaces are called *first countable.*

Definition 3.3 Let X be a space. A collection of open sets $\mathcal{B}$ is called a *neighborhood base* for $x \in X$ if and only if for every open set O containing x, there exists an open set $U \in \mathcal{B}$ with $x \in U \subseteq O$. A space X is called *first countable* if and only if every point has a countable neighborhood base. A space X is called *second countable* if and only if it has a countable basis.

Example 3.7 By definition, the set $\mathcal{T}_x$ of open neighborhoods of a point x in a space X is a neighborhood base for x.

Example 3.8 Every metric space is first countable since the open balls around x of radius $1, \frac{1}{2}, \frac{1}{3}, \ldots$ form a countable neighborhood base.

Example 3.9 An n-dimensional *manifold* is a second countable Hausdorff topological space with the property that every point has a neighborhood homeomorphic to $\mathbb{R}^n$.

Two examples of nonfirst countable spaces were given in example 3.4, example 3.5, and example 3.6; namely, $\mathbb{R}$ with the cocountable topology and $[0,1]^{[0,1]}$ with the product topology. But in first countable spaces such as metric spaces, sequences *do* suffice to characterize separation, closure, and continuity properties. In other words, theorem 3.2, theorem 3.3, and theorem 3.4 do have if-and-only-if versions in this special context.

Theorem 3.5 Let X be a first countable space. Then X is Hausdorff if and only if every sequence has at most one limit.

Proof. Suppose that X is first countable. If X is not Hausdorff, there exist points x and y that cannot be separated by open sets. Let $U_1, U_2, \ldots$ be a neighborhood base of x and $V_1, V_2, \ldots$ be a neighborhood base for y. For every n choose a point $x_n \in U_n \cap V_n \neq \varnothing$. The sequence $\{x_n\}$ has a subsequence that converges to x and to y. □

Theorem 3.6 Let X be a first countable space and let $A \subseteq X$. A point $x \in \overline{A}$ if and only if there exists a sequence $\{x_n\}$ in A with $\{x_n\} \to x$.

Proof. Exercise. □

Theorem 3.7 Suppose X and Y are first countable. A function $f\colon X \to Y$ is continuous if and only if for every sequence $\{x_n\}$ in X with $\{x_n\} \to x$, the sequence $\{f x_n\} \to f x$.

Proof. Exercise. □

The reason sequences characterize separation, closure, and continuity in first countable spaces but not in arbitrary spaces is simply because sequences are countable. If, however, we want the previous three if-and-only-if theorems to hold in a wider context, then we'll want a generalization of sequences—one that more accurately reflects the range of possibilities for convergence. This generalization is the topic of the next section. To help introduce the ideas, let's make a simple observation.

Given a sequence $\{x_n\}$ in a space X, consider the collection of sets the sequence is eventually inside of:

$$\mathcal{E}_{x_n} := \{A \subseteq X \mid \text{there exists an } N \text{ so that } x_n \in A \text{ for all } n \geq N\} \tag{3.1}$$

Now here's the key:

> The sequence $\{x_n\}$ converges to a point x if and only if the neighborhood base $\mathcal{T}_x$ is contained in $\mathcal{E}_{x_n}$.

Understanding convergence, therefore, amounts to understanding the set $\mathcal{E}_{x_n}$. This immediately suggests how one might attempt to generalize sequences: abstract the notion of "the set of sets the sequence is eventually inside of." Cartan (1937b) did just that in 1937, when he introduced an appropriate generalization of sequences well suited to studying convergence: *filters*. As we'll soon see, filters are just what's needed to obtain the three analogous if-and-only-if theorems for *all* spaces, not just first countable ones.

3.3 Filters and Convergence

A filter is like an algebraic road map with (perhaps rough) directions to points or places in a space. More precisely, it is a certain subset of a poset (a partially ordered set). In this chapter, the poset we focus on is the powerset of a set X. That is, we consider *filters of subsets of a given set*. Here is the definition.

Definition 3.4 A *filter* on a set X is a collection $\mathcal{F} \subseteq 2^X$ that is

(i) downward directed: $A, B \in \mathcal{F}$ implies there exists $C \in \mathcal{F}$ such that $C \subseteq A \cap B$

(ii) nonempty: $\mathcal{F} \neq \varnothing$

(iii) upward closed: $A \in \mathcal{F}$ and $A \subseteq B$ implies $B \in \mathcal{F}$

An additional property is often useful:

(iv) proper: there exists $A \subseteq X$ such that $A \notin \mathcal{F}$

So a filter on X is a downward directed, nonempty, upward closed subset of the powerset 2^X. A couple of comments are in order. Being downward directed and upward closed implies that filters are closed under finite intersections. We can also rephrase properness as the requirement that $\varnothing \notin \mathcal{F}$. For example, 2^X is itself a filter and is well named the *improper filter* on X. A set that is only downward directed and nonempty is called a *filterbase*. Any filterbase generates a filter; just take the upward closure of the base.

