CHARALAMBOS D. ALIPRANTIS
KIM C. BORDER

# Infinite Dimensional Analysis

## A Hitchhiker's Guide

Third Edition

Springer

Charalambos D. Aliprantis
Kim C. Border

# Infinite Dimensional Analysis

## A Hitchhiker's Guide

Third Edition

With 38 Figures
and 1 Table

Springer

Professor Charalambos D. Aliprantis
Department of Economics
Krannert School of Management
Rawls Hall, Room 4003
Purdue University
100 S. Grant Street
West Lafayette IN 47907-2076
USA
E-mail: aliprantis@mgmt.purdue.edu

Professor Kim C. Border
California Institute of Technology
Division of the Humanities and Social Sciences
228–77
1200 E. California Boulevard
Pasadena CA 91125
USA
E-mail: kcborder@caltech.edu

Cataloging-in-Publication Data
Library of Congress Control Number: 2006921177

ISBN-10 3-540-32696-0 3rd ed. Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32696-0 3rd ed. Springer Berlin Heidelberg New York
ISBN 3-540-65854-8 2nd ed. Springer Berlin Heidelberg New York



Springer is a part of Springer Science+Business Media
springeronline.com


Printed in Germany



Cover design: Erich Kirchner
Production: Helmut Petri
Printing: Strauss Offsetdruck

SPIN 11681052 Printed on acid-free paper – 42/3153 – 5 4 3 2 1 0

# Contents

# A foreword to the practical

## Why use infinite dimensional analysis?

Why should practical people, such as engineers and economists, learn about infinite dimensional spaces? Isn't the world finite dimensional? How can infinite dimensional analysis possibly help to understand the workings of real economies?

Infinite dimensional models have become prominent in economics and finance because they capture natural aspects of the world that cannot be examined in finite dimensional models. It has become clear in the last couple of decades that economic models capable of addressing real policy questions must be both stochastic and dynamic. There are fundamental aspects of the economy that static models cannot capture. Deterministic models, even chaotically deterministic models, seem unable to explain our observations of the world.

Dynamic models require infinite dimensional spaces. If time is modeled as continuous, then time series of economic data reside in infinite dimensional function spaces. Even if time is modeled as being discrete, there is no natural terminal period. Furthermore, models including fiat money with a terminal period lead to conclusions that are not tenable. If we are to make realistic models of money or growth, we are forced to use infinite dimensional models.

Another feature of the world that arguably requires infinite dimensional modeling is uncertainty. The future is uncertain, and infinitely many resolutions of this uncertainty are conceivable. The study of financial markets requires models that are both stochastic and dynamic, so there is a double imperative for infinite dimensional models.

There are other natural contexts in which infinite dimensional models are natural. A prominent example is commodity differentiation. While there are only finitely many types of commodities actually traded and manufactured, there are conceivably infinitely many that are not. Any theory that hopes to explain which commodities are manufactured and marketed and which are not must employ infinite dimensional analysis. A special case of commodity differentiation is the division of land. There are infinitely many ways to subdivide a parcel of land, and each subdivision can be regarded as a separate commodity.

Let us take a little time to briefly introduce some infinite dimensional spaces commonly used in economics. We do not go into any detail on their properties here—indeed we may not even define all our terms. We introduce these spaces now as a source of examples. In their own way each of these spaces can be thought of as an infinite dimensional generalization of the finite dimensional Euclidean space $\mathbb{R}^n$, and each of them captures some salient aspects of $\mathbb{R}^n$.

## Spaces of sequences

When time is modeled as a sequence of discrete dates, then economic time series are sequences of real numbers. A particularly important family of sequence spaces is the family of $\ell_p$-spaces. For $1 \leqslant p < \infty$, $\ell_p$ is defined to be the set of all sequences $x = (x_1, x_2, \ldots)$ for which $\sum_{n=1}^{\infty} |x_n|^p < \infty$. The the **$\ell_p$-norm** of the sequence $x$ is the number $\|x\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}$.

As $p$ becomes larger, the larger values of $x_n$ tend to dominate in the calculation of the $\ell_p$-norm and indeed, $\lim_{p\to\infty} \|x\|_p = \sup\{|x_n|\}$. This brings us to $\ell_\infty$. This space is defined to be the set of all real sequences $x = (x_1, x_2, \ldots)$ satisfying $\sup\{|x_n|\} < \infty$. This supremum is called the **$\ell_\infty$-norm** of $x$ and is denoted $\|x\|_\infty$. This norm is also called the **supremum norm** or sometimes the **uniform norm**, because a sequence of sequences converges uniformly to a limiting sequence in $\ell_\infty$ if and only if it converges in this norm.

All of these spaces are vector spaces under the usual (pointwise) addition and scalar multiplication. Furthermore, these spaces are nested. If $p \leqslant q$, then $\ell_p \subset \ell_q$.

There are a couple of other sequence spaces worth noting. The space of all convergent sequences is denoted $c$. The space of all sequences converging to zero is denoted $c_0$. Finally the collection of all sequences with only finitely many nonzero terms is denoted $\varphi$. All of these collections are vector spaces too, and for $1 \leqslant p < \infty$ we have the following vector subspace inclusions:

$$\varphi \subset \ell_p \subset c_0 \subset c \subset \ell_\infty \subset \mathbb{R}^{\mathbb{N}}.$$

Chapter 16 discusses the properties of these spaces at length.

The space $\ell_\infty$ plays a major role in the neoclassical theory of growth. Under commonly made assumptions in the one sector growth model, capital/labor ratios are uniformly bounded over time. If there is an exhaustible resource in fixed supply, then $\ell_1$ may be an appropriate setting for time series.

## Spaces of functions

One way to think of $\mathbb{R}^n$ is as the set of all real functions on $\{1, \ldots, n\}$. If we replace $\{1, \ldots, n\}$ by an arbitrary set $X$, the set of all real functions on $X$, denoted $\mathbb{R}^X$, is a natural generalization of $\mathbb{R}^n$. In fact, sequence spaces are a special case of function spaces, where $X$ is the set of natural numbers $\{1, 2, 3, \ldots\}$. When $X$ has a topological structure (see Chapter 2), it may be acceptable to restrict attention to $C(X)$, the continuous real functions on $X$.

Function spaces arise in models of uncertainty. In this case $X$ represents the set of *states of the world*. Functions on $X$ are then state-contingent variables. In statistical modeling it is common practice to denote the set of states by $\Omega$ and to endow it with additional structure, namely a $\sigma$-algebra $\Sigma$ and a probability measure $\mu$. In this case it is natural to consider the $L_p$-spaces. For $1 \leqslant p < \infty$, $L_p(\mu)$ is defined to be the collection of all ($\mu$-equivalence classes of) $\mu$-measurable functions $f$ for which $\int_\Omega |f|^p \, d\mu < \infty$. (These terms are all explained in Chapter 11. It is okay to think of these integrals as $\int_0^1 |f(x)|^p \, dx$ for now.) The number $\|f\|_p = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p}$ is the **$L_p$-norm** of $f$. The $L_\infty$-norm is defined by

$$\|f\|_\infty = \text{ess sup} f = \sup \{t : \mu(\{x : |f(x)| \geqslant t\}) > 0\}.$$

This norm is also known as the **essential supremum** of $f$. The space $L_\infty$ is the space of all $\mu$-measurable functions with finite essential supremum. Chapter 13 covers the $L_p$-spaces.

## Spaces of measures

Given a vector $x$ in $\mathbb{R}^n$ and a subset $A$ of indices $\{1, \ldots, n\}$ define the set function $x(A) = \sum_{i \in A} x_i$. If $A \cap B = \varnothing$, then $x(A \cup B) = x(A) + x(B)$. In this way we can think of $\mathbb{R}^n$ as the collection of additive functions on the subsets of $\{1, \ldots, n\}$. The natural generalization of $\mathbb{R}^n$ from this point of view is to consider the spaces of measures or charges on an algebra of sets. (These terms are all defined in Chapter 11.) Spaces of measures on topological spaces can inherit some of the properties from the underlying space. For instance, the space of Borel probability measures on a compact metrizable space is naturally a compact metrizable space. Results of this sort are discussed in Chapters 12 and 15.

The compactness properties of spaces of measures makes them good candidates for commodity spaces for models of commodity differentiation. They are also central to models of stochastic dynamics, which are discussed in Chapter 19.

## Spaces of sets

Since set theory can be used as the foundation of almost all mathematics, spaces of sets subsume everything else. In Chapter 3 we discuss natural ways of topologizing spaces of subsets of metrizable spaces. These results are also used in Chapter 17 to discuss continuity and measurability of correspondences. The topology of **closed convergence** of sets has proven to be useful as a way of topologizing preferences and demand correspondences. Topological spaces of sets have also been used in the theory of incentive contracts.

## Prerequisites

The main prerequisite is what is often called "mathematical sophistication." This is hard to define, but it includes the ability to manipulate abstract concepts, and an understanding of the notion of "proof."

We assume that you know the basic facts about the standard model of the real numbers. These include the fact that between any two distinct real numbers there is a rational number and also an irrational number. (You can see that we already assume you know what these are. It was only a few centuries ago that this knowledge was highly protected.) We take for granted that the real numbers are complete. We assume you know what it means for sequences and series of real numbers to converge. We trust you are familiar with naïve set theory and its notation. We assume that you are familiar with arguments using induction. We hope that you are familiar with the basic results about metric spaces. Aliprantis and Burkinshaw [13, Chapter 1], Dieudonné [97, Chapter 3], and Rudin [292, Chapter 2] are excellent expositions of the theory of metric spaces. It would be nice, but not necessary, if you had heard of the Lebesgue integral; we define it in Chapter 11. We assume that you are familiar with the concept of a vector space. A good brief reference for vector spaces is Apostol [17]. A more detailed reference is Halmos [147].

## Chapter 1

# Odds and ends

One purpose of this chapter is to standardize some terminology and notation. In particular, Definition 1.1 defines what we mean by the term "function space," and Section 1.4 introduces a number of kinds of binary relations. We also use this chapter to present some useful odds and ends that should be a part of everyone's mathematical tool kit, but which don't conveniently fit anywhere else. We introduce correspondences and the notion of the evaluation duality. Our presentation is informal and we do not prove many of our claims. We also feel free to get ahead of ourselves and refer to definitions and examples that appear much later on.

We do prove a few theorems including Szpilrajn's Extension Theorem 1.9 for partial preorders, the existence of a Hamel basis (Theorem 1.8), and the Knaster–Tarski Fixed Point Theorem 1.10. These are presented as applications of Zorn's Lemma 1.7. Example 1.4 uses a standard cardinality argument to show that the lexicographic order cannot be represented by a numerical function.

We also try to present the flavor of the subtleties of modern set theory without actually proving the results. We do however prove Cantor's Diagonal Theorem 1.5 and describe Russell's Paradox. We mention some of the more esoteric aspects of the Axiom of Choice in Section 1.11 in order to convince you that you really do want to put up with it, and all it entails, such as non-measurable sets (Corollary 10.42). We also introduce the ordinals in Section 1.13.

## 1.1 Numbers

Leopold Kronecker is alleged to have remarked that, "God made the integers, all the rest is the work of man."[1] The **natural numbers** are $1, 2, 3, \ldots$, etc., and the set of natural numbers is denoted $\mathbb{N}$. (Some authors consider zero to be a natural number as well, and there are times we may do likewise.) We do not attempt to develop a construction of the real numbers, or even the natural numbers here. A very readable development may be found in E. Landau [221] or C. D. Aliprantis and O. Burkinshaw [13, Chapter 1].

[1] According to E. T. Bell [36, p. 477].

## 1.3 Relations, correspondences, and functions

Given two sets $X$ and $Y$, we can form the Cartesian product $X \times Y$, which is the collection of ordered pairs of elements from $X$ and $Y$. (We assume you know what ordered pairs are and do not give a formal definition.) A **relation** between members of $X$ and members of $Y$ can be thought of as a subset of $X \times Y$.[5] A relation between members of $X$ is called a **binary relation** on $X$. For a binary relation $R$ on a set $X$, that is, $R \subset X \times X$, it is customary to write $x \, R \, y$ rather than $(x, y) \in R$.

A near synonym for relation is **correspondence**, but the connotation is much different. We think of a correspondence $\varphi$ from $X$ to $Y$ as associating to each $x$ in $X$ a subset $\varphi(x)$ of $Y$, and we write $\varphi \colon X \twoheadrightarrow Y$. The **graph** of $\varphi$, denoted $\operatorname{Gr} \varphi$ is $\{(x, y) \in X \times Y : y \in \varphi(x)\}$. The space $X$ is the **domain** of the correspondence and $Y$ is the **codomain**. Given a subset $A \subset X$, the **image** $\varphi(A)$ of $A$ under $\varphi$ is defined by $\varphi(A) = \bigcup\{\varphi(x) : x \in A\}$. The **range** of $\varphi$ is the image of $X$ itself. We may occasionally call $Y$ the **range space** of $\varphi$. When the range space and the domain are the same, we say that a point $x$ is a **fixed point** of the correspondence $\varphi$ if $x \in \varphi(x)$. We have a lot more to say about correspondences in Chapters 17 and 18.

A special kind of relation is a **function**. A relation $R$ between $X$ and $Y$ is a function if $(x, y) \in R$ and $(x, z) \in R$ imply $y = z$. A function is sometimes called a **mapping** or **map**. We think of a function $f$ from $X$ into $Y$ as "mapping" each point $x$ in $X$ to a point $f(x)$ in $Y$, and we write $f \colon X \to Y$. We may also write $x \mapsto f(x)$ to refer to the function $f$. The **graph** of $f$, denoted $\operatorname{Gr} f$ is $\{(x, y) \in X \times Y : y = f(x)\}$. As with correspondences, the space $X$ is the **domain** of the function and $Y$ is the **codomain**. Given a subset $A \subset X$, the **image** of $A$ under $f$ is $f(A) = \{f(x) : x \in A\}$. The **range** of $f$ is the image of $X$ itself. When the range space and the domain are the same, we say that a point $x$ is a **fixed point** of the function $f$ if $x = f(x)$.

The graph of a function $f$ is also the graph of a singleton-valued correspondence $\varphi$ defined by $\varphi(x) = \{f(x)\}$, and vice versa. Clearly $f$ and $\varphi$ represent the same relation, but their values are not exactly the same objects.

A **partial function** from $X$ to $Y$ is a function from a subset of $X$ to $Y$. If $f \colon X \to Y$ and $A \subset X$, then $f|_A$ is the **restriction** of $f$ to $A$. That is, $f|_A$ has domain $A$, and for each $x \in A$, $f|_A(x) = f(x)$. We also say that $f$ is an **extension** of $f|_A$.

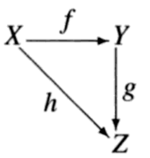
A function $x \colon \mathbb{N} \to X$, from the natural numbers to the set $X$, is called a **sequence** in the set $X$. The traditional way to denote the value $x(n)$ is $x_n$, and it is called the $n^{\text{th}}$ **term of the sequence**. Using an abused (standard) notation, we shall denote the sequence $x$ by $\{x_n\}$, and we shall consider it both as a function and

[5] Some authors, e.g., N. Bourbaki [62] and K. J. Devlin [91] pointedly make a distinction between a relation, which is a linguistic notion, and the set of ordered pairs that stand in that relation to each other, which is a set theoretic construct. In practice, there does not seem to be a compelling reason to be so picky.

as its range—a subset of $X$. A **subsequence** of a sequence $\{x_n\}$ is a sequence $\{y_n\}$ for which there exists a strictly increasing sequence $\{k_n\}$ of natural numbers (that is, $1 \leqslant k_1 < k_2 < k_3 < \cdots$) such that $y_n = x_{k_n}$ holds for each $n$.

The **indicator function** (or **characteristic function**) $\chi_A$ of a subset $A$ of $X$ is defined by $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$. The set of all functions from $X$ to $Y$ is denoted $Y^X$. Recall that the power set of $X$ is denoted $2^X$. This is also the notation for the set of all functions from $X$ into $2 = \{0, 1\}$. The rationale for this is that every subset $A$ of $X$ can be identified with its characteristic function $\chi_A$, which assumes only the values 0 and 1.

If $f : X \to Y$ and $g : Y \to Z$, the **composition** of $g$ with $f$, denoted $g \circ f$, is the function from $X$ to $Z$ defined by the formula $(g \circ f)(x) = g(f(x))$. We may also draw the accompanying sort of diagram to indicate that $h = g \circ f$. We sometimes say that this **diagram commutes** as another way of saying $h = g \circ f$.




More generally, for any two relations $R \subset X \times Y$ and $S \subset Y \times Z$, the composition relation $S \circ R$ is defined by

$$S \circ R = \{(x, z) \in X \times Z : \exists\, y \in Y \text{ with } (x, y) \in R \text{ and } (y, z) \in S\}.$$

A function $f : X \to Y$ is **one-to-one**, or an **injection**, if for every $y$ in the range space, there is at most one $x$ in the domain satisfying $y = f(x)$. The function $f$ maps $X$ **onto** $Y$, or is a **surjection**, if for every $y$ in $Y$, there is some $x$ in $X$ with $f(x) = y$. A **bijection** is a one-to-one onto function. A bijection may sometimes be referred to as a **one-to-one correspondence**. The **inverse image**, or simply **inverse**, of a subset $A$ of $Y$ under $f$, denoted $f^{-1}(A)$, is the set of $x$ with $f(x) \in A$. If $f$ is one-to-one, the inverse image of a singleton is either a singleton or empty, and there is a function $g : f(X) \to X$, called the **inverse** of $f$, that satisfies $x = g(y)$ if and only if $f(x) = y$. The inverse function is usually denoted $f^{-1}$. Note that we may write $f^{-1}(y)$ to denote the inverse image of the singleton $\{y\}$ even if the function $f$ is not one-to-one.

You should verify that the inverse image preserves the set theoretic operations. That is,

$$f^{-1}\Big(\bigcap_{i \in I} A_i\Big) = \bigcap_{i \in I} f^{-1}(A_i), \qquad f^{-1}\Big(\bigcup_{i \in I} A_i\Big) = \bigcup_{i \in I} f^{-1}(A_i),$$
$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B).$$

## 1.4 A bestiary of relations

There are many conditions placed on binary relations in various contexts, and we summarize a number of them here. Some we have already mentioned above. We gather them here largely to standardize our terminology. Not all authors use the same terminology that we do. Each of these definitions should be interpreted as if prefaced by the appropriate universal quantifiers "for every $x$, $y$, $z$," etc. The symbol $\neg$ indicates negation, and a compound expression such as $x\,R\,y$ and $y\,R\,z$ may be abbreviated $x\,R\,y\,R\,z$.

A binary relation $R$ on a set $X$ is:

- **reflexive** if $x\,R\,x$.
- **irreflexive** if $\neg(x\,R\,x)$.
- **symmetric** if $x\,R\,y$ implies $y\,R\,x$. Note that this does not imply reflexivity.
- **asymmetric** if $x\,R\,y$ implies $\neg(y\,R\,x)$. An asymmetric relation is irreflexive.
- **antisymmetric** if $x\,R\,y$ and $y\,R\,x$ imply $x = y$. An antisymmetric relation may or may not be reflexive.
- **transitive** if $x\,R\,y$ and $y\,R\,z$ imply $x\,R\,z$.
- **complete**, or **connected**, if either $x\,R\,y$ or $y\,R\,x$ or both. Note that a complete relation is reflexive.
- **total**, or **weakly connected**, if $x \neq y$ implies either $x\,R\,y$ or $y\,R\,x$ or both. Note that a total relation may or may not be reflexive. Some authors call a total relation complete.
- a **partial order** if it is reflexive, transitive, and antisymmetric. Some authors (notably J. L. Kelley [198]) do not require a partial order to be reflexive.
- a **linear order** if it is total, transitive, and antisymmetric; a total partial order, if you will. It obeys the following **trichotomy law**: For every pair $x, y$ exactly one of $x\,R\,y$, $y\,R\,x$, or $x = y$ holds.
- an **equivalence relation** if it is reflexive, symmetric, and transitive.
- a **preorder**, or **quasiorder**, if it is reflexive and transitive. An antisymmetric preorder is a partial order.
- the **symmetric part** of the relation $S$ if $x\,R\,y \iff (x\,S\,y \;\&\; y\,S\,x)$.
- the **asymmetric part** of the relation $S$ if $x\,R\,y \iff (x\,S\,y \;\&\; \neg y\,S\,x)$.
- the **transitive closure** of the relation $S$ when $x\,R\,y$ whenever either $x\,S\,y$ or there is a finite set $\{x_1, \ldots, x_n\}$ such that $x S\,x_1\,S\,x_2 \cdots x_n\,S\,y$. The transitive closure of $S$ is the intersection of all the transitive relations (as sets of ordered pairs) that include $S$. (Note that the relation $X \times X$ is transitive and includes $S$, so we are not taking the intersection of the empty set.)