In chapter 0, it was noted that any poset can be viewed as a category; objects are elements in the set, and morphisms are provided by the partial order. So there's hope that filters have a categorical description. Indeed they do. It starts with the observation that the poset 2^X has the property that every pair of elements $A, B \in 2^X$ has a *meet* (a greatest lower bound), namely their intersection $A \cap B$. We can define another poset also having this property. Consider the two-element poset $2 := \{0 \leq 1\}$. For $a, b \in 2$ define their meet $a \wedge b$ by

$$0 \wedge 0 = 0 \qquad 0 \wedge 1 = 0 \qquad 1 \wedge 0 = 0 \qquad 1 \wedge 1 = 1$$

Every monotone function $f\colon 2^X \to 2$ that respects this structure—that is, that satisfies $f(A \cap B) = fA \wedge fB$—defines a filter, namely the preimage $f^{-1}1$. Verifying this claim is a simple exercise. In the language of order theory, f is called a *meet-semilattice homomorphism*. In the language of category theory, it is called a *continuous functor*.

Indeed, the posets 2^X and 2 are categories, and a functor between them is precisely a monotone function. We'll see in chapter 4 that a meet is an example of a more general categorical construction called a *limit*, and a functor that respects limits is, with inspiration from theorem 3.4, called *continuous*. Filters thus arise from continuous functors $2^X \to 2$.

Example 3.10 For any set X, the *trivial filter* $\mathcal{F} = \{X\}$ is a proper filter. More generally, for any nonempty subset $A \subseteq X$, the set of all sets containing A is a proper filter.

Example 3.11 Another example of a proper filter is the *eventuality filter* $\mathcal{E}_{x_n}$ associated to the sequence $\{x_n\}$ from equation 3.1. It's aptly named since it's the collection of all sets that the sequence eventually remains in.

Example 3.12 The cofinite subsets of a set X

$$\mathcal{F} := \{A \subseteq X \mid X \smallsetminus A \text{ finite}\}$$

define a filter called the *Fréchet filter*. If X is infinite, then the Frećhet filter is proper.

Example 3.13 Given a topological space $(X, \mathcal{T})$, the open neighborhoods $\mathcal{T}_x$ of a point x form a filterbase, though they generally do not form a filter. The reason is simply that (usually) not every set containing an open neighborhood of x is open. But there is some ambiguity in the mathematics community on this point. Kelley (1955) defines "a neighborhood of x" to mean exactly "a set containing an open set containing x." Others, such as Munkres (2000), prefer all neighborhoods be open neighborhoods. In our language, the filterbase $\mathcal{T}_x$ generates the filter of (not necessarily open) neighborhoods of x.

We began our discussion about filters with an observation about convergence: a sequence in a topological space converges if and only if its eventuality filter contains the filterbase $\mathcal{T}_x$. This motivates the following definition.

Definition 3.5 A filter $\mathcal{F}$ on a topological space $(X, \mathcal{T})$ *converges* to x if and only if $\mathcal{F}$ refines $\mathcal{T}_x$, that is if $\mathcal{T}_x \subseteq \mathcal{F}$. When $\mathcal{F}$ converges to x we'll write $\mathcal{F} \to x$.

Example 3.14 The real-valued sequence $\{x_n\} := \{1, -1, \frac{1}{2}, -1, \frac{1}{4}, -1, \frac{1}{8}, \ldots\}$ does not converge, whereas the subsequence $\{x_{2n}\} = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\}$ does. We can see this by reasoning with eventuality filters, which isn't so different from reasoning with sequences. Here's the thing to notice: the eventuality filter $\mathcal{E}_{x_{2n}}$ is the set of all subsets $A \subset \mathbb{R}$ for which there exists an N so that $\frac{1}{2^n} \in A$ for all $n \geq N$. It's straightforward to check that $\mathcal{E}_{x_{2n}} \to 0$. However, the eventuality filter $\mathcal{E}_{x_n}$ has the same description as $\mathcal{E}_{x_{2n}}$, except each A must also include -1. So $\mathcal{E}_{x_n} \subseteq \mathcal{E}_{x_{2n}}$.

This example illustrates that passing to a subsequence increases the size of an eventuality filter since the membership condition is weaker. At the extreme end of this, the improper filter $\mathcal{F} = 2^X$ converges to *every point* in X! (And yet it is not the eventuality filter of any sequence.)

At this juncture, you'll recall our earlier claim that filters suffice to give a characterization of the Hausdorff property, closure, and continuity. The proofs of the first two are now available.

Theorem 3.8 A space is Hausdorff if and only if limits of convergent proper filters are unique.

Proof. Suppose X is Hausdorff and that a proper filter $\mathcal{F}$ converges to both x and y with $x \neq y$. Then there are open neighborhoods U of x and V of y with $U \cap V = \varnothing$. By convergence, $U, V \in \mathcal{F}$. Since $\mathcal{F}$ is a filter, $\varnothing = U \cap V \in \mathcal{F}$, which contradicts $\mathcal{F}$ being proper.

However, if X is not Hausdorff then there are two distinct points x and y that cannot be separated by open sets. Let

$$\mathcal{B} = \{U \cap V \mid x \in U, y \in V \text{ with } U \text{ and } V \text{ open}\}$$

Note that $\mathcal{B}$ is downward directed and nonempty, so it is a filterbase. The filter $\mathcal{F}$ generated by the collection $\mathcal{B}$ converges to both x and y. □

Theorem 3.9 Let X be a space with $A \subseteq X$. A point $x \in \overline{A}$ if and only if there exists a proper filter $\mathcal{F}$ containing A with $\mathcal{F} \to x$.