## 1.5 Equivalence relations

Equivalence relations are among the most important. As defined above, an **equivalence relation** on a set $X$ is a reflexive, symmetric, and transitive relation, often denoted $\sim$. Here are several familiar equivalence relations.

- Equality is an equivalence relation.
- For functions on a measure space, almost everywhere equality is an equivalence relation.
- In a semimetric space $(X, d)$, the relation defined by $x \sim y$ if $d(x, y) = 0$ is an equivalence relation.
- Given any function $f$ with domain $X$, we can define an equivalence relation $\sim$ on $X$ by $x \sim y$ whenever $f(x) = f(y)$.

Given an equivalence relation $\sim$ on a set $X$ we define the **equivalence class** $[x]$ of $x$ by $[x] = \{y : y \sim x\}$. If $x \sim y$, then $[x] = [y]$; and if $x \nsim y$, then $[x] \cap [y] = \varnothing$. The $\sim$-equivalence classes thus partition $X$ into disjoint sets. The collection of $\sim$-equivalence classes of $X$ is called the **quotient of $X$ modulo $\sim$**, often written as $X/\sim$. The function $x \mapsto [x]$ is called the **quotient mapping**. In many contexts, we **identify** the members of an equivalence class. What we mean by this is that we write $X$ instead of $X/\sim$, and we write $x$ instead of $[x]$. Hopefully, you (and we) will not become confused and make any mistakes when we do this. As an example, if we identify elements of a semimetric space as described above, the quotient space becomes a true metric space in the obvious way. In fact, all the $L_p$-spaces are quotient spaces defined in this manner.

A **partition** $\{D_i\}_{i \in I}$ of a set $X$ is a collection of nonempty subsets of $X$ satisfying $D_i \cap D_j = \varnothing$ for $i \neq j$ and $\bigcup_{i \in I} D_i = X$. Every partition defines an equivalence relation on $X$ by letting $x \sim y$ if $x, y \in D_i$ for some $i$. In this case, the equivalence classes are precisely the sets $D_i$.

## 1.6 Orders and such

A **partial order** (or **partial ordering**, or simply **order**) is a reflexive, transitive, and antisymmetric binary relation. It is traditional to use a symbol like $\geqslant$ to denote a partial order. The expressions $x \geqslant y$ and $y \leqslant x$ are synonyms. A set $X$ equipped with a partial order is a **partially ordered set**, sometimes called a **poset**. Two elements $x$ and $y$ in a partially ordered set are **comparable** if either $x \geqslant y$ or $y \geqslant x$ (or both, in which case $x = y$). A **total order** or **linear order** $\geqslant$ is a partial order where every two elements are comparable. That is, a total order is a partial order that is total. A **chain** in a partially ordered set is a subset that is totally ordered—any two elements of a chain are comparable. In a partially ordered set the notation $x > y$ means $x \geqslant y$ and $x \neq y$. The **order interval** $[x, y]$ is the set $\{z \in X : x \leqslant z \leqslant y\}$. Note that if $y \ngeq x$, then $[x, y] = \varnothing$.

Let $(X, \geqslant)$ be a partially ordered set. An **upper bound** for a set $A \subset X$ is an element $x \in X$ satisfying $x \geqslant y$ for all $y \in A$. An element $x$ is a **maximal element** of $X$ if there is no $y$ in $X$ for which $y > x$. Similarly, a **lower bound** for $A$ is an $x \in X$ satisfying $y \geqslant x$ for all $y \in A$. **Minimal elements** are defined analogously. A **greatest element** of $A$ is an $x \in A$ satisfying $x \geqslant y$ for all $y \in A$. **Least elements** are defined in the obvious fashion. Clearly a nonempty subset of $X$ has at most one greatest element and a greatest element if it exists is maximal. If the partial order is complete, then a maximal element is also the greatest. The **supremum** of a set is its least upper bound and the **infimum** is its greatest lower bound. The supremum and infimum of a set need not exist. We write $x \vee y$ for the supremum, and $x \wedge y$ for the infimum, of the two point set $\{x, y\}$. For linear orders, $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$. A **lattice** is a partially ordered set in which every pair of elements has a supremum and an infimum. It is easy to show (by induction) that every finite set in a lattice has a supremum and an infimum. A **sublattice** of a lattice is a subset that is closed under pairwise infima and suprema. A **complete lattice** is a lattice in which every nonempty subset $A$ has a supremum $\bigvee A$ and an infimum $\bigwedge A$. In particular, a complete lattice itself has an infimum, denoted 0, and a supremum denoted 1. The monograph by D. M. Topkis [331] provides a survey of some of the uses of lattices in economics.

A function $f\colon X \to Y$ between two partially ordered sets is **monotone** if $x \geqslant y$ in $X$ implies $f(x) \geqslant f(y)$ in $Y$. Some authors use the term **isotone** instead. The function $f$ is **strictly monotone** if $x > y$ in $X$ implies $f(x) > f(y)$ in $Y$. Monotone functions are also called **increasing** or **nondecreasing function.**[6] We may also say that $f$ is **decreasing** or **nonincreasing** if $x \geqslant y$ in $X$ implies $f(y) \geqslant f(x)$ in $Y$. Strictly decreasing functions are defined in the obvious way.

## 1.7 Real functions

A function whose range space is the real numbers is called a **real function** or a **real-valued function**. A function whose range space is the extended real numbers is called an **extended real function**. If an extended real function satisfies $f(x) = 0$ for all $x$ in a set $A$, we say that $f$ **vanishes** on $A$. Or if $x \notin B$ implies $f(x) = 0$, we say that $f$ vanishes outside $B$. For traditional reasons we also use the term **functional** to indicate a real linear or sublinear function on a vector space. (These terms are defined in Chapter 5.)

The **epigraph** of an (extended) real function $f$ on a set $X$, denoted $\operatorname{epi} f$, is the set in $X \times \mathbb{R}$ defined by $\operatorname{epi} f = \{(x, \alpha) \in X \times \mathbb{R} : \alpha \geqslant f(x)\}$. That is, $\operatorname{epi} f$ is the set of points lying on or above the graph of $f$. Notice that if $f(x) = \infty$, then the

[6] We use this terminology despite the fact, as D. M. Topkis [331] points out, the negation of "$f$ is increasing" is not "$f$ is nonincreasing." Do you see why?

In particular, the set of rational numbers is countable. (Why?) The following fact is an immediate consequence of those above.

- The set of all finite subsets of a countable set is again countable.

We use the countability of the rationals to jump ahead and prove the following well-known and important result.

**1.3 Theorem (Discontinuities of increasing functions)** *Let $I$ be an interval in $\mathbb{R}$ and let $f\colon I \to \mathbb{R}$ be nondecreasing, that is, $x > y$ implies $f(x) \geqslant f(y)$. Then $f$ has at most countably many points of discontinuity.*

*Proof*: For each $x$, since $f$ is nondecreasing,

$$\sup\{f(y) : y < x\} = f(x_-) \leqslant f(x) \leqslant f(x_+) = \inf\{f(y) : y > x\}.$$

Clearly $f$ is continuous at $x$ if and only if $f(x_-) = f(x) = f(x_+)$. So if $x$ is a point of discontinuity, then there is a rational number $q_x$ satisfying $f(x_-) < q_x < f(x_+)$. Furthermore if $x$ and $y$ are points of discontinuity and $x < y$, then $q_x < q_y$. (Why?) Thus $f$ has at most countably many points of discontinuity. ∎

Not every infinite set is countable; some are larger. G. Cantor showed that the set of real numbers is not countable using a technique now referred to as the **Cantor diagonal process**. It works like this. Suppose the unit interval $[0, 1]$ were countable. Then we could list the decimal expansion of the reals in $[0, 1]$ in order. We now construct a real number that does not appear on the list by romping down the diagonal and making sure our number is different from each number on the list. One way to do this is to choose a real number $b$ whose decimal expansion $0.b_1b_2b_3\ldots$ satisfies $b_n = 7$ unless $a_{n,n} = 7$ in which case we choose $b_n = 3$. In this way, $b$ differs from every number on the list. This shows that it is impossible to enumerate the unit interval with the integers. It also shows that $\mathbb{N}^{\mathbb{N}}$, the set of all sequences of natural numbers, is uncountable.

| $\mathbb{N}$ | $\mathbb{R}$ |
|---|---|
| 1 | $0.a_{11}a_{12}a_{13}\ldots$ |
| 2 | $0.a_{21}a_{22}a_{23}\ldots$ |
| 3 | $0.a_{31}a_{32}a_{33}\ldots$ |
| 4 | $0.a_{41}a_{42}a_{43}\ldots$ |
| $\vdots$ | $\vdots \quad \ddots$ |

A corollary of the uncountability of the reals is that there are well behaved linear orderings that have no real-valued representation.

**1.4 Example (An order with no utility)** Define the linear order $\succcurlyeq$ on $\mathbb{R}^2$ by $(x_1, x_2) \succcurlyeq (y_1, y_2)$ if and only if either $x_1 > y_1$ or $x_1 = y_1$ and $x_2 \geqslant y_2$. (This order is called the **lexicographic order** on the plane.) A **utility** for this order is a function $u\colon \mathbb{R}^2 \to \mathbb{R}$ satisfying $x \succcurlyeq y$ if and only if $u(x) \geqslant u(y)$. Now suppose by way of contradiction that this order has a utility. Then for each real number $x$, we have $u(x, 1) > u(x, 0)$. Consequently there must be some rational number $r_x$ satisfying $u(x, 1) > r_x > u(x, 0)$. Furthermore, if $x > y$, then $r_x > r_y$. Thus

$x \leftrightarrow r_x$ is a one-to-one correspondence between the real numbers and a set of rational numbers, implying that the reals are countable. This contradiction proves the claim. ∎

The cardinality of the set of real numbers $\mathbb{R}$ is called the cardinality of the **continuum**, written card $\mathbb{R} = \mathfrak{c}$. Here are some familiar sets with cardinality $\mathfrak{c}$.

- The intervals $[0, 1]$ and $(0, 1)$ (and as a matter of fact any nontrivial subinterval of $\mathbb{R}$).
- The Euclidean spaces $\mathbb{R}^n$.
- The set of irrational numbers in any nontrivial subinterval of $\mathbb{R}$.
- The collection of all subsets of a countably infinite set.
- The set $\mathbb{N}^{\mathbb{N}}$ of all sequences of natural numbers.

For more about the cardinality of sets see, for instance, T. Jech [185].

## 1.10 The Diagonal Theorem and Russell's Paradox

The diagonal process used by Cantor to show that the real numbers are not countable can be viewed as a special case of the following more general argument.

**1.5 Cantor's Diagonal Theorem** *Let $X$ be a set and let $\varphi\colon X \twoheadrightarrow X$ be a correspondence. Then the set $A = \{x \in X : x \notin \varphi(x)\}$ of non-fixed points of $\varphi$ is not a value of $\varphi$. That is, there is no $x$ satisfying $\varphi(x) = A$.*[9]

*Proof*: Assume by way of contradiction that there is some $x_0 \in X$ satisfying $\varphi(x_0) = A$. If $x_0$ is not a fixed point of $\varphi$, that is, $x_0 \notin \varphi(x_0)$, then by definition of $A$, we have $x_0 \in A = \varphi(x_0)$, a contradiction. On the other hand, if $x_0$ is a fixed point of $\varphi$, that is, $x_0 \in \varphi(x_0)$, then by definition of $A$, we have $x_0 \notin A = \varphi(x_0)$, also a contradiction. Hence $A$ is not the value of $\varphi$ at any point. ∎

**Russell's Paradox** is a clever argument devised by Bertrand Russell as an attack on the validity of the proof of the Diagonal Theorem. It goes like this. Let $S$ be the set of all sets, and let $\varphi\colon S \twoheadrightarrow S$ be defined by $\varphi(A) = \{B \in S : B \in A\}$ for every $A \in S$. Since $\varphi(A)$ is just the set of members of $A$, we have $\varphi(A) = A$. That is, $\varphi$ is the identity on $S$, so the set of its values is just $S$ again. By Cantor's Diagonal Theorem, the set $C = \{A \in S : A \notin \varphi(A)\}$ is not a value of $\varphi$, so it cannot be a set, which is a contradiction.

[9] Descriptive set theorists state the theorem as "$A$ is not in the range of $\varphi$," but they think of $\varphi$ as a function from $X$ to its power set $2^X$. For them the range is a subset of $2^X$, namely $\{\varphi(x) : x \in X\}$, but by our definition, the range is a subset of $X$, namely $\bigcup\{\varphi(x) : x \in X\}$.

The paradox was resolved not by repudiating the Diagonal Theorem, but by the realization that $S$, the collection of all sets, cannot itself be a set. What this means is that we have to be very much more careful about deciding what is a set and what is not a set.

## 1.11 The axiom of choice and axiomatic set theory

In Section 1.2, we were sloppy, even for us, but we were hoping you would not notice. For instance, we took it for granted that the union of a set of sets was a set, and that $I$-tuples (whatever they are) existed. Russell's Paradox tells us we should worry if these really are sets. Well maybe not we, but someone should worry. If you are worried, we recommend P. R. Halmos [149], or A. Shen and N. K. Vereshchagin [303] for "naïve set theory." For an excellent exposition of "axiomatic set theory," we recommend K. J. Devlin [92] or T. Jech [185].

Axiomatic set theory is viewed by many happy and successful people as a subject of no practical relevance. Indeed you may never have been exposed to the most popular axioms of set theory, the **Zermelo–Frankel (ZF) set theory**. For your edification we mention that ZF set theory proper has eight axioms. For instance, the Axiom of Infinity asserts the existence of an infinite set. There is also a ninth axiom, the **Axiom of Choice**, and ZF set theory together with this axiom is often referred to as ZFC set theory. We shall not list the others here, but suffice it to say that the first eight axioms are designed so that the collection of objects that we call sets is closed under certain set theoretic operations, such as unions and power sets. They were also designed to ward off Russell's Paradox.

The ninth axiom of ZFC set theory, the Axiom of Choice, is a seemingly innocuous set theoretic axiom with much hidden power.

**1.6 Axiom of Choice** *If $\{A_i : i \in I\}$ is a nonempty set of nonempty sets, then there is a function $f\colon I \to \bigcup_{i \in I} A_i$ satisfying $f(i) \in A_i$ for each $i \in I$. In other words, the Cartesian product of a nonempty set of nonempty sets is itself a nonempty set.*

The function $f$, whose existence the axiom asserts, chooses a member of $A_i$ for each $i$. Hence the term "Axiom of Choice." This axiom is both consistent with and independent of ZF set theory proper. That is, if the Axiom of Choice is dropped as an axiom of set theory, it cannot be proven by using the remaining eight axioms that the Cartesian product of nonempty sets is a nonempty set. Furthermore, adding the Axiom of Choice does not make the axioms of ZF set theory inconsistent. (A collection of axioms is inconsistent if it is possible to deduce both a statement $P$ and its negation $\neg P$ from the axioms.)

There has been some debate over the desirability of assuming the Axiom of Choice. (G. Moore [251] presents an excellent history of the Axiom of Choice and the controversy surrounding it.) Since there may be no way to describe the choice function, why should we assume it exists? Further, the Axiom of Choice has some unpleasant consequences. The Axiom of Choice makes it possible, for instance, to prove the existence of non-Lebesgue measurable sets of real numbers (Corollary 10.42). R. Solovay [316] has shown that by dropping the Axiom of Choice, it is possible to construct models of set theory in which all subsets of the real line are Lebesgue measurable. Since measurability is a major headache in integration and probability theory, it would seem that dropping the Axiom of Choice would be desirable. Along the same lines is the **Banach–Tarski Paradox** due to S. Banach and A. Tarski [32]. They prove, using the Axiom of Choice, that the unit ball $U$ in $\mathbb{R}^3$ can be partitioned into two disjoint sets $X$ and $Y$ with the property that $X$ can be partitioned into five disjoint sets, which can be reassembled (after translation and rotation) to make a copy of $U$, and the same is true of $Y$. That is, the ball can be cut up into pieces and reassembled to make two balls of the same size! (These pieces are obviously not Lebesgue measurable. Worse yet, this paradox shows that it is impossible to define a finitely additive volume in any reasonable manner on $\mathbb{R}^3$.) For a proof of this remarkable result, see, e.g., T. Jech [184, Theorem 1.2, pp. 3–6].

On the other hand, dropping the Axiom of Choice also has some unpleasant side effects. For example, without some version of the Axiom of Choice, our previous assertion that the countable union of countable sets is countable ceases to be true. Its validity can be restored by assuming the Countable Axiom of Choice, a weaker assumption that says only that a countable product of sets is a set. Without the Countable Axiom of Choice, there exist infinite sets that have no countably infinite subset. (See, for instance, T. Jech [184, Section 2.4, pp. 20–23].)

From our point of view, the biggest problem with dropping the Axiom of Choice is that some of the most useful tools of analysis would be thrown out with it. J. L. Kelley [197] has shown that the Tychonoff Product Theorem 2.61 would be lost. Most proofs of the Hahn–Banach Extension Theorem 5.53 make use of the Axiom of Choice, but it is not necessary. The Hahn–Banach theorem, which is central to linear analysis, can be proven using the Prime Ideal Theorem of Boolean Algebra, see W. A. J. Luxemburg [232]. The Prime Ideal Theorem is equivalent to the Ultrafilter Theorem 2.19, which we prove using Zorn's Lemma 1.7 (itself equivalent to the Axiom of Choice). J. D. Halpern [152] has shown that the Ultrafilter Theorem does not imply the Axiom of Choice. Nevertheless, M. Foreman and F. Wehrung [126] have shown that if the goal is to eliminate non-measurable sets, then we have to discard the Hahn–Banach Extension Theorem. That is, any superset of the ZF axioms strong enough to prove the Hahn–Banach theorem is strong enough to prove the existence of non-measurable sets. We can learn to live with non-measurable sets, but not without the Hahn–Banach theorem. So we might as well assume the Axiom of Choice. For more on the Axiom of Choice, we recommend the monograph by P. Howard and J. E. Rubin [170]. In addition, P. R. Halmos [149] and J. L. Kelley [198, Chapter 0] have extended discussions of the Axiom of Choice.

## 1.12 Zorn's Lemma

A number of propositions are equivalent to the Axiom of Choice. One of these is Zorn's Lemma, due to M. Zorn [350]. That is, Zorn's Lemma is a theorem if the Axiom of Choice is assumed, but if Zorn's Lemma is taken as an axiom, then the Axiom of Choice becomes a theorem.

**1.7 Zorn's Lemma** *If every chain in a partially ordered set $X$ has an upper bound, then $X$ has a maximal element.*

We indicate the power of Zorn's Lemma by employing it to prove a number of useful results from mathematics and economics. In addition to the results that we present in this section, we also use Zorn's Lemma to prove the Ultrafilter Theorem 2.19, the Tychonoff Product Theorem 2.61, the Hahn–Banach Extension Theorem 5.53, and the Krein–Milman Theorem 7.68.

The first use of Zorn's Lemma is the well-known fact that vector spaces possess Hamel bases. Recall that a **Hamel basis** or simply a **basis** of a vector space $V$ is a linearly independent set $B$ (every finite subset of $B$ is linearly independent) such that for each nonzero $x \in V$ there are $b_1, \dots, b_k \in B$ and nonzero scalars $\alpha_1, \dots, \alpha_k$ (all uniquely determined) such that $x = \sum_{i=1}^{k} \alpha_i b_i$.