Proof. First recall that $x \in \overline{A}$ if and only if every neighborhood of x nontrivially intersects A or equivalently if and only if the filterbase $\mathcal{B} = \{U \cap A\}_{U \in \mathcal{T}_x}$ does not contain the empty set. So if $x \in \overline{A}$, then simply generate a proper filter from $\mathcal{B}$. Conversely, if there is a proper filter $\mathcal{F}$ that converges to x and contains A, then $\mathcal{B} \subseteq \mathcal{F}$ and so $\mathcal{B}$ cannot contain the empty set. □

In the next theorem we'll show that filters also suffice to detect continuity. But first we need to understand functions in the context of filters.

Definition 3.6 Given a filter $\mathcal{F}$ on X and a function $f\colon X \to Y$, the set $\{fA \mid A \in \mathcal{F}\}$ of images of elements of $\mathcal{F}$ form a filterbase. The filter $f_*\mathcal{F}$ generated by this base is the *pushforward of* $\mathcal{F}$ *with respect to* f. Explicitly:

$$f_*\mathcal{F} := \{B \subseteq Y \mid \text{there exists } A \in \mathcal{F} \text{ with } fA \subseteq B\}$$

In this definition, the "generated by" is necessary since the set of images itself may not form a filter. For example, if f is not surjective, then the images don't contain Y and therefore cannot be upward closed.

Example 3.15 As a simple example, $f_*\mathcal{E}_{x_n} = \mathcal{E}_{fx_n}$. In other words, the pushforward of the eventuality filter of a sequence is the eventuality filter of the pushforward of that sequence.

Now here's the desired theorem.

Theorem 3.10 A function $f\colon X \to Y$ is continuous if and only if for every filter $\mathcal{F}$ on X, if $\mathcal{F} \to x$, then $f_*\mathcal{F} \to fx$.

Proof. Let $\mathcal{F}$ be a filter on X with $\mathcal{F} \to x$, and suppose $f\colon X \to Y$ is continuous. We want to show $\mathcal{T}_{fx} \subseteq f_*\mathcal{F}$; that is, for any $B \in \mathcal{T}_{fx}$ there exists a set $A \in \mathcal{F}$ with $fA \subseteq B$. So choose $A = f^{-1}B$. Continuity implies $f^{-1}\mathcal{T}_{fx} \subseteq \mathcal{T}_x$, which means $A \in \mathcal{T}_x$. The statement $\mathcal{F} \to x$ means $\mathcal{T}_x \subseteq \mathcal{F}$ and so $A \in \mathcal{F}$.

Conversely, suppose that whenever $\mathcal{F} \to x$ we have $f_*\mathcal{F} \to fx$ for any filter $\mathcal{F}$. Take $\mathcal{F}$ to be the filter generated by $\mathcal{T}_x$ to find that $f_*\mathcal{F} \to fx$, which means $\mathcal{T}_{fx} \subseteq f_*\mathcal{F}$. Thus for every $B \in \mathcal{T}_{fx}$, there exists a set A in $\mathcal{T}_x$ with $fA \subseteq B$. This proves that f is continuous. □

Filters, therefore, do indeed give the triad of theorems for all spaces.

all spaces (with filters)		*first countable spaces (with sequences)*
Theorem 3.8	**Hausdorff**	Theorem 3.5
Theorem 3.9	**closure**	Theorem 3.6
Theorem 3.10	**continuity**	Theorem 3.7

Having reached our goal, you might expect the chapter to conclude here. But not so fast. There's much more to filters. We've used them to study convergence, thereby promoting some "if, then" theorems about sequences to "if and only if" theorems about filters. But filters also shine particularly bright in discussions of compactness. To illustrate, we will use filters in the next section to provide a wonderfully short proof of Tychonoff's theorem, which was introduced—but not proven—in chapter 2.

3.4 Tychonoff's Theorem

The goal of this section is to prove the following theorem.

Tychonoff's Theorem 2 Given any collection $\{X_\alpha\}_{\alpha\in A}$ of compact spaces, the product $\prod_{\alpha\in A} X_\alpha$ is compact.

It's easier to prove that the product of finitely many compact spaces is compact than it is to prove the general case. For example, in Munkres' *Topology* (2000), compactness is introduced in chapter 3, where it is proven that the product of finitely many compact spaces is compact (Theorem 26.7). For the proof of the general case, the intrepid reader must wait until chapter 5 (Theorem 37.3), with a full chapter on countability and separation interrupting. Schaum's Outline (Lipschutz, 1965) states Tychonoff's theorem in chapter 12, but the proof is banished to the exercises. One must use the axiom of choice (or its equivalent) to prove the general case (see our theorem 3.14).

We will present a variation on Cartan's proof (1937a) by way of a little more filter technology. That technology is a particular kind of filter called an *ultrafilter*, which we introduce next. We'll take a leisurely stroll through the ideas, pointing out notable results along the way. In a grand finale, Tychonoff's theorem is proven in a few short lines in section 3.4.2.

3.4.1 Ultrafilters and Compactness

An *ultrafilter* is simply a filter that is *maximal*. The terms are synonymous.

Definition 3.7 A proper filter on a set is an *ultrafilter* if and only if it is not properly contained in any other proper filter.

This definition is second order since it deals with a quantification over subsets of a set. In practice, we'd like to work with a first-order definition—a characterization of an ultrafilter that doesn't require us to compare it to all other filters. Happily, such a characterization exists.