**1.8 Theorem** *Every nontrivial vector space has a Hamel basis.*

*Proof*: Let $V$ be a nontrivial vector space, that is, $V \neq \{0\}$. Let $X$ denote the collection of all linearly independent subsets of $V$. Since $\{x\} \in X$ for each $x \neq 0$, we see that $X \neq \varnothing$. Note that $X$ is partially ordered by set inclusion. In addition, note that an element of $X$ is maximal if and only if it is a basis. (Why?) Now if $\mathcal{C}$ is a chain in $X$, then $A = \bigcup_{C \in \mathcal{C}} C$ is a linearly independent subset of $V$, so $A$ belongs to $X$ and is an upper bound for $\mathcal{C}$. By Zorn's Lemma 1.7, $X$ has a maximal element. Thus $V$ has a basis. ∎

As another example of the use of Zorn's Lemma, we present the following result, essentially due to E. Szpilrajn [327]. It is used to prove the key results in the theory of revealed preference, see M. K. Richter [283, Lemma 2, p. 640]. The proof of the result is not hard, but we present it in agonizing detail because the argument is so typical of how to use Zorn's Lemma.

It is always possible to extend any binary relation $R$ on a set $X$ to the total relation $S$ defined by $x \, S \, y$ for all $x, y$. But this is not very interesting since it destroys any asymmetry present in $R$. Let us say that the binary relation $S$ on a set $X$ is a **compatible extension** of the relation $R$ if $S$ extends $R$ and preserves the asymmetry of $R$. That is, $x \, R \, y$ implies $x \, S \, y$, and together $x \, R \, y$ and $\neg(y \, R \, x)$ imply $\neg(y \, S \, x)$.

**1.9 Theorem (Total extension of preorders)** *Any preorder has a compatible extension to a total preorder.*

Some care must be taken in the interpretation of this result. The theorem does *not* assert that the set $F$ of fixed points is a sublattice of $X$. It may well be that the supremum of a set in the lattice $(F, \geqslant)$ is not the same as its supremum in the lattice $(X, \geqslant)$. For example, let $X = \{0, 1, a, b, b'\}$ and define the partial order $\geqslant$ by $1 \geqslant a \geqslant b \geqslant 0$ and $1 \geqslant a \geqslant b' \geqslant 0$ (and all the other comparisons implied by transitivity and reflexivity). Note that $b$ and $b'$ are not comparable. Define the monotone function $f\colon X \to X$ by $f(x) = x$ for $x \neq a$ and $f(a) = 1$. The set of $F$ of fixed points of $f$ is $\{0, b, b', 1\}$, which is a complete lattice. Let $B = \{b, b'\}$ and note that $\bigvee B = 1$ when $B$ viewed as a subset of $F$, but $\bigvee B = a$, when $B$ viewed as a subset of $X$.

In a converse direction, any incomplete lattice has a fixed point-free monotone function into itself. For a proof, see A. C. Davies [81]. Tarksi's Theorem has been extended to cover increasing correspondences by R. E. Smithson [314] and X. Vives [336]. See F. Echenique [113] for more constructive proofs of these and related results.

## 1.13 Ordinals

We now apply Zorn's Lemma to the proof of the Well Ordering Principle, which is yet another equivalent of the Axiom of Choice.

**1.12 Definition** *A set $X$ is* ***well ordered*** *by the linear order $\leq$ if every nonempty subset of $X$ has a first element. An element $x$ of $A$ is* ***first*** *in $A$ if $x \leq y$ for all $y \in A$.*

*An* ***initial segment*** *of $(X, \leq)$ is any set of the form $I(x) = \{y \in X : y \leq x\}$.*

*An* ***ideal*** *in a well ordered set $X$ is a nonempty subset $A$ of $X$ such that for each $a \in A$ the initial segment $I(a)$ is included in $A$.*

**1.13 Well Ordering Principle** *Every nonempty set can be well ordered.*

*Proof*: Let $X$ be a nonempty set, and let

$$\mathcal{X} = \{(A, \leq_A) : A \subset X \text{ and } \leq_A \text{ well orders } A\}.$$

Note that $\mathcal{X}$ is nonempty, since every finite set is well ordered by any linear order. Define the partial order $\succcurlyeq$ on $\mathcal{X}$ by $(A, \leq_A) \succcurlyeq (B, \leq_B)$ if $B$ is an ideal in $A$ and $\leq_A$ extends $\leq_B$. If $\mathcal{C}$ is a chain in $\mathcal{X}$, set $C = \bigcup\{A : (A, \leq_A) \in \mathcal{C}\}$, and define $\leq_C$ on $C$ by $x \leq_C y$ if $x \leq_A y$ for some $(A, \leq_A) \in \mathcal{C}$. Then $\leq_C$ is a well defined order on $C$, and $(C, \leq_C)$ belongs to $\mathcal{X}$ (that is, $\leq_C$ well orders $C$) and is an upper bound for $\mathcal{C}$. (Why?) Therefore, by Zorn's Lemma 1.7, the partially ordered set $\mathcal{X}$ has a maximal element $(A, \leq)$. We claim that $A = X$, so that $X$ is well ordered by $\leq$. For if there is some $x \notin A$, extend $\leq$ to $A \cup \{x\}$ by $y \leq x$ for all $y \in A$. This extended relation well orders $A \cup \{x\}$ and $A$ is an ideal in $A \cup \{x\}$ (why?), contradicting the maximality of $(A, \leq)$. ∎

We now prove the existence of a remarkable and useful well ordered set.

**1.14 Theorem** *There is an ordered set $(\Omega, \leq)$ satisfying the following properties.*

1. *$\Omega$ is uncountable and well ordered by $\leq$.*
2. *$\Omega$ has a greatest element $\omega_1$.*
3. *If $x < \omega_1$, then the initial segment $I(x)$ is countable.*
4. *If $x < \omega_1$, then $\{y \in \Omega : x \leq y \leq \omega_1\}$ is uncountable.*
5. *Every nonempty subset of $\Omega$ has a least upper bound.*
6. *A nonempty subset of $\Omega \setminus \{\omega_1\}$ has a least upper bound in $\Omega \setminus \{\omega_1\}$ if and only if it is countable. In particular, the least upper bound of every uncountable subset of $\Omega$ is $\omega_1$.*

*Proof*: Let $(X, \leq)$ be an uncountable well ordered set, and consider the set $A$ of elements $x$ of $X$ such that the initial segment $I(x) = \{y \in X : y \leq x\}$ is uncountable. Without loss of generality we may assume $A$ is nonempty, for if $A$ is empty, append a point $y$ to $X$, and extend the ordering $\leq$ by $x \leq y$ for all $x \in X$. This order well orders $X \cup \{y\}$. Under the extension, $A$ is now nonempty. The set $A$ has a first element, traditionally denoted $\omega_1$. Set $\Omega = I(\omega_1)$, the initial segment generated by $\omega_1$. Clearly $\Omega$ is an uncountable well ordered set with greatest element $\omega_1$.

The proofs of the other properties except (6) are straightforward, and we leave them as exercises. So suppose $C = \{x_1, x_2, \dots\}$ is a countable subset of $\Omega \setminus \{\omega_1\}$. Then $\bigcup_{n=1}^{\infty} I(x_n)$ is countable, so there is some $x < \omega_1$ not belonging to this union. Such an $x$ is clearly an upper bound for $C$ so its least upper bound $b$ (which exists by (5)), satisfies $b \leq x < \omega_1$. For the converse, observe that if $b < \omega_1$ is a least upper bound for a set $C$, then $C$ is included in the countable set $I(b)$. ∎

The elements of $\Omega$ are called **ordinals**, and $\omega_1$ is called the **first uncountable ordinal**. The set $\Omega_0 = \Omega \setminus \{\omega_1\}$ is the set of **countable ordinals**. Also note that we can think of the natural numbers $\mathbb{N} = \{1, 2, \dots\}$ as a subset of $\Omega$: Identify 1 with the first element of $\Omega$, and recursively identify $n$ with the first element of $\Omega \setminus \{1, 2, \dots, n-1\}$. In interval notation we may write $\Omega = [1, \omega_1]$ and $\Omega_0 = [1, \omega_1)$.

The first element of $\Omega \setminus \mathbb{N}$ is denoted $\omega_0$. It is the **first infinite ordinal.**[10] Clearly, $n < \omega_0$ for each $n \in \mathbb{N}$. The names are justified by the fact that if we take any other well ordered uncountable set with a greatest element and find the first uncountable initial segment $\Omega' = [1', \omega']$, then there is a strictly monotone function $f$ from $\Omega$ onto $\Omega'$. To establish the existence of such a function $f$ argue as follows. Let

$$\mathcal{X} = \{(x, g) \mid x \in \Omega \text{ and } g\colon I(x) \to \Omega' \text{ is strictly monotone and has range } I(g(x))\}.$$

[10] Be aware that some authors use $\Omega$ to denote the first uncountable ordinal and $\omega$ to denote the first infinite ordinal.

If $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}' = \{1', 2', \ldots\}$ are the natural numbers of $\Omega$ and $\Omega'$ respectively, and $g\colon \mathbb{N} \to \Omega'$ is defined by $g(n) = n'$, then $(n, g) \in \mathcal{X}$ for each $n \in \mathbb{N}$. This shows that $\mathcal{X}$ is nonempty. Next, define a partial order $\succcurlyeq$ on $\mathcal{X}$ by $(x, g) \succcurlyeq (y, h)$ if $x \geq y$ and $g = h$ on $I(y)$. Now let $\{(x_\alpha, g_\alpha)\}_{\alpha \in A}$ be a chain in $\mathcal{X}$. Put $x = \sup_{\alpha \in A} x_\alpha$ in $\Omega$ and define $g\colon I(x) \to \Omega'$ by $g(y) = g_\alpha(y)$ if $y < x_\alpha$ for some $\alpha$ and $g(x) = \sup_{\alpha \in A} g(x_\alpha)$. Notice that $g$ is well defined, strictly monotone, and satisfies $g(I(x)) = I(g(x))$ and $(x, g) \succcurlyeq (x_\alpha, g_\alpha)$ for each $\alpha \in A$. This shows that every chain in $\mathcal{X}$ has an upper bound. By Zorn's lemma, $\mathcal{X}$ has a maximal element, say $(x, f)$. We now leave it as an exercise to you to verify that $x = \omega_1$ and that $f(\omega_1) = \omega_1'$. You should also notice that $f$ is uniquely determined and, in fact, $f(x)$ is the first element of the set $\Omega' \setminus \{f(y) : y < x\}$.

In the next chapter we make use of the following result.

**1.15 Interlacing Lemma** *Suppose $\{x_n\}$ and $\{y_n\}$ are interlaced sequences in $\Omega_0$. That is, $x_n \leq y_n \leq x_{n+1}$ for all $n$. Then both sequences have the same least upper bound in $\Omega_0$.*

*Proof*: By Theorem 1.14 (6), each sequence has a least upper bound in $\Omega_0$. Call the least upper bounds $x$ and $y$ respectively. Since $y_n \geq x_n$ for all $n$, we have $y \geq x$. Since $x_{n+1} \geq y_n$ for all $n$, we have $x \geq y$. Thus $x = y$. ∎

As an aside, here is how the Well Ordering Principle implies the Axiom of Choice. Let $\{A_i : i \in I\}$ be a nonempty family of nonempty sets. Well order $\bigcup_{i \in I} A_i$ and let $f(i)$ be the first element of $A_i$. Then $f$ is a choice function.

## Chapter 2

# Topology

We begin with a chapter on what is now known as general topology. Topology is the abstract study of convergence and approximation. We presume that you are familiar with the notion of convergence of a sequence of real numbers, and you may even be familiar with convergence in more general normed or metric spaces. Recall that a sequence $\{x_n\}$ of real numbers converges to a real number $x$ if $\{|x_n - x|\}$ converges to zero. That is, for every $\varepsilon > 0$, there is some $n_0$ such that $|x_n - x| < \varepsilon$ for all $n \geqslant n_0$. In metric spaces, the general notion of the distance between two points (given by the *metric*) plays the role of the absolute difference between real numbers, and the theory of convergence and approximation in metric spaces is not all that different from the theory of convergence and approximation for real numbers. For instance, a sequence $\{x_n\}$ of points in a metric space converges to a point $x$ if the distance $d(x_n, x)$ between $x_n$ and $x$ converges to zero as a sequence of real numbers. That is, if for every $\varepsilon > 0$, there is an $n_0$ such that $d(x_n, x) < \varepsilon$ for all $n \geqslant n_0$. However, metric spaces are inadequate to describe approximation and convergence in more general settings. A very real example of this is given by the notion of pointwise convergence of real functions on the unit interval. It turns out there is no way to define a metric on the space of all real functions on the interval $[0, 1]$ so that a sequence $\{f_n\}$ of functions converges pointwise to a function $f$ if and only if the distance between $f_n$ and $f$ converges to zero. Nevertheless, the notion of pointwise convergence is extremely useful, so it is imperative that a general theory of convergence should include it.

There are many equivalent ways we could develop a general theory of convergence.[1] In some ways, the most natural place to start is with the notion of a *neighborhood* as a primitive concept. A neighborhood of a point $x$ is a collection of points that includes all those "sufficiently close" to $x$. (In metric spaces, "sufficiently close" means within some positive distance $\varepsilon$.) We could define the collection of all neighborhoods and impose axioms on the family of neighborhoods. Instead of this, we start with the concept of an open set. An *open* set is a set that is a neighborhood of all its points. It is easier to impose axioms on

[1] The early development of topology used many different approaches to capture the notion of approximation: closure operations, proximity spaces, $L$-spaces, uniform spaces, etc. Some of these notions were discarded, while others were retained because of their utility.

the family of open sets than it is to impose them directly on neighborhoods. The family of all open sets is called a *topology*, and a set with a topology is called a *topological space*.

Unfortunately for you, a theory of convergence for topological spaces that is adequate to deal with pointwise convergence has a few quirks. Most prominent is the inadequacy of using sequences to describe continuity of functions. A function is continuous if it carries points sufficiently close in the domain to points sufficiently close in the range. For metric spaces, continuity of $f$ is equivalent to the condition that the sequence $\{f(x_n)\}$ converges to $f(x)$ whenever the sequence $\{x_n\}$ converges to $x$. This no longer characterizes continuity in the more general framework of topological spaces. Instead, we are forced to introduce either *nets* or *filters*. A net is like a sequence, except that instead of being indexed by the natural numbers, the index set can be much larger. Two particularly important techniques for indexing nets include indexing the net by the family of neighborhoods of a point, and indexing the net by the class of all finite subsets of a set.

There are offsetting advantages to working with general topological spaces. For instance, we can define topologies to make our favorite functions continuous. These are called *weak* topologies. The topology of pointwise convergence is actually a weak topology, and weak topologies are fundamental to understanding the equilibria of economies with an infinite dimensional commodity space.

Another important topological notion is compactness. Compact sets can be approximated arbitrarily well by finite subsets. (In Euclidean spaces, the compact sets are the closed and bounded sets.) Two of the most important theorems in this chapter are the Weierstrass Theorem 2.35, which states that continuous functions achieve their maxima on compact sets, and the Tychonoff Product Theorem 2.61, which asserts that the product of compact sets is compact in the product topology (the topology of pointwise convergence). This latter result is the basis of the Alaoglu Theorem 5.105, which describes a general class of compact sets in infinite dimensional spaces.

Liberating the notions of neighborhood and convergence from their metric space setting often leads to deeper insights into the structure of approximation methods. The idea of weak convergence and the keystone Tychonoff Product Theorem are perhaps the most important contributions of general topology to analysis—although at least one of us has heard the complaint that "topology is killing analysis." We collect a few fundamental topological definitions and results here. In the interest of brevity, we have included only material that we use later on, and have neglected other important and potentially useful results. We present no discussion of algebraic or differential topology, and have omitted discussion of quotient topologies, projective and inductive limits, metrizability theorems, extension theorems, and a variety of other topics. For more detailed treatments of general topology, there are a number of excellent standard references, including Dugundji [106], Kelley [198], Kuratowski [218], Munkres [256], and Willard [342]. Willard's historical notes are especially thorough.

The intersection of a family of topologies on a set is again a topology. (Why?) If $\mathcal{A}$ is an arbitrary nonempty family of subsets of a set $X$, then there exists a smallest (with respect to set inclusion) topology that includes $\mathcal{A}$. It is the intersection of all topologies that include $\mathcal{A}$. (Note that the discrete topology always includes $\mathcal{A}$.) This topology is called the **topology generated by $\mathcal{A}$** and consists precisely of $\varnothing$, $X$ and all sets of the form $\bigcup_\alpha V_\alpha$, where each $V_\alpha$ is a finite intersection of sets from $\mathcal{A}$.

A **base** for a topology $\tau$ is a subfamily $\mathcal{B}$ of $\tau$ such that each $U \in \tau$ is a union of members of $\mathcal{B}$. Equivalently, $\mathcal{B}$ is a base for $\tau$ if for every $x \in X$ and every open set $U$ containing $x$, there is a basic open set $V \in \mathcal{B}$ satisfying $x \in V \subset U$. Conversely, if $\mathcal{B}$ is a family of sets that is closed under finite intersections and $\bigcup \mathcal{B} = X$, then the family $\tau$ of all unions of members of $\mathcal{B}$ is a topology for which $\mathcal{B}$ is a base. A subfamily $\mathcal{S}$ of a topology $\tau$ is a **subbase** for $\tau$ if the collection of all finite intersections of members of $\mathcal{S}$ is a base for $\tau$. Note that if $\varnothing$ and $X$ belong to a collection $\mathcal{S}$ of subsets, then $\mathcal{S}$ is a subbase for the topology it generates. A topological space is called **second countable** if it has a countable base. (Note that a topology has a countable base if and only if it has a countable subbase.)

If $Y$ is a subset of a topological space $(X, \tau)$, then an easy argument shows that the collection $\tau_Y$ of subsets of $Y$, defined by

$$\tau_Y = \{V \cap Y : V \in \tau\},$$

is a topology on $Y$. This topology is called the **relative topology** or the **topology induced by $\tau$** on $Y$. When $Y \subset X$ is equipped with its relative topology, we call $Y$ a **(topological) subspace** of $X$. A set in $\tau_Y$ is called **(relatively) open** in $Y$. For example, since $X \in \tau$ and $Y \cap X = Y$, then $Y$ is relatively open in itself. Note that the relatively closed subsets of $Y$ are of the form

$$Y \setminus (Y \cap V) = Y \setminus V = Y \cap (X \setminus V),$$

where $V \in \tau$. That is, the relatively closed subsets of $Y$ are the restrictions of the closed subsets of $X$ to $Y$. Also note that for a semimetric topology, the relative topology is derived from the same semimetric restricted to the subset at hand. Unless otherwise stated, a subset $Y$ of $X$ carries its relative topology.

Part of the definition of a topology requires that a finite intersection of open sets is also an open set. However, a countable intersection of open sets need not be an open set. For instance, $\{0\} = \bigcap_{n=1}^{\infty}(-\frac{1}{n}, \frac{1}{n})$ is a countable intersection of open sets in $\mathbb{R}$ that is not open. Similarly, although finite unions of closed sets are closed sets, an arbitrary countable union of closed sets need not be closed; for instance, $(0, 1] = \bigcup_{n=1}^{\infty}[\frac{1}{n}, 1]$ is a countable union of closed sets in $\mathbb{R}$ that is neither open nor closed. The sets that are countable intersections of open sets or countable unions of closed sets are important enough that they have been given two special, albeit curious, names.

**2.3 Definition** *A subset of a topological space is:*

- *a $\mathcal{G}_\delta$-**set**, or simply a $\mathcal{G}_\delta$, if it is a countable intersection of open sets.*
- *an $\mathcal{F}_\sigma$-**set**, or simply an $\mathcal{F}_\sigma$, if it is a countable union of closed sets.*[2]

The example $(0,1] = \bigcup_{n=1}^{\infty}[\frac{1}{n},1] = \bigcap_{n=1}^{\infty}(0,1+\frac{1}{n})$ shows that a set can be simultaneously a $\mathcal{G}_\delta$- and an $\mathcal{F}_\sigma$-set.

## 2.2 Neighborhoods and closures

Let $(X,\tau)$ be a topological space, and let $A$ be any subset of $X$. The topology $\tau$ defines two sets intimately related to $A$. The **interior** of $A$, denoted $A^\circ$, is the largest (with respect to inclusion) open set included in $A$. (It is the union of all open subsets of $A$.) The interior of a nonempty set may be empty. The **closure** of $A$, denoted $\overline{A}$, is the smallest closed set including $A$; it is the intersection of all closed sets including $A$. It is not hard to verify that $A \subset B$ implies $A^\circ \subset B^\circ$ and $\overline{A} \subset \overline{B}$. Also, it is obvious that a set $A$ is open if and only if $A = A^\circ$, and a set $B$ is closed if and only if $B = \overline{B}$. Consequently, for any set $A$, $\overline{(\overline{A})} = \overline{A}$ and $(A^\circ)^\circ = A^\circ$.