Proposition 3.1 A filter $\mathcal{U}$ on a set X is an ultrafilter if and only if for every subset $A \subseteq X$ the following condition holds: $A \notin \mathcal{U}$ if and only if there exists $B \in \mathcal{U}$ with $A \cap B = \varnothing$.

Proof. Let $\mathcal{U}$ be an ultrafilter. Then $A \notin \mathcal{U}$ if and only if the filter generated by $\mathcal{U} \cup \{A\}$ is the powerset 2^X. Since the generated filter consists of all sets containing an intersection of the form $B \cap A$ for some $B \in \mathcal{U}$, this is equivalent to the statement that the empty set contains $B \cap A$ for some $B \in \mathcal{U}$. Since the empty set is a subset of any set, the result follows.

Conversely, suppose $\mathcal{U}$ is a filter on X satisfying the condition, and let $\mathcal{F}$ be a filter properly containing $\mathcal{U}$. So there is at least one $A \in \mathcal{F}$ which is not in $\mathcal{U}$. By hypothesis, there must also exist a $B \in \mathcal{U}$ disjoint from A. But $\varnothing = A \cap B \in \mathcal{F}$, and so $\mathcal{F}$ is the improper filter by upward closure. $\square$

Here's a nonexample followed by an example.

Example 3.16 If X has more than one point, then the *trivial filter* $\{X\}$ is not an ultrafilter.

Example 3.17 Given any x in a set X, the *principal filter at* x defined as $\{A \subseteq X \mid x \in A\}$ is an ultrafilter.

The next example highlights a point we wish to emphasize. In this chapter, we defined filters on powersets, but the definition of "downward directed, nonempty, upward closed" makes perfect sense in *any* poset. Keep this in mind. Filters in more general posets are useful, natural objects of study. This is illustrated well by *Riemann integration*.

Example 3.18 Let $f\colon [a, b] \to \mathbb{R}$ be a bounded function and let $(\mathcal{P}, \leq)$ be the poset of partitions of $[a, b]$ ordered by refinement. Any partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ yields two real numbers:

$$u_P := \textstyle\sum_{i=1}^n \sup\,(f|_{[x_{i-1},x_i]})(x_i - x_{i-1}) \qquad l_P := \textstyle\sum_{i=1}^n \inf\,(f|_{[x_{i-1},x_i]})(x_i - x_{i-1})$$

Similarly, any filter of partitions in $\mathcal{P}$ defines a pair of filters on $\mathbb{R}$:

$$\mathcal{B}_u(\mathcal{F}) = \{U \subseteq \mathbb{R} \mid \text{ there exists } Q \in \mathcal{F} \text{ such that } Q \leq P \text{ implies } u_P \in U\}$$

and

$$\mathcal{B}_l(\mathcal{F}) = \{U \subseteq \mathbb{R} \mid \text{ there exists } Q \in \mathcal{F} \text{ such that } Q \leq P \text{ implies } l_P \in U\}$$

For any ultrafilter $\mathcal{U}$ in $\mathcal{P}$, the filters $\mathcal{B}_u(\mathcal{U})$ and $\mathcal{B}_l(\mathcal{U})$ converge to real numbers. The function f is Riemann integrable if and only if they converge to the same value. In such a case, that value is called the integral $\int_a^b f$. So ultrafilters and the natural ordering of partitions allow us to replace the typical "partition norms" or "meshes" with a direct handling of the underlying orders. (As is often the case with categorical constructions, filters are reflective of what we have, not necessarily the computation of what we want.)

So just remember that a thorough discussion of filters can—and does—exist outside the context of powersets. But the powerset context is a very nice one: filters on powersets have special properties not shared by more general filters (see exercise 3.13 at the end of the chapter). In particular, because our filters are on powersets, the property of maximality is equivalent to another property—primality. In other words, ultrafilters on X are equivalent to *prime* filters on X. This reformulation will give us a simple characterization of compactness and ultimately a clean proof of Tychonoff's theorem.

Definition 3.8 A filter $\mathcal{F}$ on a set X is *prime* if and only if it is proper and if for all $A, B \subseteq X$,

$$A \cup B \in \mathcal{F} \text{ implies } A \in \mathcal{F} \text{ or } B \in \mathcal{F}$$

Theorem 3.11 A filter on X is maximal if and only if it is prime.

Proof. Suppose $\mathcal{F}$ is an ultrafilter on X and fails to be prime. Then there are $A, B \subseteq X$ such that $A \cup B \in \mathcal{F}$, but neither A nor B are in $\mathcal{F}$. The latter holds if and only if there exist sets $A', B' \in \mathcal{F}$ with $A \cap A' = \varnothing = B \cap B'$. This implies the intersection $(A \cup B) \cap (A' \cup B')$ is empty, which is true if and only if $A \cup B \notin \mathcal{F}$, a clear contradiction.

Now suppose $\mathcal{F}$ is prime but not a maximal. Then $\mathcal{F}$ is properly contained in a proper filter $\mathcal{G}$. So there exists a nonempty $A \in \mathcal{G}$ with $A \notin \mathcal{F}$. Notice that $X \setminus A \notin \mathcal{F}$, for otherwise $X \setminus A \in \mathcal{G}$ and if G contains both A and $X \setminus A$, then $\mathcal{G}$ would not be proper. But $A \cup (X \setminus A) = X \in \mathcal{F}$, contradicting the hypothesis that $\mathcal{F}$ is prime. □

Below, we'll use this theorem to give a succinct characterization of compact spaces. But despite the theorem, it will be good to keep a distinction between prime and maximal in our minds. There are several reasons why. First, as alluded to above, prime and maximal are not equivalent in more general settings. Keeping the two distinct in our minds strengthens our intuition. Second, a priori prime filters are difficult to construct. But as we'll soon see, we can always consider a maximal extension of a proper filter and then use the fact that ultrafilters are prime. Finally, the theorems proven below are naturally phrased in terms of prime filters because images commute with unions and therefore prime filters pushforward. Keeping the distinction thus makes proving theorems easier: if you need a prime filter, just construct one by extending a proper filter. If your construction involves pushing a filter forward, then prime filters are your friend. Now, en route to compactness let's prove our claim that every proper filter can be extended to a maximal one. We'll call on *Zorn's lemma*, whose statement we recount here.

Zorn's Lemma If every chain in a nonempty poset P has an upper bound, then P has a maximal element.

The Ultrafilter Lemma Every proper filter is contained in an ultrafilter.

Proof. Any set of filters $\{\mathcal{F}_\alpha\}_{\alpha \in A}$ is bounded above by the filter generated by finite intersections of elements of the $\{\mathcal{F}_\alpha\}$. When the set is a chain of proper filters, this upper bound is itself proper. Given a proper filter $\mathcal{F}$, chains of proper filters containing $\mathcal{F}$ thus have proper upper bounds. By Zorn's Lemma, there is a maximal filter containing $\mathcal{F}$. □

Corollary 3.11.1 Any infinite set has a non-principal ultrafilter.

Proof. Consider the Fréchet filter $\mathcal{F} := \{A \subseteq X \mid X \setminus A \text{ is finite}\}$ and appeal to the Ultrafilter Lemma to extend $\mathcal{F}$ to an ultrafilter $\mathcal{U}$. Were $\mathcal{U}$ to contain any finite set, it would contain its (cofinite) complement; hence $\varnothing \in \mathcal{U}$, which contradicts that $\mathcal{U}$ is a proper filter. □

To appreciate this result, think back to example 3.17. When pressed, it's hard to come up with ultrafilters that are not principal. The fact that any *nonprincipal ultrafilters* exist is not at all obvious. To produce one, we needed to appeal to the Ultrafilter Lemma. The Ultrafilter Lemma also gives the promised characterization of compactness.

Theorem 3.12 A space X is compact if and only if every prime filter converges.

Proof. Suppose $\mathcal{F}$ is a prime filter that fails to converge to any $x \in X$. Equivalently, suppose for all x there exists $U_x \in \mathcal{T}_x - \mathcal{F}$. The set $\{U_x\}_{x \in X}$ is an open cover. By compactness, choose a finite subcover $\{U_{x_i}\}_{i=1}^n$. Then $U_{x_1} \cup \cdots \cup U_{x_n} = X \in \mathcal{F}$. By primality, there exists an i such that $U_{x_i} \in \mathcal{F}$, a contradiction.

Now suppose X is not compact. Choose a collection $\mathcal{V}$ of closed sets with the finite intersection property and empty intersection. Note that for all x there exists $V_x \in \mathcal{V}$ with $x \notin V_x$. Further, by the Ultrafilter Lemma, $\mathcal{V}$ is contained in an ultrafilter $\mathcal{U}$. However, $\mathcal{U} \not\rightarrow x$ for any x, for otherwise it would imply $\varnothing = V_x \cap V_x^c \in \mathcal{U}$ contradicting properness of $\mathcal{U}$. □

You'll recognize this theorem as a generalization of the Bolzano-Weierstrass theorem introduced in chapter 2. Restated, it says:

If X is compact, then every sequence has a convergent subsequence.

So recalling from example 3.14 that subsequences correspond to larger (eventuality) filters and that filters are good at promoting "if then" theorems to "if and only if" ones, you might hope for a promotion of Bolzano-Weierstrass for filters:

X is compact if and only if every proper filter is contained in a convergent proper filter.

As a result of theorem 3.12, this is indeed the case. Where we once used convergent subsequences, we now use prime filters. Another consequence of theorem 3.12 comes for free when paired with theorem 3.8.

Corollary 3.12.1 A space X is compact Hausdorff if and only if every prime filter has exactly one limit point.

This characterization of compact Hausdorff spaces is the beginning of a long categorical tale. Sharing the full story would take us too far off course, so instead we'll tell an abridged version. It starts with the fact that ultrafilters define a functor from the category Set to itself, a consequence of this next theorem.

Theorem 3.13 Let $\mathcal{U}$ be an ultrafilter on X and let $f: X \to Y$. The pushforward $f_*\mathcal{U}$ is an ultrafilter on Y.

Proof. Exercise. □

Since the pushforward of an ultrafilter is an ultrafilter, the assignment $\beta\colon \mathsf{Set} \to \mathsf{Set}$ that sends a set X to βX, the set of ultrafilters on X, defines a functor. For a morphism $f: X \to Y$ of sets, the function $\beta f: \beta X \to \beta Y$ sends an ultrafilter to its pushforward. In the special case when X is a compact Hausdorff space, every ultrafilter on X converges to exactly one point. So you might wonder whether the assignment $\alpha\colon \beta X \to X$ that sends an ultrafilter to its unique limit point is of any interest. It is. It's the key to unlocking an important categorical statement:

> The category of compact Hausdorff spaces is equivalent to the category of algebras for the ultrafilter monad.