**2.4 Lemma** *For any subset $A$ of a topological space, $A^\circ = (\overline{A^c})^c$.*

*Proof*: Clearly,

$$A^\circ \subset A \implies A^c \subset (A^\circ)^c \implies \overline{A^c} \subset \overline{(A^\circ)^c} = (A^\circ)^c \implies A^\circ \subset (\overline{A^c})^c.$$

Also, $A^c \subset \overline{A^c}$ implies $(\overline{A^c})^c \subset A$. Since $(\overline{A^c})^c$ is an open set and $A^\circ$ is the largest open set included in $A$, we see that $A^\circ = (\overline{A^c})^c$. ∎

The following property of the closure of the union of two sets easy to prove.

**2.5 Lemma** *If $A$ and $B$ are subsets of a topological space, then $\overline{A \cup B} = \overline{A} \cup \overline{B}$.*

A **neighborhood** of a point $x$ is any set $V$ containing $x$ in its interior. In this case we say that $x$ is an **interior point** of $V$. According to our definition, a neighborhood need not be an open set, but some authors define neighborhoods to be open.

**2.6 Lemma** *A set is open if and only if it is a neighborhood of each of its points.*

[2] This terminology seems to be derived from the common practice of using $G$ to denote open sets and $F$ for closed sets. The use of $F$ probably comes from the French *fermé*, and $G$ follows $F$. The letter $\sigma$ probably comes from the word sum, which was often the way unions are described. According to H. L. Royden [290, p. 53], the letter $\delta$ is for the German *durchschnitt*.

The collection of all neighborhoods of a point $x$, called the **neighborhood base**, or **neighborhood system**, at $x$, is denoted $\mathcal{N}_x$. It is easy to verify that $\mathcal{N}_x$ satisfies the following properties.

1. $X \in \mathcal{N}_x$.
2. For each $V \in \mathcal{N}_x$, we have $x \in V$ (so $\varnothing \notin \mathcal{N}_x$).
3. If $V, U \in \mathcal{N}_x$, then $V \cap U \in \mathcal{N}_x$.
4. If $V \in \mathcal{N}_x$ and $V \subset W$, then $W \in \mathcal{N}_x$.

**2.7 Definition** *A topology on X is called* ***Hausdorff*** *(or* ***separated****) if any two distinct points can be separated by disjoint neighborhoods of the points. That is, for each pair $x, y \in X$ with $x \neq y$ there exist neighborhoods $U \in \mathcal{N}_x$ and $V \in \mathcal{N}_y$ such that $U \cap V = \varnothing$.*

It is easy to see that singletons are closed sets in a Hausdorff space. (Why?) Topologies defined by metrics are Hausdorff. The trivial topology and the topologies in Examples 2.2.8 and 2.2.9 are not Hausdorff.

A **neighborhood base** at $x$ is a collection $\mathcal{B}$ of neighborhoods of $x$ with the property that if $U$ is any neighborhood of $x$, then there is a neighborhood $V \in \mathcal{B}$ with $V \subset U$. A topological space is called **first countable** if every point has a countable neighborhood base.[3] Every semimetric space is first countable: the balls of radius $\frac{1}{n}$ around $x$ form a countable neighborhood base at $x$. Clearly every second countable space is also first countable, but the converse is not true. (Consider an uncountable set with the discrete metric.)

A point $x$ is a **point of closure** or **closure point** of the set $A$ if every neighborhood of $x$ meets $A$. Note that $\overline{A}$ coincides with the set of all closure points of $A$. A point $x$ is an **accumulation point** (or a **limit point**, or a **cluster point**) of $A$ if for each neighborhood $V$ of $x$ we have $(V \setminus \{x\}) \cap A \neq \varnothing$.

To see the difference between closure points and limit points, consider the subset $A = [0, 1) \cup \{2\}$ of $\mathbb{R}$. Then 2 is a closure point of $A$ in $\mathbb{R}$, but not a limit point. The point 1 is both a closure point and a limit point of $A$.

We say that $x$ is a **boundary point** of $A$ if each neighborhood $V$ of $x$ satisfies both $V \cap A \neq \varnothing$ and $V \cap A^c \neq \varnothing$. Clearly, accumulation and boundary points of $A$ belong to its closure $\overline{A}$. Let $A'$ denote the set of all accumulation points of $A$ (called the **derived set** of $A$) and $\partial A$ denote the **boundary** of $A$, the set of all boundary points of $A$. We have the following identities:

$$\overline{A} = A^\circ \cup \partial A \quad \text{and} \quad \partial A = \partial A^c = \overline{A} \cap \overline{A^c}.$$

From the above identities, we see that a set $A$ is closed if and only if $A' \subset A$ (and also if and only if $\partial A \subset A$). In other words, we have the following result.

[3] Now you know why the term "second countable" exists.

**2.8 Lemma** *A set is closed if and only if it contains all its limit points.*

To illustrate this morass of definitions, again let $A = [0, 1) \cup \{2\}$ be viewed as a subset of $\mathbb{R}$. Then the boundary of $A$ is $\{0, 1, 2\}$ and its derived set is $[0, 1]$. The closure of $A$ is $[0, 1] \cup \{2\}$ and its interior is $(0, 1)$. Also note that the boundary of the set of rationals in $\mathbb{R}$ is the entire real line.

A subset $A$ of a topological space $X$ is **perfect** (in $X$) if it is closed and every point in $A$ is an accumulation point of $A$. In particular, every neighborhood of a point $x$ in $A$ contains a point of $A$ different from $x$. The space $X$ is perfect if all of its points are accumulation points. A point $x \in A$ is an **isolated point** of $A$ if there is a neighborhood $V$ of $x$ with $(V \setminus \{x\}) \cap A = \varnothing$. That is, if $\{x\}$ is a relatively open subset of $A$. A set is perfect if and only if it is closed and has no isolated points. Note that if $A$ has no isolated points, then its closure, $\overline{A}$, is perfect in $X$. (Why?) Also, note that the empty set is perfect.

## 2.3 Dense subsets

A subset $D$ of a topological space $X$ is **dense** (in $X$) if $\overline{D} = X$. In other words, a set $D$ is dense if and only if every nonempty open subset of $X$ contains a point in $D$. In particular, if $D$ is dense in $X$ and $x$ belongs to $X$, then every neighborhood of $x$ contains a point in $D$. This means that any point in $X$ can be approximated arbitrarily well by points in $D$. A set $N$ is **nowhere dense** if its closure has empty interior. A topological space is **separable** if it includes a countable dense subset.

**2.9 Lemma** *Every second countable space is separable.*

*Proof*: Let $\{B_1, B_2, \ldots\}$ be a countable base for the topology, and pick $x_i \in B_i$ for each $i$. Then $\{x_1, x_2, \ldots\}$ is dense. (Why?) ∎

The converse is true for metric spaces (Lemma 3.4), but not in general.

**2.10 Example (A separable space with no countable base)** We give two examples of separable spaces that do not have countable bases. The first example is highly artificial, but easy to understand. The second example is both natural and important, but it requires some material that we do not cover till later.

1. Let $X$ be an uncountable set and fix $x_0 \in X$. Take the topology consisting of the empty set and all sets containing $x_0$, cf. Example 2.2 (8). The set $\{x_0\}$ is dense in $X$, so $X$ is separable. Furthermore, each set of the form $\{x_0, x\}$, $x \in X$, is open, so there is no countable base.

2. In order to understand this example you need some knowledge of weak topologies (Section 2.13) and the representation of linear functionals on

sequence spaces (see Chapter 16). The example is the space $\ell_1$ of all absolutely summable real sequences equipped with the weak topology $\sigma(\ell_1, \ell_\infty)$. The countable set of all eventually zero sequences with rational components is a dense subset of $\ell_1$ (why?), so $(\ell_1, \sigma(\ell_1, \ell_\infty))$ is a separable Hausdorff space. However, $\sigma(\ell_1, \ell_\infty)$ is not first countable; see Theorem 6.26. ∎

## 2.4 Nets

A **sequence** in $X$ is a function from the natural numbers $\mathbb{N} = \{1, 2, \ldots\}$ into $X$. We usually think of a sequence as a subset of $X$ indexed by $\mathbb{N}$. A net is a direct generalization of the notion of a sequence. Instead of the natural numbers, the index set can be more general. The key issue is that the index set have a sense of direction. A **direction** $\geq$ on a (not necessarily infinite) set $D$ is a reflexive transitive binary relation with the property that each pair has an upper bound. That is, for each pair $\alpha, \beta \in D$ there exists some $\gamma \in D$ satisfying $\gamma \geq \alpha$ and $\gamma \geq \beta$. Note that a direction need not be a partial order since we do not require it to be antisymmetric. In practice, though, most directions are partial orders. Also note that for a direction, every finite set has an upper bound. A **directed set** is any set $D$ equipped with a direction $\geq$. Here are a few examples.

1. The set of all natural numbers $\mathbb{N} = \{1, 2, \ldots\}$ with the direction $\geq$ defined by $m \geq n$ whenever $m \geqslant n$.

2. The set $(0, \infty)$ under the direction $\geq$ defined by $x \geq y$ whenever $x \geqslant y$.

3. The set $(0, 1)$ under the direction $\geq$ defined by $x \geq y$ whenever $x \leqslant y$.

4. The neighborhood system $\mathcal{N}_x$ of a point $x$ in a topological space under the direction $\geq$ defined by $V \geq W$ whenever $V \subset W$. (The fact that the neighborhood system of a point is a directed set is the reason nets are so useful.)

5. The collection $\Phi$ of all finite subsets of a set $X$ under the direction $\geq$ defined by $A \geq B$ whenever $A \supset B$.

If $D$ is a directed set, then it is customary to denote the direction of $D$ by $\geqslant$ instead of $\geq$. The context in which the symbol $\geqslant$ is employed indicates whether or not it represents the direction of a set. If $A$ and $B$ are directed sets, then their Cartesian product $A \times B$ is also a directed set under the **product direction** defined by $(a, b) \geqslant (c, d)$ whenever $a \geqslant c$ and $b \geqslant d$. As a matter of fact, if $\{D_i : i \in I\}$ is an arbitrary family of directed sets, then their Cartesian product $D = \prod_{i \in I} D_i$ is also a directed set under the product direction defined by $(a_i)_{i \in I} \geqslant (b_i)_{i \in I}$ whenever $a_i \geqslant b_i$ for each $i \in I$. Unless otherwise indicated, the Cartesian product of a family of directed sets is directed by the product direction.

$\varphi\colon \Lambda \to A$ be the mapping appearing in the definition of the subnet. Also, pick some $\lambda_0 \in \Lambda$ satisfying $y_\lambda \in V$ for each $\lambda \geqslant \lambda_0$. Next, choose some $\lambda_1 \in \Lambda$ such that $\varphi_\lambda \geqslant \alpha_0$ for each $\lambda \geqslant \lambda_1$. If $\lambda_2 \in \Lambda$ satisfies $\lambda_2 \geqslant \lambda_1$ and $\lambda_2 \geqslant \lambda_0$, then the index $\beta = \varphi_{\lambda_2}$ satisfies $\beta \geqslant \alpha_0$ and $x_\beta = x_{\varphi_{\lambda_2}} = y_{\lambda_2} \in V$, so that $x$ is a limit point of the net $\{x_\alpha\}$. ∎

**2.17 Lemma** *In a topological space, a net converges to a point if and only if every subnet converges to that same point.*

*Proof*: Let $\{x_\alpha\}$ be a net in the topological space $X$ converging to $x$. Clearly, for every subnet $\{y_\lambda\}$ of $\{x_\alpha\}$ we have $y_\lambda \to x$. For the converse, assume that every subnet of $\{x_\alpha\}$ converges to $x$, and assume by way of contradiction that $\{x_\alpha\}$ does not converge to $x$. Then, there exists a neighborhood $V$ of $x$ such that for any index $\alpha \in A$ there exists some $\varphi_\alpha \geqslant \alpha$ with $x_{\varphi_\alpha} \notin V$. Now if $y_\alpha = x_{\varphi_\alpha}$, then $\{y_\alpha\}_{\alpha\in A}$ is a subnet of $\{x_\alpha\}$ that fails to converge to $x$. This is a contradiction, so $x_\alpha \to x$, as desired. Note that limits do not need to be unique for this result. ∎

As with sequences, every bounded net $\{x_\alpha\}$ of real numbers has a largest and a smallest limit point. The largest limit point of $\{x_\alpha\}$ is called the **limit superior**, written $\limsup_\alpha x_\alpha$, and the smallest is called the **limit inferior**, written $\liminf_\alpha x_\alpha$. It is not difficult to show that

$$\liminf_{\alpha} x_\alpha = \sup_{\alpha} \inf_{\beta\geqslant\alpha} x_\beta \leqslant \limsup_{\alpha} x_\alpha = \inf_{\alpha} \sup_{\beta\geqslant\alpha} x_\beta.$$

Also, note that $x_\alpha \to x$ in $\mathbb{R}$ if and only if

$$x = \liminf_{\alpha} x_\alpha = \limsup_{\alpha} x_\alpha.$$

## 2.5 Filters

The canonical example of a filter (and the reason filters are important in topology) is the neighborhood system $\mathcal{N}_x$ of a point $x$ in a topological space. We introduce filters not to maximize the number of new concepts, but because they are genuinely useful in their own right, see for instance, Theorem 2.86.

**2.18 Definition** *A* ***filter*** *on a set $X$ is a family $\mathfrak{F}$ of subsets of $X$ satisfying:*

1. *$\varnothing \notin \mathfrak{F}$ and $X \in \mathfrak{F}$;*
2. *If $A, B \in \mathfrak{F}$, then $A \cap B \in \mathfrak{F}$; and*
3. *If $A \subset B$ and $A \in \mathfrak{F}$, then $B \in \mathfrak{F}$.*

*A* ***free filter*** *is a filter $\mathfrak{F}$ with empty intersection, that is, $\bigcap_{A\in\mathfrak{F}} A = \varnothing$. Filters that are not free are called* ***fixed***.

Here are two more examples of filters.

- Let $X$ be an arbitrary set, and let $S$ be a nonempty subset of $X$. Then the collection of sets

$$\mathcal{F} = \{A \subset X : S \subset A\}$$

is a filter. Note that this filter is fixed.

- Let $X$ be an infinite set and consider the collection $\mathcal{F}$ of cofinite sets. (A set is **cofinite** if it is the complement of a finite set.) That is,

$$\mathcal{F} = \{A \subset X : A^c \text{ is a finite set}\}.$$

Observe that $\mathcal{F}$ is a free filter.

A filter $\mathcal{G}$ is a **subfilter** of another filter $\mathcal{F}$ if $\mathcal{F} \subset \mathcal{G}$. In this case we also say that $\mathcal{G}$ is **finer** than $\mathcal{F}$. Note that despite the term *sub*filter, this partial order on filters is the opposite of inclusion. A filter $\mathcal{U}$ is an **ultrafilter** if $\mathcal{U}$ has no proper subfilter. That is, $\mathcal{U}$ is an ultrafilter if $\mathcal{U} \subset \mathcal{G}$ for a filter $\mathcal{G}$ implies $\mathcal{U} = \mathcal{G}$.

**2.19 Ultrafilter Theorem** *Every filter is included in at least one ultrafilter. Consequently, every infinite set has a free ultrafilter.*

*Proof*: Let $\mathcal{F}$ be a filter on a set $X$, and let $\mathcal{C}$ be the nonempty collection of all subfilters of $\mathcal{F}$. That is,

$$\mathcal{C} = \{\mathcal{G} : \mathcal{G} \text{ is a filter and } \mathcal{F} \subset \mathcal{G}\}.$$

The collection $\mathcal{C}$ is partially ordered by inclusion. Given a chain $\mathcal{B}$ in $\mathcal{C}$, the family $\{A : A \in \mathcal{G} \text{ for some } \mathcal{G} \in \mathcal{B}\}$ is a filter that is an upper bound for $\mathcal{B}$ in $\mathcal{C}$. Thus the hypotheses of Zorn's Lemma 1.7 are satisfied, so $\mathcal{C}$ has a maximal element. Note that every maximal element of $\mathcal{C}$ is an ultrafilter including $\mathcal{F}$.

For the last part, note that if $X$ is an infinite set, then

$$\mathcal{F} = \{A \subset X : A^c \text{ is finite}\}$$

is a free filter. Any ultrafilter that includes $\mathcal{F}$ is a free ultrafilter. ■

Several useful properties of ultrafilters are included in the next three lemmas.

**2.20 Lemma** *Every fixed ultrafilter on a set $X$ is of the form*

$$\mathcal{U}_x = \{A \subset X : x \in A\}$$

*for a unique $x \in X$.*

*Proof*: Let $\mathcal{U}$ be a fixed ultrafilter on $X$ and let $x \in \bigcap_{A \in \mathcal{U}} A$. Then the family $\mathcal{U}_x = \{A \subset X : x \in A\}$ is a filter on $X$ satisfying $\mathcal{U} \subset \mathcal{U}_x$. Hence $\mathcal{U} = \mathcal{U}_x$. ■

A nonempty collection $\mathcal{B}$ of subsets of a set $X$ is a **filter base** if

1. $\varnothing \notin \mathcal{B}$; and
2. if $A, B \in \mathcal{B}$, then there exists some $C \in \mathcal{B}$ with $C \subset A \cap B$. (That is, $\mathcal{B}$ is directed by $\subset$.)

Every filter is, of course, a filter base. On the other hand, if $\mathcal{B}$ is a filter base for a set $X$, then the collection of sets

$$\mathcal{F}_{\mathcal{B}} = \{A \subset X : B \subset A \text{ for some } B \in \mathcal{B}\}$$

is a filter, called the **filter generated by** $\mathcal{B}$. For instance, the open neighborhoods at a point $x$ of a topological space form a filter base $\mathcal{B}$ satisfying $\mathcal{F}_{\mathcal{B}} = \mathcal{N}_x$ (the filter of all neighborhoods at $x$).

**2.21 Lemma** *An ultrafilter $\mathcal{U}$ on a set X satisfies the following:*

1. *If $A_1 \cup \cdots \cup A_n \in \mathcal{U}$, then $A_i \in \mathcal{U}$ for some $i$.*
2. *If $A \cap B \neq \varnothing$ for all $B \in \mathcal{U}$, then $A \in \mathcal{U}$.*

*Proof*: (1) Let $\mathcal{U}$ be an ultrafilter on $X$ and let $A \cup B \in \mathcal{U}$. If $A \notin \mathcal{U}$, then the collection of sets $\mathcal{F} = \{C \subset X : A \cup C \in \mathcal{U}\}$ is a filter satisfying $B \in \mathcal{F}$ and $\mathcal{U} \subset \mathcal{F}$. Hence, $\mathcal{F} = \mathcal{U}$, so $B \in \mathcal{U}$. The general case follows by induction.

(2) Assume that $A \cap B \neq \varnothing$ for all $B \in \mathcal{U}$. If $\mathcal{B} = \{A \cap B : B \in \mathcal{U}\}$, then $\mathcal{B}$ is a filter base and the filter $\mathcal{F}$ it generates satisfies $\mathcal{U} \subset \mathcal{F}$ and $A \in \mathcal{F}$. Since $\mathcal{U}$ is an ultrafilter, we see that $\mathcal{F} = \mathcal{U}$, so $A \in \mathcal{U}$. ∎

**2.22 Lemma** *If $\mathcal{U}$ is a free ultrafilter on a set X, then $\mathcal{U}$ contains no finite subsets of X. In particular, only infinite sets admit free ultrafilters.*

*Proof*: We first note that a free filter $\mathcal{U}$ contains no singletons. For if $\{x\} \in \mathcal{U}$, then $\{x\} \cap A \neq \varnothing$ for each $A \in \mathcal{U}$, so $x \in A$ for each $A \in \mathcal{U}$. Hence $\bigcap_{A \in \mathcal{U}} A \neq \varnothing$, a contradiction.