What's the ultrafilter monad? And what does it mean to be an algebra for one? We won't get into the details, but we will try to give you an idea about what the statement means. Principal filters play a principal role: since the pushforward of a principal filter is principal, they assemble into a natural transformation $\eta : \mathrm{id}_{\mathsf{Set}} \to \beta$ defined by

$$\eta_X(x) = P_x$$

where P_x is the principal filter at $x \in X$. There is another natural transformation $\mu\colon \beta \circ \beta \to \beta$ that comes into play. We will not describe μ except to say that it works like a kind of multiplication, and the natural transformation η behaves like a unit for this multiplication. The triple (β, η, μ) defines something called a *monad*. (It should remind you of a *monoid*, which also consists of three things: a set X, an associative multiplication map $m\colon X \times X \to X$, and an element $u\colon * \to X$ behaving as a unit for m.) Once you have a monad, you can define something called an algebra for that monad, and the algebras for a monad form a category. And *this* category, one can show, is equivalent to the category of compact Hausdorff spaces. Behind the curtain of this categorical connection between ultrafilters and compact Hausdorff spaces is the rich theory of *adjunctions*. For this reason, the story will naturally resurface later in this book: it's closely related to the discussion of the Stone-Čech compactification in chapter 5. For an introduction to monads, see Riehl (2016), and for a fuller account of the ultrafilter monad tale, see Manes (1969) and the article on compacta at the nLab (Stacey et al., 2019). The ambitious reader may further enjoy Leinster (2013).

After this leisurely stroll through ultrafilters, prime filters, Bolzano-Weierstrass, and monads, we are now ready to prove that which we set out to prove: Tychonoff's theorem.

3.4.2 A Proof of Tychonoff's Theorem

There is a conservation of difficulty in mathematics. One theorem may have many proofs, and more sophisticated tools will give more elegant proofs. Historically, the difficulties in Tychonoff's theorem were in finding the correct definition of the product topology and in characterizing compactness. By using ultrafilters to characterize compactness, we are using a sophisticated theoretical tool. The proof we share is correspondingly elegant (Chernoff, 1992).

If sequences had been sufficient for discussions of convergence, then we could use a ready-made argument. Just use continuity of projection maps from the product to push a sequence forward. In each factor, pass to a subsequence, which converges by Bolzano-

Weierstrass. Then in the product, form a subsequence by taking the indices common to all the convergent subsequences in the projections, and conclude this sequence converges. Here's the thrill: by replacing sequences with filters and doing the work needed to develop some understanding of the theory, this ready-made argument becomes a genuine proof.

Proof of Tychonoff's theorem. Let $\{X_\alpha\}_{\alpha\in A}$ be a family of compact spaces, define $X := \prod_{\alpha\in A} X_\alpha$, and let $\mathcal{F}$ be an ultrafilter on X. We must show that $\mathcal{F}$ converges.

The pushforward of an ultrafilter is an ultrafilter and, since the X_α are compact, there exists x_α such that $(\pi_\alpha)_*\mathcal{F} \to x_\alpha$. So by definition, for all open neighborhoods U of x_α, there exists $B \in \mathcal{F}$ with $\pi_\alpha B \subseteq U$. Equivalently, $B \subseteq \pi_\alpha^{-1}U$, and so $\pi_\alpha^{-1}U \in \mathcal{F}$. And every open neighborhood of $(x_\alpha)_{\alpha\in A} \in X$ is a union of finite intersections of the $\pi_\alpha^{-1}U$. Therefore, $\mathcal{F} \to (x_\alpha)_{\alpha\in A}$. □

We'll close this chapter with the remark that any treatment of Tychonoff's theorem requires some machinery from set theory. In our presentation, we hid the machinery in Zorn's lemma, which we used to prove the Ultrafilter Lemma. The reason set theory is unavoidable is because Tychonoff's theorem is equivalent to the axiom of choice. Without digressing too much, we'd like to give you some idea of why *Zermelo-Frankel-Choice* is equivalent to *Zermelo-Frankel-Tychonoff*.

3.4.3 A Little Set Theory

In any imaginable proof that the Tychonoff theorem implies the axiom of choice, one begins with an arbitrary collection of sets and then creates a collection of compact topological spaces. The compactness of the product leads to the existence of a choice function. In 1950, Kelley proved that the Tychonoff theorem implies the axiom of choice (Kelley, 1950) using augmented cofinite topologies. Here, we give an easier proof. First, we recall the axiom of choice.

The Axiom of Choice For any collection of nonempty sets $\{X_\alpha\}_{\alpha\in A}$, the product $\prod_{\alpha\in A} X_\alpha$ is nonempty.

Theorem 3.14 Tychonoff's theorem is equivalent to the axiom of choice.

Proof. We used Zorn's lemma to prove Tychonoff's theorem. Although we don't prove it, the axiom of choice implies Zorn's lemma (see exercise 3.14 at the end of the chapter), from which it follows that Tychonoff's theorem is implied by the axiom of choice.

To prove that Tychonoff's theorem implies the axiom of choice, let $\{X_\alpha\}_{\alpha\in A}$ be a collection of nonempty sets. We need to make a bunch of compact spaces so we can apply the Tychonoff theorem. First, add a new element to X_α called "∞_α," letting $Y_\alpha = X_\alpha \cup \{\infty_\alpha\}$. Each set Y_α can be made into a space by defining the topology to be $\{\varnothing, \{\infty_\alpha\}, X_\alpha, Y_\alpha\}$. Note that Y_α is compact—there are only finitely many open sets so every open cover is finite. Thus by Tychonoff's theorem, $Y := \prod_{\alpha\in A} Y_\alpha$ is compact.

Now consider a collection of open sets $\{U_\beta\}_{\beta\in A}$ of Y where U_β is the basic open set in Y obtained by taking the product of all Y_αs for $\alpha \neq \beta$ and putting the open set $\{\infty_\beta\}$ in the βth factor. Notice that any finite subcollection $\{U_{\beta_1}, \ldots, U_{\beta_n}\}$ cannot cover Y, for the function f defined as follows is not in $\cup_{i=1}^{n} U_{\beta_i}$. Choose a partial function $\overline{f} \in \prod_{i=1}^{n} X_{\beta_i}$, which is possible without the axiom of choice since the product is finite. Then extend $\overline{f}$ to a function $f \in Y$ by setting $f(\alpha) = \infty_\alpha$ for all $\alpha \neq \beta_1, \ldots, \beta_n$, which is possible since we're not making any choices.

Therefore, the collection $\{U_\beta\}$ cannot cover Y. So there is a function $f \in Y$ not in the $\cup_{\alpha\in A} U_\alpha$. This says that for no $\alpha \in A$ does $f\alpha = \infty_\alpha$. Therefore, $f\alpha \in X_\alpha$ for each α, which is a desired choice function. $\square$

Exercises

1. Suppose A is a subspace of X. We say a map $f\colon A \to Y$ can be *extended* to X if there is a continuous map $g\colon X \to Y$ with $g = f$ on A.

a) Prove that if A is dense in X and Y is Hausdorff, then f can be extended to X in at most one way.

b) Give an example of spaces X and Y, a dense subset A, and a map $f\colon A \to Y$ such that f can be extended to X in more than one way.

c) Give an example of spaces X and Y, a dense subset A, and a map $f\colon A \to Y$ such that f cannot be extended.

2. Prove that $\mathbb{R}$ with the cocountable topology (sets with countable complement are open) is a non-Hausdorff space in which convergent sequences have unique limits.

3. Check all the details of example 3.14.

4. Check all the details of example 3.5.

5. Nets are an earlier generalization of sequences introduced by Moore and Smith (1922); Moore (1915) used to address the insufficiency of sequences. This exercise demonstrates that nets and proper filters are formally interchangeable.

Definition 3.9 A *net* is a function $\eta\colon S \to X$ whose domain is a directed set.

A *directed set* is defined to be a pair $(S, \leq)$, where S is a set and $\leq$ is a relation on S satisfying:

- for all $s \in S$, $s \leq s$,
- for all $s, t, u \in S$, $s \leq t$ and $t \leq u$ imply $s \leq u$,
- for all $s, t \in S$, there exists $u \in S$ with $s \leq u$ and $t \leq u$.

We say that η *converges* to x if and only if its eventuality filter

$$\mathcal{E}_\eta := \{A \subseteq X \mid \text{ there exists } t \in S \text{ such that } \ s \geq t \text{ implies } \eta s \in A\}$$

contains $\mathcal{T}_x$ in which case we write $\eta \to x$.

a) A sequence is an example of a net. Show that a subsequence of a sequence is a subnet, but not all subnets of a sequence are subsequences. For an interesting example, use the family of sawtooth functions from example 3.5 whose corners have rational coordinates.

b) Given a proper filter $\mathcal{F}$, let $\mathcal{D} := \{(A, a) \in 2^X \times X \mid a \in A \in \mathcal{F}\}$. Show that $\mathcal{D}$ is directed by the relation $(A, a) \leq (B, b)$ if and only if $B \subseteq A$.

c) Let $\pi_\mathcal{F}\colon \mathcal{D} \to X$ be the net sending $(A, a) \mapsto a$. Prove that $\mathcal{E}_{\pi_\mathcal{F}} = \mathcal{F}$.

d) Conclude that $\pi_\mathcal{F} \to x$ if and only if $\mathcal{F} \to x$.

6. Check all the details of example 3.6.

7. **Pushforward of Filters:** Given $f\colon X \to Y$ and a filter $\mathcal{F}$ on X, prove that the set

$$\{B \subseteq Y \mid \text{ there exists } A \in \mathcal{F} \text{ such that } \ fA \subseteq B\}$$

is a filterbase.

8. Prove theorem 3.13.

9. Here are two variations of Hausdorff. Call a space KC if all its compact sets are closed. Call a space US if the limits of convergent sequences are unique. Prove that Hausdorff implies KC implies US, but that the implications are strict (Wilansky, 1967).

10. Show that a countable intersection of open dense sets in a complete metric space is dense. (This is called the *Baire category theorem*.)

11. Let X be a compact space and let $\{f_n\}$ be an increasing sequence in $\mathsf{Top}(X, \mathbb{R})$. Prove that if $\{f_n\}$ converges pointwise, then $\{f_n\}$ converges uniformly.