Now for an ultrafilter $\mathcal{U}$, if the finite set $\{x_1, \ldots, x_n\} = \bigcup_{i=1}^{n} \{x_i\}$ belongs to $\mathcal{U}$, then by Lemma 2.21 (1) we have $\{x_i\} \in \mathcal{U}$ for some $i$, contrary to the preceding observation. Hence, no finite subset of $X$ can be a member of $\mathcal{U}$. ∎

We now come to the definition of convergence for filters. A filter $\mathcal{F}$ in a topological space **converges** to a point $x$, written $\mathcal{F} \to x$, if $\mathcal{F}$ includes the neighborhood filter $\mathcal{N}_x$ at $x$, that is, $\mathcal{N}_x \subset \mathcal{F}$. Similarly, a filter base $\mathcal{B}$ converges to some point $x$, denoted $\mathcal{B} \to x$, if the filter generated by $\mathcal{B}$ converges to $x$. Clearly, $\mathcal{N}_x \to x$ for each $x$.

An element $x$ in a topological space is a **limit point** of a filter $\mathcal{F}$ whenever $x \in \overline{A}$ for each $A \in \mathcal{F}$. The set of all limit points of $\mathcal{F}$ is denoted Lim $\mathcal{F}$. Clearly, $\operatorname{Lim} \mathcal{F} = \bigcap_{A \in \mathcal{F}} \overline{A}$. As with nets, the limit points of a filter are precisely the limits of its subfilters.

**2.23 Theorem** *In a topological space, a point is a limit point of a filter if and only if there exists a subfilter converging to it.*

*Proof*: Let $x$ be a limit point of a filter $\mathcal{F}$ in a topological space. That is, let $x \in \bigcap_{A \in \mathcal{F}} \overline{A}$. Then, the collection of sets

$$\mathcal{B} = \{V \cap A : V \in \mathcal{N}_x \text{ and } A \in \mathcal{F}\}$$

is a filter base. Moreover, if $\mathcal{G}$ is the filter it generates, then both $\mathcal{F} \subset \mathcal{G}$ and $\mathcal{N}_x \subset \mathcal{G}$. That is, $\mathcal{G}$ is a subfilter of $\mathcal{F}$ converging to $x$.

For the converse, assume that $\mathcal{G}$ is a subfilter of $\mathcal{F}$ (that is, $\mathcal{F} \subset \mathcal{G}$) satisfying $\mathcal{G} \to x$ (that is, $\mathcal{N}_x \subset \mathcal{G}$). Then each $V \in \mathcal{N}_x$ and each $A \in \mathcal{F}$ both belong to $\mathcal{G}$. Consequently, $V \cap A \neq \varnothing$. Therefore, $x \in \bigcap_{A \in \mathcal{F}} \overline{A}$. ∎

We state without proof the following characterization of convergence.

**2.24 Lemma** *In a topological space, a filter converges to a point if and only if every subfilter converges to that same point.*

## 2.6 Nets and Filters

There is an intimate connection between nets and filters. Let $\{x_\alpha\}_{\alpha \in D}$ be a net in a topological space $X$. For each $\alpha$ define the **section** or **tail** $F_\alpha = \{x_\beta : \beta \geqslant \alpha\}$ and consider the family of sets $\mathcal{B} = \{F_\alpha : \alpha \in D\}$. It is a routine matter to verify that $\mathcal{B}$ is a filter base. The filter $\mathcal{F}$ generated by $\mathcal{B}$ is called the **section filter** of $\{x_\alpha\}$ or the **filter generated by the net** $\{x_\alpha\}$.

The net $\{x_\alpha\}_{\alpha \in D}$ and its section filter $\mathcal{F}$ have the same limit points. That is, $\operatorname{Lim}\{x_\alpha\} = \operatorname{Lim} \mathcal{F}$. Indeed, if $x \in \operatorname{Lim}\{x_\alpha\}$, then $x$ is (by Theorem 2.16) the limit of some subnet $\{y_\lambda\}$ of $\{x_\alpha\}$. A simple argument shows that the filter $\mathcal{G}$ generated by $\{y_\lambda\}$ is a subfilter of $\mathcal{F}$ and $\mathcal{G} \to x$. Conversely, if $x \in \operatorname{Lim} \mathcal{F}$, then for each index $\alpha$ and each $V \in \mathcal{N}_x$ we have $V \cap F_\alpha \neq \varnothing$. Thus if we choose some $y_{\alpha,V} \in V \cap F_\alpha$, then $\{y_{\alpha,V}\}_{(\alpha,V) \in D \times \mathcal{N}_x}$ defines a subnet of $\{x_\alpha\}$ satisfying $y_{\alpha,V} \to x$, so $x \in \operatorname{Lim}\{x_\alpha\}$.

Next, consider an arbitrary filter $\mathcal{F}$ in a topological space $X$ and then define the set $D = \{(a, A) : A \in \mathcal{F} \text{ and } a \in A\}$. The set $D$ has a natural direction $\geqslant$ defined by $(a, A) \geqslant (b, B)$ whenever $A \subset B$, so the formula $x_{a,A} = a$ defines a net in $X$, called the **net generated by the filter** $\mathcal{F}$. Observe that the section $F_{a,A} = A$, so the filter generated by the net $\{x_{a,A}\}$ is precisely $\mathcal{F}$. In particular, we have $\operatorname{Lim}\{x_{a,A}\} = \operatorname{Lim} \mathcal{F}$.

This argument establishes the following important equivalence result for nets and filters.

**2.25 Theorem (Equivalence of nets and filters)** *In a topological space, a net and the filter it generates have the same limit points. Similarly, a filter and the net it generates have the same limit points.*

## 2.7 Continuous functions

One of the most important duties of topologies is defining the class of continuous functions.

**2.26 Definition** *A function $f\colon X \to Y$ between topological spaces is* ***continuous*** *if $f^{-1}(U)$ is open in $X$ for each open set $U$ in $Y$.*

*We say that $f$ is* ***continuous at the point*** *$x$ if $f^{-1}(V)$ is a neighborhood of $x$ whenever $V$ is an open neighborhood of $f(x)$.*

In a metric space, continuity at a point $x$ reduces to the familiar $\varepsilon$-$\delta$ definition: For each $\varepsilon > 0$, the $\varepsilon$-ball at $f(x)$ is a neighborhood of $f(x)$. The inverse image of the ball is a neighborhood of $x$, so for some $\delta > 0$, the $\delta$-ball at $x$ is in the inverse image. That is, if $y$ is within $\delta$ of $x$, then $f(y)$ is within $\varepsilon$ of $f(x)$. The next two theorems give several other characterizations of continuity.

**2.27 Theorem** *For a function $f\colon X \to Y$ between topological spaces the following statements are equivalent.*

1. *$f$ is continuous.*
2. *$f$ is continuous at every point.*
3. *If $C$ is a closed subset of $Y$, then $f^{-1}(C)$ is a closed subset of $X$.*
4. *If $B$ is an arbitrary subset of $Y$, then $f^{-1}(B^\circ) \subset \big[f^{-1}(B)\big]^\circ$.*
5. *If $A$ is an arbitrary subset of $X$, then $f(\overline{A}) \subset \overline{f(A)}$.*
6. *$f^{-1}(V)$ is open in $X$ for each $V$ in some subbase for the topology on $Y$.*

*Proof*: (1) $\implies$ (2) This is obvious.

(2) $\implies$ (3) Let $C$ be a closed subset of $Y$ and let $x \in [f^{-1}(C)]^c = f^{-1}(C^c)$. So $f(x) \in C^c$. Since $C^c$ is an open set, the continuity of $f$ at $x$ guarantees the existence of some neighborhood $V$ of $x$ such that $y \in V$ implies $f(y) \in C^c$. The latter implies $V \subset f^{-1}(C^c)$, so $f^{-1}(C^c)$ is a neighborhood of all of its points. Thus $f^{-1}(C^c)$ is open, which implies that $f^{-1}(C) = [f^{-1}(C^c)]^c$ is closed.

(3) $\implies$ (4) Let $B$ be a subset of $Y$. Since $B^\circ$ is open, the set $(B^\circ)^c$ is closed, so by hypothesis $\big[f^{-1}(B^\circ)\big]^c = f^{-1}((B^\circ)^c)$ is also closed. This means that $f^{-1}(B^\circ)$ is open, and since $f^{-1}(B^\circ) \subset f^{-1}(B)$ is true, we see that $f^{-1}(B^\circ) \subset \big[f^{-1}(B)\big]^\circ$.

(4) $\implies$ (5) Let $A$ be an arbitrary subset of $X$ and let $y \in f(\overline{A})$. Then, there exists some $x \in \overline{A}$ with $y = f(x)$. If $V$ is an open neighborhood of $y$, then $f^{-1}(V) = f^{-1}(V^\circ) \subset [f^{-1}(V)]^\circ$, so $f^{-1}(V) = \big[f^{-1}(V)\big]^\circ$, proving that $f^{-1}(V)$ is an open neighborhood of $x$. Since $x \in \overline{A}$, we see that $f^{-1}(V) \cap A \neq \varnothing$, so $V \cap f(A) \neq \varnothing$. Therefore $y \in \overline{f(A)}$.

A family of sets has the **finite intersection property** if every finite subfamily has a nonempty intersection. Every filter has the finite intersection property, and an ultrafilter is a maximal family with the finite intersection property. Compactness can also be characterized in terms of the finite intersection property.

**2.31 Theorem** *For a topological space X, the following are equivalent.*

1. *$X$ is compact.*
2. *Every family of closed subsets of $X$ with the finite intersection property has a nonempty intersection.*
3. *Every net in $X$ has a limit point (or, equivalently, every net has a convergent subnet).*
4. *Every filter in $X$ has a limit point, (or, equivalently, every filter has a convergent subfilter).*
5. *Every ultrafilter in $X$ is convergent.*

*Proof*: (1) $\iff$ (2) Assume that $X$ is compact, and let $\mathcal{E}$ be a family of closed subsets of $X$. If $\bigcap_{E\in\mathcal{E}} E = \varnothing$, then $X = \bigcup_{E\in\mathcal{E}} E^{c}$, therefore $\{E^{c} : E \in \mathcal{E}\}$ is an open cover of $X$. Thus there exist $E_1, \ldots, E_n \in \mathcal{E}$ satisfying $X = \bigcup_{i=1}^{n} E_i^{c}$. This implies $\bigcap_{i=1}^{n} E_i = \varnothing$, so $\mathcal{E}$ does not have the finite intersection property. Thus, if $\mathcal{E}$ possesses the finite intersection property, then $\bigcap_{E\in\mathcal{E}} E \neq \varnothing$.

For the converse, assume that (2) is true and that $\mathcal{V}$ is an open cover of $X$. Then $\bigcap_{V\in\mathcal{V}} V^{c} = \varnothing$, so the finite intersection property must be violated. That is, there exist $V_1, \ldots, V_n \in \mathcal{V}$ satisfying $\bigcap_{j=1}^{n} V_j^{c} = \varnothing$, or $X = \bigcup_{j=1}^{n} V_j$, which proves that $X$ is compact.

(3) $\iff$ (4) This equivalence is immediate from Theorem 2.25.

(4) $\iff$ (5) This equivalence follows from Theorems 2.23 and 2.19.

(4) $\iff$ (2) Assume first that $\mathcal{G}$ is a family of closed subsets of $X$ with the finite intersection property. Then $\mathcal{G}$ is a filter base, so by hypothesis the filter $\mathcal{F}$ it generates has a limit point. Now note that $\bigcap_{G\in\mathcal{G}} G = \bigcap_{A\in\mathcal{F}} \overline{A} = \operatorname{Lim} \mathcal{F} \neq \varnothing$.

For the converse, assume that (2) is true and that $\mathcal{F}$ is a filter on $X$. Then the family of closed sets $\mathcal{G} = \{\overline{A} : A \in \mathcal{F}\}$ satisfies the finite intersection property, so $\operatorname{Lim} \mathcal{F} = \bigcap_{A\in\mathcal{F}} \overline{A} \neq \varnothing$. ∎

A subset $A$ of a topological space is **sequentially compact** if every sequence in $A$ has a subsequence converging to an element of $A$. A topological space $X$ is **sequentially compact** if $X$ itself is a sequentially compact set.

In many ways compactness can be viewed as a topological generalization of finiteness. There is an informal principle that compact sets behave like points in many instances. We list a few elementary properties of compact sets.

- Finite sets are compact.
- Finite unions of compact sets are compact.
- Closed subsets of compact sets are compact.
- If $K \subset Y \subset X$, then $K$ is a compact subset of $X$ if and only if $K$ is a compact subset of $Y$ (in the relative topology).

We note the following result, which we use frequently without any special mention. It is an instance of how compact sets act like points.

**2.32 Lemma** *If $K$ is a compact subset of a Hausdorff space, and $x \notin K$, then there are disjoint open sets $U$ and $V$ with $K \subset U$ and $x \in V$. In particular, compact subsets of Hausdorff spaces are closed.*

*Proof*: Since $X$ is Hausdorff, for each $y$ in $K$, there are disjoint open neighborhoods $U_y$ of $y$ and $V_y$ of $x$. The $U_y$s cover $K$, so there is a finite subfamily $U_{y_1}, \ldots, U_{y_n}$ covering $K$. Now note that the disjoint open sets $U = \bigcup_{i=1}^n U_{y_i}$ and $V = \bigcap_{i=1}^n V_{y_i}$ have the desired properties. ■

Compact subsets of non-Hausdorff spaces need not be closed.

**2.33 Example (A compact set that is not closed)** Let $X$ be a set with at least two elements, endowed with the indiscrete topology. Any singleton is compact, but $X$ is the only nonempty closed set. ■

**2.34 Theorem** *Every continuous function between topological spaces carries compact sets to compact sets.*

*Proof*: Let $f\colon X \to Y$ be a continuous function between two topological spaces, and let $K$ be a compact subset of $X$. Also, let $\{V_i : i \in I\}$ be an open cover of $f(K)$. Then $\{f^{-1}(V_i) : i \in I\}$ is an open cover of $K$. By the compactness of $K$ there exist indexes $i_1, \ldots, i_n$ satisfying $K \subset \bigcup_{j=1}^n f^{-1}(V_{i_j})$. Hence,

$$f(K) \subset f\Big(\bigcup_{j=1}^n f^{-1}(V_{i_j})\Big) = \bigcup_{j=1}^n f\big(f^{-1}(V_{i_j})\big) \subset \bigcup_{j=1}^n V_{i_j},$$

which shows that $f(K)$ is a compact subset of $Y$. ■

Since a subset of the real line is compact if and only if it is closed and bounded, the preceding lemma yields the following fundamental result.

**2.35 Corollary (Weierstrass)** *A continuous real-valued function defined on a compact space achieves its maximum and minimum values.*

A function $f: X \to Y$ between topological spaces is **open** if it carries open sets to open sets ($f(U)$ is open whenever $U$ is), and **closed** if it carries closed sets to closed sets ($f(F)$ is closed whenever $F$ is). If $f$ has an inverse, then $f^{-1}$ is continuous if and only if $f$ is open (and also if and only if $f$ is closed).

The following is a simple but very useful result.

**2.36 Theorem** *A one-to-one continuous function from a compact space onto a Hausdorff space is a homeomorphism.*

*Proof*: Assume that $f: X \to Y$ satisfies the hypotheses. If $C$ is a closed subset of $X$, then $C$ is a compact set, so by Theorem 2.34 the set $f(C)$ is also compact. Since $Y$ is Hausdorff, it follows that $f(C)$ is also a closed subset of $Y$. That is, $f$ is a closed function. Now note that $(f^{-1})^{-1}(C) = f(C)$, and by Theorem 2.27, the function $f^{-1}: Y \to X$ is also continuous. ∎

We close with an example of a compact Hausdorff space whose unusual properties are exploited in Examples 12.9 and 14.13.

**2.37 Example (Space of ordinals)** The set $\Omega = [1, \omega_1]$ of ordinals is a Hausdorff topological space with its **order topology**. A subbase for this topology consists of all sets of the form $\{y \in \Omega : y < x\}$ or $\{y \in \Omega : y > x\}$ for some $x \in \Omega$. Recall that any increasing sequence in $\Omega$ has a least upper bound. The least upper bound is also the limit of the sequence in the order topology.

The topological space $\Omega$ is compact. To see this, let $\mathcal{V}$ be an open cover of $\Omega$. Since $\omega_1$ is contained in some open set, then for some ordinal $x_0 < \omega_1$ the interval $(x_0, \omega_1] = \{y \in \Omega : x_0 < y \leq \omega_1\}$ is included in some member of the cover. Let $x_1$ be the first such ordinal, and let $V_1 \in \mathcal{V}$ satisfy $(x_1, \omega_1] \subset V_1$. By the same reasoning, unless $x_1 = 1$ there is a first ordinal $x_2 < x_1$ with $(x_2, x_1]$ included in some $V_2 \in \mathcal{V}$. Proceeding inductively, as long as $x_{n-1} \neq 1$, we can find $x_n < x_{n-1}$, the first ordinal with $(x_n, x_{n-1}] \subset V_n \in \mathcal{V}$. We claim that $x_n = 1$ for some $n$, so this process stops. Otherwise the set $\{x_1 > x_2 > \cdots\}$ has no first element. Thus $V_1, \ldots, V_n$ cover $\Omega$ with the possible exception of the point 1, which belongs to some member of $\mathcal{V}$.

Note that $\Omega$ is not separable: Let $C$ be any countable subset of $\Omega$, and let $b$ be the least upper bound of $C \setminus \{\omega_1\}$. Then any $x$ with $b < x < \omega_1$ cannot lie in the closure of $C$, so $C$ is not dense. A consequence of this is that $\Omega$ is not metrizable, since by Lemma 3.26 below, a compact metrizable space must be separable. ∎

## 2.9 Nets vs. sequences

So far, we have seen several similarities between nets and sequences, and you may be tempted to think that for most practical purposes nets and sequences behave alike. This is a mistake. We warn you that there are subtle differences between nets and sequences that you need to be careful of. The most important of them is highlighted by the following theorem and example.

**2.38 Theorem** *In a topological space, if a sequence $\{x_n\}$ converges to a point $x$, then the set $\{x, x_1, x_2, \ldots\}$ of all terms of the sequence together with the limit point $x$ is compact.*

*Proof*: Let $\{U_i\}_{i \in I}$ be an open cover of $S = \{x, x_1, x_2, \ldots\}$. Pick some index $i_0$ with $x \in U_{i_0}$ and note that there exists some $m$ such that $x_n \in U_{i_0}$ for all $n > m$. Now for each $1 \leqslant k \leqslant m$ pick an index $i_k$ with $x_k \in U_{i_k}$ and note that $S \subset \bigcup_{k=0}^{m} U_{i_k}$, which shows that $S$ is compact. ∎

Nets need not exhibit this property.

**2.39 Example (A convergent net without compact tails)** Let $D$ be the set of rational numbers in the interval $(0, 1)$, directed by the usual ordering $\geqslant$ on the real numbers. It defines a net $\{x_\alpha\}_{\alpha \in D}$ in the compact metric space $[0, 1]$ by letting $x_\alpha = \alpha$. Clearly, $x_\alpha \to 1$ in $[0, 1]$. If $\alpha_0 \in D$, then note that

$$\{x_\alpha : \alpha \geqslant \alpha_0\} \cup \{1\} = \{r \in [\alpha_0, 1] : r \text{ is a rational number}\},$$

which fails to be compact (or even closed) for any $\alpha_0 \in D$.

It is also interesting to note that for any $\alpha_0 \in D$, every real number $z \in [\alpha_0, 1)$ is an accumulation point of the set $\{x_\alpha : \alpha \geqslant \alpha_0\}$. However, note that there is no subnet of $\{x_\alpha\}$ that converges to $z$. (Every subnet of $\{x_\alpha\}$ converges to 1.) ∎

Whenever possible, it is desirable to replace nets with sequences, and theorems to this effect are very useful. One case that allows us to replace nets with sequences is the case of a first countable topology (each point has a countable neighborhood base). This class of spaces includes all metric spaces.

**2.40 Theorem** *Let $X$ be a first countable topological space.*

1. *If $A$ is a subset of $X$, then $x$ belongs to the closure of $A$ if and only if there is a sequence in $A$ converging to $x$.*

2. *A function $f: X \to Y$, where $Y$ is another topological space, is continuous if and only if $x_n \to x$ in $X$ implies $f(x_n) \to f(x)$ in $Y$.*

*Proof*: (1) Let $x \in \overline{A}$. Let $\{V_1, V_2, \ldots\}$ be a countable base for the neighborhood system $\mathcal{N}_x$ at $x$. Since $x \in \overline{A}$, we have $\left(\bigcap_{k=1}^{n} V_k\right) \cap A \neq \varnothing$ for each $n$. Pick $x_n \in \left(\bigcap_{k=1}^{n} V_k\right) \cap A$ and note that $x_n \to x$.