12. Verify that the following definition of filters in posets specializes to the definition we gave for filters on sets.

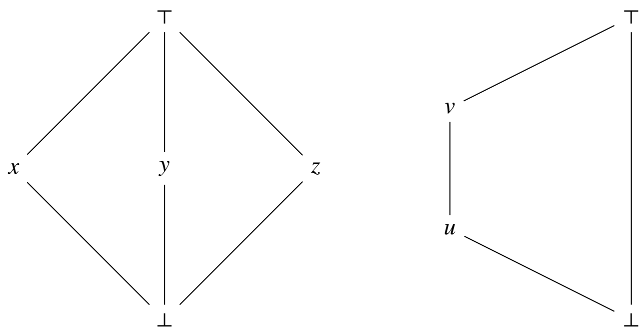
Definition 3.10 A *filter in a poset* $(\mathcal{P}, \preceq)$ is a collection $\mathcal{F} \subseteq \mathcal{P}$ such that it is:

Downward directed: $a, b \in \mathcal{F}$ implies there exists $c \in \mathcal{P}$ such that $c \preceq a$ and $c \preceq b$,

Nonempty: $\mathcal{F}$ is nonempty,

Upward closed: $a \in \mathcal{F}$ and $a \preceq b$ implies $b \in \mathcal{F}$.

13. Following up on our discussion of prime and maximal filters, consider the following pair of lattices:



Find a maximal filter that is not prime in the lattice on the right and a filter that is prime but not maximal in the lattice on the left.

Note: Lattices without any sublattice isomorphic to one of these satisfy the distributive property $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$. And in particular, we may conclude that for distributive lattices, maximal filters are prime by recycling our proof substituting unions for joins and intersections for meets.

14. Transfinite Induction: Axiom of choice implies Zorn's lemma

Ordinals are set theory's response to the question of how things may be ordered "one after another." Consequently, they form the setting in which induction may be defined.

Definition 3.11 A *well ordering* on a set S is a linear (or total) order "$\leq$" in which every nonempty subset has a least element. Together with order preserving—that is, monotone—functions, well-orders form a category. *Ordinals* are defined to be the isomorphism classes of objects in this category. Following von Neumann, we associate to each ordinal $[\alpha]$ a representative well-ordering:

$$\alpha := \{ \text{ ordinals } \beta < \alpha\}$$

and will feel free to refer to "the" ordinal α as this representative well-order.

Being well-ordered amounts to having two pieces of information: there's always a starting point and at every element there is an unambiguously defined next element—precisely the information needed to carry out induction. It's worth looking at the first few familiar ordinals to get a feel for them:

Example 3.19 First observe that there is a least ordinal, typically called 0, namely the initial well-order $\varnothing$ with the empty relation. This is the seed from which we may freely generate ordinals.

Names	*Representatives*	*Orders*
0	$\varnothing$	
1	$\{0\}$	0
2	$\{0,1\}$	$0 \to 1$
$\vdots$	$\vdots$	$\vdots$
ω	$\mathbb{N}$	$0 \to 1 \to \cdots$
$\omega+1$		$0 \to 1 \to \cdots \omega$
$\omega+2$		$0 \to 1 \to \cdots \omega \to \omega+1$
$\vdots$	$\vdots$	$\vdots$

Some caution is in order. Note that the underlying sets of, say, ω and $\omega + 1$ are in bijection, but as orders they are distinct. For example, the ordinal $\omega + 1$ has a nonzero element that is not the immediate successor of any other element. Such elements are called *limit ordinals* and play an important role in the theory of ordinals. Second, keep in mind that not every total order is a well-order; the rationals are not well ordered, for example. Finally, it is tempting to guess that the ordinals are themselves well ordered. After all, they are linearly ordered by "is an initial segment of," and in any set of ordinals, there is a least element. However, this leads to the *Burali-Forti paradox*: were Ω the *set* of all ordinals, then it's well ordered by our above comments and hence is (order isomorphic to) an ordinal. So $\Omega < \Omega$, contradicting trichotomy—we are forced to conclude that the collection of all ordinals is not a set.

(i) Use the Burali-Forti paradox to prove that there are ordinals of arbitrarily large cardinality. (Hint: try contradiction.)

(ii) Demonstrate the following: to prove a property $P(-)$ holds for all ordinals, it is sufficient to demonstrate that:

- *Base Case*: $P(0)$ holds.
- *Successor Step*: $P(\alpha)$ implies $P(\alpha + 1)$.
- *Limit Step*: Given limit ordinal λ, for all $\alpha < \lambda$ such that $P(\alpha)$ holds, then $P(\lambda)$ holds.

(iii) Using the following strategy, prove that the axiom of choice implies Zorn's lemma:

Given the axiom of choice and a nonempty partially ordered set $(\mathcal{P}, \leq)$ in which every chain has an upper bound, for each $a \in \mathcal{P}$ define

$$E_a := \{b \in \mathcal{P} \mid a < b\}$$

There are two cases: if there is some a for which $E_a = \varnothing$, then a is maximal and we are done; otherwise, the axiom of choice guarantees that there is a function $f : \mathcal{P} \to \mathcal{P}$ such that $fa \in E_a$ for all a.

Use transfinite induction to prove that there are chains of all ordinal lengths in $\mathcal{P}$. State a contradiction and conclude that the axiom of choice implies Zorn's lemma.