(2) If $f: X \to Y$ is continuous, then $x_n \to x$ implies $f(x_n) \to f(x)$. For the converse, assume that $x_n \to x$ in $X$ implies $f(x_n) \to f(x)$ in $Y$ and let $A \subset X$. By Theorem 2.27 (5), it suffices to show that $f(\overline{A}) \subset \overline{f(A)}$. So let $x \in \overline{A}$. By part (1), there exists a sequence $\{x_n\} \subset A$ satisfying $x_n \to x$. By hypothesis, $f(x_n) \to f(x)$, so $f(x) \in \overline{f(A)}$. ∎

## 2.10 Semicontinuous functions

A function $f\colon X \to [-\infty, \infty]$ on a topological space $X$ is:

- **lower semicontinuous** if for each $c \in \mathbb{R}$ the set $\{x \in X : f(x) \leqslant c\}$ is closed (or equivalently, the set $\{x \in X : f(x) > c\}$ is open).

- **upper semicontinuous** if for each $c \in \mathbb{R}$ the set $\{x \in X : f(x) \geqslant c\}$ is closed (or equivalently, the set $\{x \in X : f(x) < c\}$ is open).

Clearly, a function $f$ is lower semicontinuous if and only $-f$ is upper semicontinuous, and vice versa. Also, a real function is continuous if and only if it is both upper and lower semicontinuous.

**2.41 Lemma** *The pointwise supremum of a family of lower semicontinuous functions is lower semicontinuous. Similarly, the pointwise infimum of a family of upper semicontinuous functions is upper semicontinuous.*

*Proof*: We prove the lower semicontinuous case only. To this end, let $\{f_\alpha\}$ be a family of lower semicontinuous functions defined on a topological space $X$, and let $f(x) = \sup_\alpha f_\alpha(x)$ for each $x \in X$. From the identity

$$\{x \in X : f(x) \leqslant c\} = \bigcap_\alpha \{x \in X : f_\alpha(x) \leqslant c\},$$

we see that $\{x \in X : f(x) \leqslant c\}$ is closed for each $c \in \mathbb{R}$. ∎

The next characterization of semicontinuity is sometimes used as a definition. Later, in Corollary 2.60, we present another characterization of semicontinuity.

**2.42 Lemma** *Let $f\colon X \to [-\infty, \infty]$ be a function on a topological space. Then:*

$$f \text{ is lower semicontinuous if and only if } x_\alpha \to x \implies \liminf_\alpha f(x_\alpha) \geqslant f(x).$$

$$f \text{ is upper semicontinuous if and only if } x_\alpha \to x \implies \limsup_\alpha f(x_\alpha) \leqslant f(x).$$

*When $X$ is first countable, nets can be replaced by sequences.*

*Proof*: We establish the lower semicontinuous case. So assume first that $f$ is lower semicontinuous, and let $x_\alpha \to x$ in $X$. If $f(x) = -\infty$, then the desired inequality is trivially true. So suppose $f(x) > -\infty$. Fix $c < f(x)$ and note that (by the lower semicontinuity of $f$) the set $V = \{y \in X : f(y) > c\}$ is open. Since $x \in V$, there is some $\alpha_0$ such that $x_\beta \in V$ for all $\beta \geqslant \alpha_0$, that is, $f(x_\beta) > c$ for all $\beta \geqslant \alpha_0$. Hence,

$$\liminf_\alpha f(x_\alpha) = \sup_\alpha \inf_{\beta \geqslant \alpha} f(x_\beta) \geqslant \inf_{\beta \geqslant \alpha_0} f(x_\beta) \geqslant c$$

for all $c < f(x)$. This implies that $\liminf_\alpha f(x_\alpha) \geqslant f(x)$.

**2.48 Theorem** *Every compact Hausdorff space is normal, and therefore completely regular.*

*Proof*: Let $X$ be a compact Hausdorff space and let $E$ and $F$ be disjoint nonempty closed subsets of $X$. Then both $E$ and $F$ are compact. Choose a point $x \in E$. By Lemma 2.32 for each $y \in F$, there exist disjoint open sets $V_y$ and $U_y$ with $y \in V_y$ and $E \subset U_y$. Since $\{V_y : y \in F\}$ is an open cover of $F$, which is compact, there exist $y_1, \ldots, y_k \in F$ such that $F \subset \bigcup_{i=1}^k V_{y_i}$. Now note that the open sets $V = \bigcup_{i=1}^k V_{y_i}$ and $U = \bigcap_{i=1}^k U_{y_i}$ satisfy $E \subset U$, $F \subset V$, and $U \cap V = \varnothing$. ∎

We can modify the proof of Theorem 2.48 in order to prove a slightly stronger result. Before we can state the result we need the following definition. A topological space is a **Lindelöf space** if every open cover has a countable subcover. Clearly every second countable space is a Lindelöf space.

**2.49 Theorem** *Every regular Lindelöf space is normal.*

*Proof*: Let $A$ and $B$ be nonempty disjoint closed subsets of a Lindelöf space $X$. The regularity of $X$ implies that for each $x \in A$ there exists an open neighborhood $V_x$ of $x$ such that $\overline{V_x} \cap B = \varnothing$. Similarly, for each $y \in B$ there exists an open neighborhood $W_y$ of $y$ such that $\overline{W_y} \cap A = \varnothing$. Clearly the collection of open sets $\{V_x : x \in A\} \cup \{W_y : y \in B\} \cup \{X \setminus A \cup B\}$ covers $X$. Since $X$ is a Lindelöf space, there exist a countable subcollection $\{V_n\}$ of $\{V_x\}_{x \in A}$ and a countable subcollection $\{W_n\}$ of $\{W_y\}_{y \in B}$ such that $A \subset \bigcup_{n=1}^\infty V_n$ and $B \subset \bigcup_{n=1}^\infty W_n$.

Now for each $n$ let $V_n^* = V_n \setminus \bigcup_{i=1}^n \overline{W_i}$ and $W_n^* = W_n \setminus \bigcup_{i=1}^n \overline{V_i}$. Then the sets $V_n^*$ and $W_n^*$ are open, $V_n^* \cap W_m^* = \varnothing$ for all $n$ and $m$, $A \subset \bigcup_{n=1}^\infty V_n^* = V$, and $B \subset \bigcup_{n=1}^\infty W_n^* = W$. To finish the proof note that $V \cap W = \varnothing$. ∎

In addition to the properties already mentioned, there is another classification of topological spaces that you may run across, but which we eschew. A topological space is called a ***$T_0$*-space** if for each pair of distinct points, there is a neighborhood of one of them that does not contain the other. A ***$T_1$*-space** is one in which for each pair of distinct points, each has a neighborhood that does not contain the other. This is equivalent to each singleton being closed. A ***$T_2$*-space** is another name for a Hausdorff space. A ***$T_3$*-space** is a regular $T_1$-space. A ***$T_4$*-space** is a normal $T_1$-space. Finally, a ***$T_{3\frac{1}{2}}$*-space** or a **Tychonoff space** is a completely regular $T_1$-space.[6]

Here are some of the relations among the properties: Every Hausdorff space is $T_1$, and every $T_1$-space is $T_0$. A regular or normal space need not be Hausdorff: consider any two point set with the trivial topology. Every normal $T_1$-space is Hausdorff. A Tychonoff space is Hausdorff. For other separation axioms see A. Wilansky [340].

[6] If we had our way, the Hausdorff property would be part of the definition of a topology, and life would be much simpler.

## 2.12 Comparing topologies

The following two lemmas are trivial applications of the definitions, but they are included for easy reference. We feel free to refer to these results without comment. The proofs are left as an exercise.

**2.50 Lemma** *For two topologies $\tau'$ and $\tau$ on a set $X$ the following statements are equivalent.*

1. *$\tau'$ is weaker than $\tau$, that is, $\tau' \subset \tau$.*

2. *The identity mapping $x \mapsto x$, from $(X, \tau)$ to $(X, \tau')$, is continuous.*

3. *Every $\tau'$-closed set is also $\tau$-closed.*

4. *Every $\tau$-convergent net is also $\tau'$-convergent to the same point.*

5. *The $\tau$-closure of any subset is included in its $\tau'$-closure.*

**2.51 Lemma** *If $\tau'$ is weaker than $\tau$, then each of the following holds.*

1. *Every $\tau$-compact set is also $\tau'$-compact.*

2. *Every $\tau'$ continuous function on $X$ is also $\tau$ continuous.*

3. *Every $\tau$-dense set is also $\tau'$-dense.*

When we have a choice of what topology to put on a set, there is the following rough tradeoff. The finer the topology, the more open sets there are, so that more functions are continuous. On the other hand, there are also more insidious open covers of a set, so there tend to be fewer compact sets. There are a number of useful theorems involving continuous functions and compact sets. One is the Weierstrass Theorem 2.35, which asserts that a real continuous function on a compact set attains its maximum and minimum. The Brouwer–Schauder–Tychonoff Fixed Point Theorem 17.56 says that a continuous function from a compact convex subset of a locally convex linear space into itself has a fixed point. Another example is a Separating Hyperplane Theorem 5.79 that guarantees the existence of a continuous linear functional strongly separating a compact convex set from a disjoint closed convex set in a locally convex linear space.

## 2.13 Weak topologies

There are two classes of topologies that by and large include everything of interest. The first and most familiar is the class of topologies that are generated by a metric. The second class is the class of weak topologies.

Let $X$ be a nonempty set, let $\{(Y_i, \tau_i)\}_{i\in I}$ be a family of topological spaces and for each $i \in I$ let $f_i : X \to Y_i$ be a function. The **weak topology** or **initial topology** on $X$ generated by the family of functions $\{f_i\}_{i\in I}$ is the weakest topology on $X$ that makes all the functions $f_i$ continuous. It is the topology generated by the family of sets

$$\{f_i^{-1}(V) : i \in I \text{ and } V \in \tau_i\}.$$

Another subbase for this topology consists of

$$\{f_i^{-1}(V) : i \in I \text{ and } V \in \mathcal{S}_i\},$$

where $\mathcal{S}_i$ is a subbase for $\tau_i$. Let $w$ denote this weak topology. A base for the weak topology can be constructed out of the finite intersections of sets of this form. That is, the collection of sets of the form $\bigcap_{k=1}^n f_{i_k}^{-1}(V_{i_k})$, where each $V_{i_k}$ belongs to $\tau_{i_k}$ and $\{i_1, \ldots, i_n\}$ is an arbitrary finite subset of $I$, is a base for the weak topology. The next lemma is an important tool for working with weak topologies.

**2.52 Lemma** *A net satisfies $x_\alpha \xrightarrow{w} x$ for the weak topology $w$ if and only if $f_i(x_\alpha) \xrightarrow{\tau_i} f_i(x)$ for each $i \in I$.*

*Proof*: Since each $f_i$ is $w$-continuous, if $x_\alpha \xrightarrow{w} x$, then $f_i(x_\alpha) \xrightarrow{\tau_i} f_i(x)$ for all $i \in I$. Conversely, let $V = \bigcap_{k=1}^n f_{i_k}^{-1}(V_{i_k})$ be a basic neighborhood of $x$, where each $V_{i_k} \in \tau_{i_k}$. For each $k$, if $f_{i_k}(x_\alpha) \xrightarrow{\tau_{i_k}} f_{i_k}(x)$, then there is $\alpha_{i_k}$ such that $\alpha \geqslant \alpha_{i_k}$ implies $x_\alpha \in f_{i_k}^{-1}(V_{i_k})$. Pick $\alpha_0 \geqslant \alpha_{i_k}$ for all $k$. Then $\alpha \geqslant \alpha_0$ implies $x_\alpha \in V$. That is, $x_\alpha \xrightarrow{w} x$. ∎

An important special case is the weak topology generated by a family of real functions. For a family $\mathcal{F}$ of real functions on $X$, the weak topology generated by $\mathcal{F}$ is denoted $\sigma(X, \mathcal{F})$. It is easy to see that a subbase for $\sigma(X, \mathcal{F})$ can be found by taking all sets of the form

$$U(f, x, \varepsilon) = \{y \in X : |f(y) - f(x)| < \varepsilon\},$$

where $f \in \mathcal{F}$, $x \in X$, and $\varepsilon > 0$.

We say that a family $\mathcal{F}$ of real functions on $X$ is **total**, or **separates points** in $X$, if $f(x) = f(y)$ for all $f$ in $\mathcal{F}$ implies $x = y$. Another way to say the same thing is that $\mathcal{F}$ separates points in $X$ if for every $x \neq y$ there is a function $f$ in $\mathcal{F}$ satisfying $f(x) \neq f(y)$. The weak topology $\sigma(X, \mathcal{F})$ is Hausdorff if and only if $\mathcal{F}$ is total.

Here is a subtle point about weak topologies. Let $\mathcal{F}$ be a family of real-valued functions on a set $X$. Every subset $A \subset X$ has a relative topology induced by the $\sigma(X, \mathcal{F})$ weak topology on $X$. It also has its own weak topology, the $\sigma(A, \mathcal{F}|_A)$ topology, where $\mathcal{F}|_A$ is the family of restrictions of the functions in $\mathcal{F}$ to $A$. Are these topologies the same? Conveniently the answer is yes.

**2.53 Lemma (Relative weak topology)** *Let $\mathcal{F}$ be a family of real-valued functions on a set X, and let A be a subset of X. The $\sigma(A, \mathcal{F}|_A)$ weak topology on A is the relative topology on A induced by the $\sigma(X, \mathcal{F})$ weak topology on X.*

*Proof*: Use Lemma 2.52 to show that the convergent nets in each topology are the same. This implies that the identity is a homeomorphism. ∎

We employ the following standard notation throughout this monograph:

- $\mathbb{R}^X$ denotes the vector space of real-valued functions on a nonempty set $X$.

- $C(X)$ denotes the vector space of continuous real-valued functions on the topological space $(X, \tau)$. We may occasionally use the abbreviation $C$ for $C(X)$ when $X$ is clear from the context. We also use the common shorthand $C[0, 1]$ for $C([0, 1])$, the space of continuous real functions on the unit interval $[0, 1]$.

- $C_b(X)$ is the space of bounded continuous real functions on $(X, \tau)$. It is a vector subspace of $C(X)$.[7]

- The **support** of a real function $f \colon X \to \mathbb{R}$ on a topological space is the closure of the set $\{x \in X : f(x) \neq 0\}$, denoted $\operatorname{supp} f$. That is,

$$\operatorname{supp} f = \overline{\{x \in X : f(x) \neq 0\}}.$$

$C_c(X)$ denotes the vector space of all continuous real-valued functions on $X$ with compact support.

The vector space $\mathbb{R}^X$ coincides, of course, with the vector space $C(X)$ when $X$ is equipped with the discrete topology.

We now make a simple observation about weak topologies.

**2.54 Lemma** *The weak topology on the topological space X generated by $C(X)$ is the same as the weak topology generated by $C_b(X)$.*

*Proof*: Consider a subbasic open set $U(f, x, \varepsilon) = \{y \in X : |f(y) - f(x)| < \varepsilon\}$, where $f \in C(X)$. Define the function $g \colon X \to \mathbb{R}$ by

$$g(z) = \min\{f(x) + \varepsilon, \max\{f(x) - \varepsilon, f(z)\}\}.$$

Then $g \in C_b(X)$ and $U(g, x, \varepsilon) = U(f, x, \varepsilon)$. Thus $\sigma(X, C_b)$ is as strong as $\sigma(X, C)$. The converse is immediate. Therefore $\sigma(X, C_b) = \sigma(X, C)$. ∎

We can use weak topologies to characterize completely regular spaces.

**2.55 Theorem** *A topological space $(X, \tau)$ is completely regular if and only if $\tau = \sigma(X, C(X)) = \sigma(X, C_b(X))$.*

---

[7] The notation $C^*$ is used in some specialties for denoting $C_b$.

*Proof*: For any topological space $(X, \tau)$, we have $\sigma(X, C) \subset \tau$.

Assume first that $(X, \tau)$ is completely regular. Let $x$ belong to the $\tau$-open set $U$. Pick $f \in C(X)$ satisfying $f(x) = 0$ and $f(U^c) = \{1\}$. Then $\{y \in X : f(y) < 1\}$ is a $\sigma(X, C)$-open neighborhood of $x$ included in $U$. Thus $U$ is also $\sigma(X, C)$-open, so $\sigma(X, C) = \tau$.

Suppose now that $\tau = \sigma(X, C)$. Let $F$ be closed and $x \notin F$. Since $F^c$ is $\sigma(X, C)$-open, there is a neighborhood $U \subset F^c$ of $x$ of the form

$$U = \bigcap_{i=1}^{m} \{y \in X : |f_i(y) - f_i(x)| < 1\},$$

where each $f_i \in C(X)$. For each $1 \leqslant i \leqslant m$ let $g_i(z) = \min\{1, |f_i(z) - f_i(x)|\}$ and $g(z) = \max_i g_i(z)$. Then $g$ continuously maps $X$ into $[0, 1]$, and satisfies $g(x) = 0$ and $g(F) = \{1\}$. Thus $X$ is completely regular. ∎

**2.56 Corollary** *The completely regular spaces are precisely those whose topology is the weak topology generated by a family of real functions.*

*Proof*: If $(X, \tau)$ is completely regular, then $\tau = \sigma(X, C(X))$.

Conversely, suppose $\tau = \sigma(X, \mathcal{F})$ for a family $\mathcal{F}$ of real functions. Then $\mathcal{F} \subset C(X)$, so $\tau = \sigma(X, \mathcal{F}) \subset \sigma(X, C(X))$. But on the other hand, $\tau$ always includes $\sigma(X, C(X))$. Thus $\tau = \sigma(X, C(X))$, so by Theorem 2.55, $(X, \tau)$ is completely regular. ∎

The next easy corollary of Theorem 2.55 and Lemma 2.52 characterizes convergence in completely regular spaces.

**2.57 Corollary** *If $X$ is completely regular, then a net $x_\alpha \to x$ in $X$ if and only if $f(x_\alpha) \to f(x)$ for all $f \in C_b(X)$.*

For additional results on completely regular spaces see Chapter 3 of the excellent book by L. Gillman and M. Jerison [138].

## 2.14 The product topology

Let $\{(X_i, \tau_i)\}_{i \in I}$ be a family of topological spaces and let $X = \prod_{i \in I} X_i$ denote its Cartesian product. A typical element $x$ of the product may also be denoted $(x_i)_{i \in I}$ or simply $(x_i)$. For each $j \in I$, the **projection** $P_j \colon X \to X_j$ is defined by $P_j(x) = x_j$. The **product topology** $\tau$, denoted $\prod_{i \in I} \tau_i$, is the weak topology on $X$ generated by the family of projections $\{P_i : i \in I\}$. That is, $\tau$ is the weakest topology on $X$ that makes each projection $P_i$ continuous. A subbase for the product topology consists of all sets of the form $P_j^{-1}(V_j) = \prod_{i \in I} V_i$ where $V_i = X_i$ for all $i \neq j$ and $V_j$ is open in $X_j$. A base for the product topology consists of all sets of the form

$$V = \prod_{i \in I} V_i,$$

and note that $U_x$ is an open neighborhood of $x$. Similarly, for every $y \in K_2$ set $V_y = \bigcap_{y \in V_j} V_j$. Observe that for each $(x, y)$, the neighborhood $U_x \times V_y$ is included in one of the original $U_i \times V_i$. (Why?) From the compactness of $K_1$ and $K_2$, there exist elements $x_1, \dots, x_m \in K_1$ and $y_1, \dots, y_\ell \in K_2$ with $K_1 \subset \bigcup_{j=1}^m U_{x_j}$ and $K_2 \subset \bigcup_{r=1}^\ell V_{y_r}$. Next, note that the open sets $U = \bigcup_{j=1}^m U_{x_j}$ and $V = \bigcup_{r=1}^\ell V_{y_r}$ satisfy

$$K_1 \times K_2 \subset U \times V \subset \bigcup_{j=1}^n U_j \times V_j \subset G.$$

So the conclusion is true for a family of two topological spaces. By induction, the claim is true for any finite family of topological spaces. (Why?) For the general case, pick a finite collection $\{\prod_{i \in I} V_i^j\}_{j=1,\dots,k}$ of basic open sets such that $K = \prod_{i \in I} K_i \subset \bigcup_{j=1}^k (\prod_{i \in I} V_i^j) \subset G$. (This is possible since $K$ is compact by the Tychonoff Product Theorem 2.61.) This implies that the general case can be reduced to that of a finite family of topological spaces. We leave the remaining details as an exercise. ■

## 2.15 Pointwise and uniform convergence

For a nonempty set $X$, the product topology on $\mathbb{R}^X$ is also called the **topology of pointwise convergence** on $X$ because a net $\{f_\alpha\}$ in $\mathbb{R}^X$ satisfies $f_\alpha \to f$ in $\mathbb{R}^X$ if and only if $f_\alpha(x) \to f(x)$ in $\mathbb{R}$ for each $x \in X$.

Remarkably, if $\mathcal{F}$ is a set of real-valued functions on $X$, we can also regard $X$ as a set of real-valued functions on $\mathcal{F}$. Each $x \in X$ can be regarded as an **evaluation functional** $e_x \colon \mathcal{F} \to \mathbb{R}$, where $e_x(f) = f(x)$. As such, there is also a weak topology on $\mathcal{F}$, $\sigma(\mathcal{F}, X)$. This topology is identical to the relative topology on $\mathcal{F}$ as a subset of $\mathbb{R}^X$ endowed with the product topology. We also note the following important result.

**2.63 Lemma** *If $\mathcal{F}$ is a total family of real functions on a set $X$, the function $x \mapsto e_x$, mapping $(X, \sigma(X, \mathcal{F}))$ into $\mathbb{R}^{\mathcal{F}}$ with its product topology, is an embedding.*

*Proof*: Since $\mathcal{F}$ is a total, the mapping $x \mapsto e_x$ is one-to-one. The rest is just a restatement of Lemma 2.52, using the observation that the product topology on $\mathbb{R}^{\mathcal{F}}$ is the topology of pointwise convergence on $\mathcal{F}$. ■

From the Tychonoff Product Theorem 2.61, it follows that a subset $\mathcal{F}$ of $\mathbb{R}^X$ is compact in the product topology if and only if it is closed and pointwise bounded. Since a subset of $\mathcal{F}$ is compact in $\mathcal{F}$ if and only if it is compact in $\mathbb{R}^X$, we see that a subset of $\mathcal{F}$ is weakly compact (compact in the product topology) if and only if it is pointwise bounded and contains the pointwise limits of its nets.

We are now in a position to give a natural example of the inadequacy of sequences. They cannot describe the product topology on an uncountable product.

**2.64 Example** Let $[0,1]^{[0,1]}$ be endowed with its product topology, the topology of pointwise convergence. Let $F$ denote the family of indicator functions of finite subsets of $[0,1]$. Recall that the indicator function $\chi_A$ of a set $A$ is defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Then $\mathbf{1}$, the function that is identically one, is not the pointwise limit of any sequence in $F$: Let $\chi_{A_n}$ be a sequence in $F$. Then $A = \bigcup_{n=1}^{\infty} A_n$ is countable, so there is some point $x$ not belonging to $A$. Since $\chi_{A_n}(x) = 0$ for all $n$, the sequence does not converge pointwise to $\mathbf{1}$.

However there is a net in $F$ that converges pointwise to $\mathbf{1}$: Take the family $\mathcal{F}$ of all finite subsets of $[0,1]$ directed upward by inclusion—that is, $A \geqslant B$ if $A \supset B$. Then the net $\{\chi_A : A \in \mathcal{F}\}$ converges pointwise to $\mathbf{1}$. (Do you see why?) ∎

A net $\{f_\alpha\}$ in $\mathbb{R}^X$ converges **uniformly** to a function $f \in \mathbb{R}^X$ whenever for each $\varepsilon > 0$ there exists some index $\alpha_0$ (depending upon $\varepsilon$ alone) such that

$$\left|f_\alpha(x) - f(x)\right| < \varepsilon$$

for each $\alpha \geqslant \alpha_0$ and each $x \in X$. Clearly, uniform convergence implies pointwise convergence, but the converse is not true.

**2.65 Theorem** *The uniform limit of a net of continuous real functions is continuous.*

*Proof*: Let $\{f_\alpha\}$ be a net of continuous real functions on a topological space $X$ that converges uniformly to a function $f \in \mathbb{R}^X$. Suppose $x_\lambda \to x$ in $X$. We now show that $f(x_\lambda) \to f(x)$.

Let $\varepsilon > 0$ be given, and pick some $\alpha_0$ satisfying $|f_\alpha(y) - f(y)| < \varepsilon$ for all $\alpha \geqslant \alpha_0$ and all $y \in X$. Since $f_{\alpha_0}$ is a continuous function, there exists some $\lambda_0$ such that $|f_{\alpha_0}(x_\lambda) - f_{\alpha_0}(x)| < \varepsilon$ for all $\lambda \geqslant \lambda_0$. Hence, for $\lambda \geqslant \lambda_0$ we have

$$\begin{aligned} &\left|f(x_\lambda) - f(x)\right| \\ &\leqslant \left|f(x_\lambda) - f_{\alpha_0}(x_\lambda)\right| + \left|f_{\alpha_0}(x_\lambda) - f_{\alpha_0}(x)\right| + \left|f_{\alpha_0}(x) - f(x)\right| \\ &< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon. \end{aligned}$$

Thus, $f(x_\lambda) \to f(x)$, so $f$ is a continuous function. ∎

Here is a simple sufficient condition for a net to converge uniformly.

**2.66 Dini's Theorem** *If a net of continuous real functions on a compact space converges monotonically to a continuous function pointwise, then the net converges uniformly.*

*Proof*: Let $\{f_\alpha\}$ be a net of continuous functions on the compact space $X$ satisfying $f_\alpha(x) \downarrow f(x)$ for each $x \in X$, where $f$ is continuous. Replacing $f_\alpha$ by $f_\alpha - f$ we may assume that $f$ is identically zero.

Let $\varepsilon > 0$. For each $x \in X$ pick an index $\alpha_x$ such that $0 \leqslant f_{\alpha_x}(x) < \varepsilon$. By the continuity of $f_{\alpha_x}$ there is an open neighborhood $V_x$ of $x$ such that $0 \leqslant f_{\alpha_x}(y) < \varepsilon$ for all $y \in V_x$. Since $\alpha \geqslant \alpha_x$ implies $f_\alpha \leqslant f_{\alpha_x}$, we see that $0 \leqslant f_\alpha(y) < \varepsilon$ for each $\alpha \geqslant \alpha_x$ and all $y \in V_x$.

From $X = \bigcup_{x \in X} V_x$ and the compactness of $X$, we see that there exist $x_1, \ldots, x_k$ in $X$ with $X = \bigcup_{i=1}^k V_{x_i}$. Now choose some index $\alpha_0$ satisfying $\alpha_0 \geqslant \alpha_{x_i}$ for all $i = 1, \ldots, k$ and note that $\alpha \geqslant \alpha_0$ implies $0 \leqslant f_\alpha(y) < \varepsilon$ for all $y \in X$. That is, the net $\{f_\alpha\}$ converges uniformly to zero. ∎

## 2.16 Locally compact spaces

A topological space is **locally compact** if every point has a compact neighborhood.[8] The existence of a single compact neighborhood at each point is enough to guarantee many more.

**2.67 Theorem (Compact neighborhood base)** *In a locally compact Hausdorff space, every neighborhood of a point includes a compact neighborhood of the point. Consequently, in a locally compact Hausdorff space, each point has a neighborhood base of compact neighborhoods.*

*Proof*: Let $G$ be an open neighborhood of $x$ and let $W$ be a compact neighborhood of $x$. If $W \subset G$, we are done, so assume $A = W \cap G^c \neq \varnothing$. For each $y \in A$ choose an open neighborhood $U_y$ of $y$ and an open neighborhood $W_y$ of $x$ satisfying $W_y \subset W$ and $U_y \cap W_y = \varnothing$. Since $A$ $(= W \cap G^c)$ is compact, there exist $y_1, \ldots, y_k \in A$ such that $A \subset \bigcup_{i=1}^k U_{y_i}$. Put $V = \bigcap_{i=1}^k W_{y_i}$ and $U = \bigcup_{i=1}^k U_{y_i}$. Now $V$ is an open neighborhood of $x$, and we claim that $\overline{V}$ is compact and included in $G$.

To see this, note first that $\overline{V} \subset W$ implies that $\overline{V}$ is compact. Now, since $U$ and $V$ are both open and $V \cap U = \varnothing$, it follows that $\overline{V} \cap U = \varnothing$. Consequently, from

$$\overline{V} \cap G^c = \overline{V} \cap (W \cap G^c) = \overline{V} \cap A \subset \overline{V} \cap U = \varnothing,$$

we see that $\overline{V} \cap G^c = \varnothing$. Hence $\overline{V} \subset G$ is a compact neighborhood of $x$. ∎

Every compact space is locally compact. In fact, the following corollary is easily seen to be true.

**2.68 Corollary** *The intersection of an open subset with a closed subset of a locally compact Hausdorff space is locally compact.*

*In particular, every open subset and every closed subset of a locally compact Hausdorff space is locally compact.*

[8] Some authors require that a locally compact space be Hausdorff.

The next result is another useful corollary.

**2.69 Corollary** *If K is a compact subset of a locally compact Hausdorff space, and G is an open set including K, then there is an open set V with compact closure satisfying* $K \subset V \subset \overline{V} \subset G$.

*Proof*: By Theorem 2.67, each point $x$ in $K$ has an open neighborhood $V_x$ with compact closure satisfying $x \in V_x \subset \overline{V}_x \subset G$. Since $K$ is compact there is a finite subcollection $\{V_{x_1}, \dots, V_{x_n}\}$ of these sets covering $K$. Then $V = \bigcup_{i=1}^{n} V_{x_i}$ is the desired open set. (Why?) ∎

A **compactification** of a Hausdorff space $X$ is a compact Hausdorff space $Y$ where $X$ is homeomorphic to a dense subset of $Y$, so we may treat $X$ as an actual dense subset of $Y$. Note that if $X$ is already compact, then it is closed in any Hausdorff space including it, so any compactification of a compact Hausdorff space is the space itself. The locally compact Hausdorff spaces are open sets in all of their compactifications. The details follow.

**2.70 Theorem** *Let* $\hat{X}$ *be a compactification of a Hausdorff space X. Then X is locally compact if and only if X is an open subset of* $\hat{X}$.

*In particular, if X is a locally compact Hausdorff space, then X is an open subset of any of its compactifications.*

*Proof*: Let $(\hat{X}, \hat{\tau})$ be a compactification of a Hausdorff space $(X, \tau)$. If $X$ is an open subset of $\hat{X}$, then it follows from Corollary 2.68 that $X$ is locally compact. For the converse, assume that $(X, \tau)$ is locally compact and fix $x \in X$. Choose a compact $\tau$-neighborhood $U$ of $x$ and then pick an open $\tau$-neighborhood $V$ of $x$ such that $V \subset U$. Now select $W \in \hat{\tau}$ such that $V = W \cap X$ and note that

$$W = W \cap \hat{X} = W \cap \overline{X} \subset \overline{W \cap X} = \overline{V} \subset \overline{U} = U \subset X.$$

This shows that $x$ in a $\hat{\tau}$-interior point of $X$, so $X \in \hat{\tau}$. ∎

**2.71 Corollary** *Only locally compact Hausdorff spaces can possibly be compactified with a finite number of points.*

The simplest compactification of a noncompact locally compact Hausdorff space is its one-point compactification. It is obtained by appending a point $\infty$, called the **point at infinity**, that does not belong to the space $X$, and we write $X_\infty$ for $X \cup \{\infty\}$. We leave the proof of the next theorem as an exercise.

**2.72 Theorem (One-point compactification)** *Let* $(X, \tau)$ *be a noncompact locally compact Hausdorff space and let* $X_\infty = X \cup \{\infty\}$, *where* $\infty \notin X$. *Then the collection*

$$\tau_\infty = \tau \cup \{X_\infty \setminus K : K \subset X \text{ is compact}\}$$

*is a topology on* $X_\infty$. *Moreover,* $(X_\infty, \tau_\infty)$ *is a compact Hausdorff space and X is an open dense subset of* $X_\infty$, *that is,* $X_\infty$ *is a compactification of X.*

The space $(X_\infty, \tau_\infty)$ is called the **Alexandroff one-point compactification** of $X$. As an example, the one-point compactification $\mathbb{R}_\infty$ of the real numbers $\mathbb{R}$ is homeomorphic to a circle.

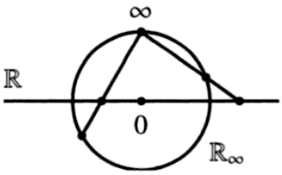


Figure 2.1. $\mathbb{R}_\infty$ is a circle.

One such homeomorphism is described by mapping the "north pole" $(0, 1)$ on the unit circle in $\mathbb{R}^2$ to $\infty$ and every other point $(x, y)$ on the circle is mapped to the point on the $x$-axis where the ray through $(x, y)$ from $\infty$ crosses the axis. See Figure 2.1. Mapmakers have long known that the one-point compactification of $\mathbb{R}^2$ is the sphere. (Look up stereographic projection in a good dictionary.)

It is immediate from Theorem 2.72 that a subset $F$ of $X$ is closed in $X_\infty$ if and only if $F$ is compact. We also have the following observation.

**2.73 Lemma** *For a subset $A$ of $X$, the set $A \cup \{\infty\}$ is closed in $X_\infty$ if and only if $A$ is closed in $X$.*

*Proof*: To see this, just note that $X_\infty \setminus (A \cup \{\infty\}) = X \setminus A$. ∎

The one-point compactification allows us to prove the following.

**2.74 Corollary** *In a locally compact Hausdorff space, nonempty compact sets can be separated from disjoint nonempty closed sets by continuous functions. In particular, every locally compact Hausdorff space is completely regular.*

*Proof*: Let $A$ be a nonempty compact subset and $B$ a nonempty closed subset of a locally compact Hausdorff space $X$ satisfying $A \cap B = \emptyset$. Then $A$ is a compact (and hence closed) subset of the one-point compactification $X_\infty$ of $X$. Let $C = B \cup \{\infty\}$. Then $C$ is a closed subset of $X_\infty$ (why?) and $A \cap C = \emptyset$.

Since $X_\infty$ is a compact Hausdorff space, it is normal by Theorem 2.48. Now by Theorem 2.46 there exists a continuous function $f\colon X_\infty \to [0, 1]$ satisfying $f(x) = 1$ for all $x \in A$ and $f(y) = 0$ for all $y \in C$. Clearly, the restriction of $f$ to $X$ has the desired properties. ∎

**2.75 Example (Topology of the extended reals)** The extended real numbers $\mathbb{R}^* = [-\infty, \infty]$ are naturally topologized as a two-point compactification of the space $\mathbb{R}$ of real numbers. A neighborhood base of $\infty$ is given by the collection of intervals of the form $(c, \infty]$ for $c \in \mathbb{R}$, and the intervals $[-\infty, c)$ constitute a neighborhood base for $-\infty$. Note that a sequence $\{x_n\}$ in $\mathbb{R}^*$ converges to $\infty$ if for every $n \in \mathbb{N}$, there exits an $n_0$ such that for all $n \geqslant n_0$ we have $x_n > m$. You should verify that this is indeed a compact space, that it is first countable, and that $\mathbb{R}$ is a dense subspace of $\mathbb{R}^*$. In fact by Theorem 3.40 it is metrizable. You should further check that an extended real-valued function that is both upper and lower semicontinuous is continuous with respect to this topology. ∎

for each $h \in C_b(Y)$, where we use the notation $\Gamma\mu$ rather than $\Gamma(\mu)$ to denote the value of $\Gamma$ at $\mu \in \mathbb{R}^{C_b(X)}$. We claim that $\Gamma$ is a continuous function. To see this, let $\{\mu_\alpha\}$ be a net in $\mathbb{R}^{C_b(X)}$ and suppose $\mu_\alpha \to \mu$ pointwise on $C_b(X)$. This means that $\mu_\alpha(f) \to \mu(f)$ in $\mathbb{R}$ for each $f$ in $C_b(X)$. In particular, $\mu_\alpha(h \circ g) \to \mu(h \circ g)$ for each $h \in C_b(Y)$. Thus

$$\Gamma\mu_\alpha(h) = \mu_\alpha(h \circ g) \to \mu(h \circ g) = \Gamma\mu(h),$$

or $\Gamma\mu_\alpha \to \Gamma\mu$ pointwise on $C_b(Y)$. Thus $\Gamma$ is continuous.

Now notice that for $x \in X$,

$$\Gamma e_x(h) = e_x(h \circ g) = h(g(x)) = e_{g(x)}(h)$$

for every $h \in C_b(Y)$, so identifying $x$ with $\varepsilon_X(x)$ and $g(x)$ with $\varepsilon_Y(g(x))$, we have

$$\Gamma(x) = g(x).$$

That is, $\Gamma$ extends $g$. Using Theorem 2.27 (5), we see that

$$\Gamma(\beta X) = \Gamma(\overline{\varepsilon_X(X)}) \subset \overline{\Gamma(\varepsilon_X(X))} \subset \overline{\varepsilon_Y(Y)} = \varepsilon_Y(Y).$$

Thus, $\Gamma$ is the unique continuous extension of $g$ to all of $\beta X$. ∎

There are a number of important corollaries.

**2.80 Corollary (Uniqueness)** *Let $K$ be a compactification of a completely regular Hausdorff space $X$ and suppose that whenever $Y$ is a compact Hausdorff space and $g\colon X \to Y$ is continuous, then $g$ has a unique continuous extension from $K$ to $Y$. Then $K$ is homeomorphic to $\beta X$.*

*Proof*: Take $Y = \beta X$ in Theorem 2.79. ∎

It is a good mental workout to imagine an element of $\beta X = \overline{\varepsilon(X)}$ that does not belong to $\varepsilon(X)$. For a real function $\mu$ on $C_b(X)$ to belong to $\overline{\varepsilon(X)}$, there must be a net $\{x_\alpha\}$ in $X$ with $e_{x_\alpha} \to \mu$ pointwise on $C_b$. That is, for each $f \in C_b(X)$, we have $f(x_\alpha) \to \mu(f)$. If $\{x_\alpha\}$ converges, say to $x$, since $\varepsilon$ is an embedding, we conclude $\mu = e_x$, which belongs to $\varepsilon(X)$. Thus if $\mu$ belongs to $\overline{\varepsilon(X)} \setminus \varepsilon(X)$ it cannot be the case that the net $\{x_\alpha\}$ converges. On the other hand, $\{x_\alpha\}$ must have a limit point in any compactification of $X$. Let $x_0$ be a limit point of $\{x_\alpha\}$ in $\beta X$. Then $\mu$ acts like an evaluation at $x_0$.

Thus we can think of the Stone–Čech compactification $\beta X$ as adding limit points to all the nets in $X$ in such a way that every $f$ in $C_b(X)$ extends continuously to $\beta X$.[10] Indeed it is characterized by this extension property.

**2.81 Corollary** *Let $K$ be a compactification of a completely regular Hausdorff space $X$ and suppose that every bounded continuous real function on $X$ has a (unique) continuous extension from $X$ to $K$. Then $K$ is homeomorphic to $\beta X$.*

[10] Professional topologists express this with the phrase "$X$ is $C^*$-embedded in $\beta X$."

*Proof*: Given any $f \in C_b(X)$, let $\hat{f}$ denote its continuous extension to $K$. Since the restriction of a continuous function on $K$ is a bounded continuous function on $X$, the mapping $f \mapsto \hat{f}$ from $C_b(X)$ to $C(K)$ is one-to-one and onto.

Define the mapping $\varphi$ from $K$ into $\mathbb{R}^{C_b(X)}$ by $\varphi_x(f) = \hat{f}(x)$. Observe that $\varphi$ is continuous. Furthermore $\varphi$ is one-to-one. To see this, suppose $\varphi_x = \varphi_y$, that is, $\hat{f}(x) = \hat{f}(y)$ for every $f \in C_b(X)$. Then $f(x) = f(y)$ for every $f \in C(K)$. But $C(K)$ separates points of $K$ (why?), so $x = y$. Consequently, $\varphi$ is a homeomorphism from $K$ to $\varphi(K)$ (Theorem 2.36).

Treating $X$ as a dense subset of $K$, observe that if $x$ belongs to $X$, then $\varphi_x$ is simply the evaluation at $x$, so by definition, $\overline{\varphi(X)}$ is the Stone–Čech compactification of $X$. Since $X$ is dense, $\varphi(X) \subset \varphi(K) \subset \overline{\varphi(X)}$. But $\varphi(K)$ is compact and therefore closed. Thus $\varphi(K) = \overline{\varphi(X)}$, and we are done. ■

We take this opportunity to describe the Stone–Čech compactification of the space $\Omega_0 = \Omega \setminus \{\omega_1\}$ of countable ordinals. Recall that it is an open subset of the compact Hausdorff space $\Omega$ of ordinals, and thus locally compact. We start with the following peculiar property of continuous functions on $\Omega_0$.

**2.82 Lemma (Continuous functions on $\Omega_0$)** *Any continuous real function on $\Omega_0 = \Omega \setminus \{\omega_1\}$ is constant on some tail of $\Omega_0$. That is, if $f$ is a continuous real function $\Omega_0$, there is an ordinal $x \in \Omega_0$ such that $y \geq x$ implies $f(y) = f(x)$.*

*Proof*: We start by making the following observation. If $f\colon \Omega_0 \to \mathbb{R}$ is continuous, and $a > b \in \mathbb{R}$, then at least one of $[f \geqslant a]$ or $[f \leqslant b]$ is countable. To see this, suppose that both are uncountable. Pick $x_1 \in \Omega_0$ so that $f(x_1) \geqslant a$. Since the initial segment $I(x_1)$ is countable, there is some $y_1 > x_1$ with $f(y_1) \leqslant b$. Proceeding in this fashion we can construct two interlaced sequences satisfying $x_n < y_n < x_{n+1}$, $f(x_n) \geqslant a$, and $f(y_n) \leqslant b$ for all $n$. By the Interlacing Lemma 1.15, these sequences have a common least upper bound $z$, which must then be the limit of each sequence. Since $f$ is continuous, we must have $f(z) = \lim f(x_n) \geqslant a$ and $f(z) = \lim f(y_n) \leqslant b$, a contradiction. Therefore at least one set is countable.

Since $\Omega_0$ is uncountable, there is some (possibly negative) integer $k$, such that the set $[k \leqslant f \leqslant k+1]$ is uncountable. Since $[f \geqslant k]$ and $[f \leqslant k+1]$ are uncountable, by the observation above we see that for each positive $n$, the sets $[f \leqslant k - \frac{1}{n}]$ and $[f \geqslant k + 1 + \frac{1}{n}]$ are countable. So except for countably many $x$, we have $k \leqslant f(x) \leqslant k+1$. Let $I_1 = [k, k+1]$. Now divide $I_1$ in half. Then either $[k \leqslant f \leqslant k + \frac{1}{2}]$ or $[k + \frac{1}{2} \leqslant f \leqslant k+1]$ is uncountable. (Both sets may be uncountable, for instance, if $f$ is constant with value $k + \frac{1}{2}$.) Without loss of generality, assume $[k \leqslant f \leqslant k + \frac{1}{2}]$ is uncountable, and set $I_2 = [k, k + \frac{1}{2}]$. Observe that $\{x \in \Omega_0 : f(x) \notin I_2\}$ is countable. Proceeding in this way we can find a nested sequence $\{I_n\}$ of closed real intervals, with the length of $I_n$ being $\frac{1}{2^n}$, and having the property that $\{x \in \Omega_0 : f(x) \notin I_n\}$ is countable. Let $a$ denote the unique point in $\bigcap_{n=1}^{\infty} I_n$. Then $\{x \in \Omega_0 : f(x) \neq a\}$ is countable. By Theorem 1.14 (6), this set has a least upper bound $b$. Now pick any $x > b$. Then $y \geq x$ implies $f(y) = a$. ■

We now come to the compactifications of $\Omega_0$.

**2.83 Theorem (Compactification of $\Omega_0$)** *The compact Hausdorff space $\Omega$ can be identified with both the Stone–Čech compactification and the one-point compactification of $\Omega_0$.*

*Proof*: The identification with the one-point compactification is straightforward. Now note that by Lemma 2.82, every continuous real function on $\Omega_0$ has a unique continuous extension to $\Omega$. Thus by Corollary 2.81, we can identify $\Omega$ with the Stone–Čech Compactification of $\Omega_0$. ∎

There are some interesting observations that follow from this. Since $\Omega$ is compact, this means that every continuous real function on $\Omega_0$ is bounded, even though $\Omega_0$ is not compact. (The open cover $\{[1, x) : x \in \Omega_0\}$ has no finite subcover.) Since every initial segment of $\Omega_0$ is countable, we also see that every continuous real function on $\Omega$ takes on only countably many values.

We observed above that $f \mapsto \hat{f}$ from $C_b(X)$ into $C(\beta X)$ is one-to-one and onto. In addition, for $f, g \in C_b(X)$ it is easy to see that:

1. $(f + g)\widehat{\ } = \hat{f} + \hat{g}$ and $(\alpha f)\widehat{\ } = \alpha \hat{f}$ for all $\alpha \in \mathbb{R}$;

2. $(\max\{f, g\})\widehat{\ } = \max\{\hat{f}, \hat{g}\}$ and $(\min\{f, g\})\widehat{\ } = \min\{\hat{f}, \hat{g}\}$; and

3. $\|f\|_\infty = \sup\{|f(x)| : x \in X\} = \sup\{|f(x)| : x \in \beta X\} = \|\hat{f}\|_\infty$.

In Banach lattice terminology (see Definition 9.16), these properties are summarized as follows.

**2.84 Corollary** *If $X$ is a completely regular Hausdorff space, then the mapping $f \mapsto \hat{f}$ is a lattice isometry from $C_b(X)$ onto $C(\beta X)$. That is, under this identification, $C_b(X) = C(\beta X)$.*

Getting ahead of ourselves a bit, we note that $C_b(X)$ is an AM-space with unit, so by Theorem 9.32 it is lattice isometric to $C(K)$ for some compact Hausdorff space $K$. According to Corollary 2.84 the space $K$ is just the Stone–Čech compactification $\beta X$.

Unlike the one-point compactification, which is often very easy to describe, the Stone–Čech compactification can be very difficult to get a handle on. For instance, the Stone–Čech compactification of $(0, 1]$ is not homeomorphic to $[0, 1]$. The real function $\sin(\frac{1}{x})$ is bounded and continuous on $(0, 1]$, but cannot be extended to a continuous function on $[0, 1]$. However, for discrete spaces, such as the natural numbers $\mathbb{N}$, there is an interesting interpretation of the Stone–Čech compactification described in the next section.

## 2.18 Stone–Čech compactification of a discrete set

In this section we characterize the Stone–Čech compactification of a discrete space. Any discrete space $X$ is metrizable by the discrete metric, and hence completely regular and Hausdorff. Thus it has a Stone–Čech compactification $\beta X$. Since every set is open in a discrete space, every such space $X$ is **extremally disconnected**, that is, it has the property that the closure of every open set is itself open. It turns out that $\beta X$ inherits this property.

**2.85 Theorem** *For an infinite discrete space $X$:*

1. *If $A$ is a subset of $X$, then $\overline{A}$ is an open subset of $\beta X$, where the bar denotes the closure in $\beta X$.*
2. *If $A, B \subset X$ satisfy $A \cap B = \varnothing$, then $\overline{A} \cap \overline{B} = \varnothing$.*
3. *The space $\beta X$ is extremally disconnected.*

*Proof*: (1 & 2) Let $A \subset X$. Put $C = X \setminus A$ and note that $A \cap C = \varnothing$. Define $f\colon X \to [0,1]$ by $f(x) = 1$ if $x \in A$ and $f(x) = 0$ if $x \in C$. Clearly, $f$ is continuous, so it extends uniquely to a continuous function $\hat{f}\colon \beta X \to [0,1]$. From $A \cup C = X$, we get $\overline{A} \cup \overline{C} = \beta X$. (Do you see why?) It follows that $\overline{A} = \hat{f}^{-1}(\{1\})$ and $\overline{C} = \hat{f}^{-1}(\{0\})$. Therefore, $\overline{A} \cap \overline{C} = \varnothing$, and $\overline{A}$ is open. Now if $B \subset X$ satisfies $A \cap B = \varnothing$, then $B \subset C$, so $\overline{A} \cap \overline{B} = \varnothing$.

(3) Let $V$ be an open subset of $\beta X$. By (1), the set $\overline{V \cap X}$ is an open subset of $\beta X$. Note that if $x \in \overline{V}$ and $W$ is an open neighborhood of $x$, then $W \cap V \neq \varnothing$, so $W \cap V \cap X \neq \varnothing$, or $x \in \overline{V \cap X}$. Therefore, $\overline{V} = \overline{V \cap X}$, so that $\overline{V}$ is open. ∎

Let $\mathfrak{U}$ denote the set of all ultrafilters on $X$. That is,

$$\mathfrak{U} = \{\mathcal{U} : \mathcal{U} \text{ is an ultrafilter on } X\}.$$

As we already know, ultrafilters on $X$ are either fixed or free. Every $x \in X$ gives rise to a unique fixed ultrafilter $\mathcal{U}_x$ on $X$ via the formula

$$\mathcal{U}_x = \{A \subset X : x \in A\},$$

and every fixed ultrafilter on $X$ is of the form $\mathcal{U}_x$.

Now let $\mathcal{U}$ be a free ultrafilter on $X$. Then $\mathcal{U}$ is a filter base in $\beta X$. Thus the filter $\mathcal{F}$ it generates has a limit point in $\beta X$ (Theorem 2.31). That is, we have $\bigcap_{F\in\mathcal{F}} \overline{F} = \bigcap_{A\in\mathcal{U}} \overline{A} \neq \varnothing$. We claim that this intersection is a singleton. To see this, assume that there exist $x, y \in \bigcap_{A\in\mathcal{U}} \overline{A}$ with $x \neq y$. Then the collections

$$\mathcal{B}_x = \{V \cap A : V \in \mathcal{N}_x, A \in \mathcal{U}\} \quad \text{and} \quad \mathcal{B}_y = \{W \cap B : W \in \mathcal{N}_y, B \in \mathcal{U}\},$$

are both filter bases on $X$. Since the filters they generate include the ultrafilter $\mathcal{U}$, it follows that $\mathcal{B}_x \cup \mathcal{B}_y \subset \mathcal{U}$. Since $\beta X$ is a Hausdorff space, there exist $V \in \mathcal{N}_x$

and $W \in \mathcal{N}_y$ such that $V \cap W = \varnothing$. This implies $\varnothing \in \mathcal{U}$, a contradiction. Hence, $\bigcap_{A\in\mathcal{U}} \overline{A}$ is a singleton.

Conversely, if $x \in \beta X \setminus X$, then the collection

$$\mathcal{B} = \{V \cap X : V \in \mathcal{N}_x\} \qquad (\star)$$

of subsets of $X$ is a filter base on $X$. By Zorn's Lemma there exists an ultrafilter $\mathcal{U}$ on $X$ including $\mathcal{B}$. Then $\mathcal{U}$ is a free ultrafilter (on $X$) satisfying $\bigcap_{A\in\mathcal{U}} \overline{A} = \{x\}$. (Why?) In other words, every point of $\beta X \setminus X$ is the limit point of a free ultrafilter on $X$.

It turns out that every point of $\beta X \setminus X$ is the limit point of exactly one free ultrafilter on $X$. To see this, let $\mathcal{U}_1$ and $\mathcal{U}_2$ be two free ultrafilters on $X$ such that $x \in \bigcap_{A\in\mathcal{U}_1} \overline{A} = \bigcap_{B\in\mathcal{U}_2} \overline{B}$. If $A \in \mathcal{U}_1$, then $A \in \mathcal{U}_2$. Otherwise, $A \notin \mathcal{U}_2$ implies $X \setminus A \in \mathcal{U}_2$, so (by Theorem 2.85) $x \in \overline{A} \cap \overline{X \setminus A} = \varnothing$, a contradiction. So $\mathcal{U}_1 \subset \mathcal{U}_2$. Similarly, $\mathcal{U}_2 \subset \mathcal{U}_1$, and hence $\mathcal{U}_1 = \mathcal{U}_2$.

For each point $x \in \beta X \setminus X$, we denote by $\mathcal{U}_x$ the unique free ultrafilter on the set $X$—whose filter base is given by ($\star$)—having $x$ as its unique limit point. Thus, we have established a one-to-one mapping $x \mapsto \mathcal{U}_x$ from $\beta X$ onto the set $\mathfrak{U}$ of all ultrafilters on $X$, where the points of $X$ correspond to the fixed ultrafilters and the points of $\beta X \setminus X$ to the free ultrafilters.

We can describe the topology on $\beta X$ in terms of $\mathfrak{U}$: For each subset $A$ of $X$, let

$$\mathfrak{U}_A = \{\mathcal{U} \in \mathfrak{U} : A \notin \mathcal{U}\}.$$

The collection $\mathcal{A} = \{\mathfrak{U}_A : A \subset X\}$ enjoys the following properties.

a. $\mathfrak{U}_\varnothing = \mathfrak{U}$ and $\mathfrak{U}_X = \varnothing$.

b. $\mathfrak{U}_A \cap \mathfrak{U}_B = \mathfrak{U}_{A\cup B}$ and $\mathfrak{U}_A \cup \mathfrak{U}_B = \mathfrak{U}_{A\cap B}$.

From properties (a) and (b), we see that $\mathcal{A}$ is a base for a topology $\tau$. This topology is called the **hull-kernel topology**.[11] The topological space $(\mathfrak{U}, \tau)$ is referred to as the **ultrafilter space** of $X$.

The ultrafilter space is a Hausdorff space. To see this, let $\mathcal{U}_1 \neq \mathcal{U}_2$. Then there exists some $A \in \mathcal{U}_1$ with $A \notin \mathcal{U}_2$ (or vice versa), so $B = X \setminus A \notin \mathcal{U}_1$. Hence $\mathcal{U}_2 \in \mathfrak{U}_A$ and $\mathcal{U}_1 \in \mathfrak{U}_B$, while $\mathfrak{U}_A \cap \mathfrak{U}_B = \mathfrak{U}_{A\cup B} = \mathfrak{U}_X = \varnothing$.

And now we have the main result of this section: The ultrafilter space with the hull-kernel topology is homeomorphic to the Stone–Čech compactification of $X$.

**2.86 Theorem** *For a discrete space $X$, the mapping $x \mapsto \mathcal{U}_x$ is a homeomorphism from $\beta X$ onto $\mathfrak{U}$. So $\beta X$ can be identified with the ultrafilter space $\mathfrak{U}$ of $X$.*

[11] See, e.g., W. A. J. Luxemburg and A. C. Zaanen [235, Chapter 1] for an explanation of the name.

**2.92 Lemma** *Let $\mathfrak{U}$ be an open cover of a compact Hausdorff space X. Then there is a locally finite family $\{f_U\}_{U\in\mathfrak{U}}$ of real functions such that:*

1. *$f_U : X \to [0,1]$ is continuous for each U.*

2. *$f_U$ vanishes on $U^c$.*

3. *$\sum_{U\in\mathfrak{U}} f_U(x) = 1$ for all $x \in X$.*

*That is, $\{f_U\}_{U\in\mathfrak{U}}$ is a continuous locally finite partition of unity subordinated to $\mathfrak{U}$.*

*Proof*: For each $x$ pick a neighborhood $U_x \in \mathfrak{U}$ of $x$. By Theorem 2.48, the space $X$ is normal, so by Urysohn's Lemma 2.46, for each $x$ there is a continuous real function $g_x : X \to [0,1]$ satisfying $g_x = 0$ on $U_x^c$ and $g_x(x) = 1$. The set $V_x = \{z \in X : g_x(z) > 0\}$ is an open neighborhood of $x$, so $\{V_x : x \in X\}$ is an open cover of $X$. Thus there is a finite subcover $\{V_{x_1}, \dots, V_{x_n}\}$. Observe that $g_{x_j}(z) > 0$ for each $z \in V_{x_j}$ and vanishes outside $U_{x_j}$. Define $g$ by $g(z) = \sum_{j=1}^n g_{x_j}(z)$ and note that $g(z) > 0$ for every $z \in X$. Replacing $g_{x_j}$ by $g_{x_j}/g$, we can assume that $\sum_{j=1}^n g_{x_j}(z) = 1$ for each $z \in X$.

Finally, put $f_U = \sum_{\{i : U_{x_i} = U\}} g_{x_i}$ (if $\{i : U_{x_i} = U\} = \varnothing$, we let $f_U = 0$), and note that the family $\{f_U\}_{U\in\mathfrak{U}}$ of real functions satisfies the desired properties. ∎

Theorem 3.22 below shows that metric spaces are paracompact.

## Chapter 3

# Metrizable spaces

In Chapter 2 we introduced topological spaces to handle problems of convergence that metric spaces could not. Nevertheless, every sane person would rather work with a metric space if they could. The reason is that the metric, a real-valued function, allows us to analyze these spaces using what we know about the real numbers. That is why they are so important in real analysis. We present here some of the more arcane results of the theory of metric spaces. Most of this material can be found in some form in K. Kuratowski's [218] tome. Many of these results are the work of Polish mathematicians in the 1920s and 1930s. For this reason, a complete separable metric space is called a *Polish space*.

Here is a guide to the major points of interest in the territory covered in this chapter. The distinguishing features of the theory of metric spaces, which are absent from the general theory of topology, are the notions of *uniform continuity* and *completeness*. These are not topological notions, in that there may be two *equivalent* metrics inducing the same topology, but they may have different uniformly continuous functions, and one may be complete while the other isn't. Nevertheless, if a topological space is *completely metrizable*, there are some topological consequences. One of these is the Baire Category Theorem 3.47, which asserts that in a completely metrizable space, the countable intersection of open dense sets is dense. Complete metric spaces are also the home of the Contraction Mapping Theorem 3.48, which is one of the fundamental theorems in the theory of dynamic programming (see the book by N. Stokey, R. E. Lucas, and E. C. Prescott [322].)

Lemma 3.23 embeds an arbitrary metric space in the Banach space of its bounded continuous real-valued functions. This result is useful in characterizing complete metric spaces. By the way, all the Euclidean spaces are complete.

In a metric space, it is easy to show that second countability and separability are equivalent (Lemma 3.4). The Urysohn Metrization Theorem 3.40 asserts that every second countable regular Hausdorff is metrizable, and that this property is equivalent to being embedded in the *Hilbert cube*. This leads to a number of properties of separable metrizable spaces. Another useful property is that in metric spaces, a set is compact if and only if it is sequentially compact (Theorem 3.28).

We also introduce the compact metric space called the *Cantor set*. It can be viewed as a subset of the unit interval, but every compact metric space is the image of the Cantor set under a continuous function. In the same vein, we study the *Baire space* of sequences of natural numbers. It is a Polish space, and every Polish space is a continuous image of it. It is also the basis for the study of *analytic sets*, which we describe in Section 12.5.

We also discuss topologies for spaces of subsets of a metric space. The most straightforward way to topologize the collection of nonempty closed subsets of a metric space is through the Hausdorff metric. Unfortunately, this technique is not topological. That is, the topology on the space of closed subsets may be different for different compatible metrics on the underlying space (Example 3.86). However, restricted to the compact subsets, the topology is independent of the compatible metric (Theorem 3.91). Since every locally compact separable metrizable space has a metrizable compactification (Corollary 3.45), for this class of spaces there is a nice topological characterization of the *topology of closed convergence* on the space of closed subsets (Corollary 3.95). Once we have a general method for topologizing subsets, our horizons are greatly expanded. For example, since binary relations are just subsets of Cartesian products, they can be topologized in a useful way; see A. Mas-Colell [240]. As another example, F. H. Page [268] uses a space of sets in order to prove the existence of an optimal incentive contract.

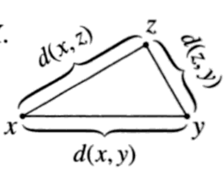
Finally, we conclude with a discussion of the space $C(X,Y)$ of continuous functions from a compact space into a metrizable space under the topology of uniform convergence. It turns out that this topology depends only on the topology of $Y$ and not on any particular metric (Lemma 3.98). The space $C(X,Y)$ is complete if $Y$ is complete, and separable if $Y$ is separable; see Lemmas 3.97 and 3.99.

## 3.1 Metric spaces

Recall the following definition from Chapter 2.

**3.1 Definition** *A **metric** (or **distance**) on a set $X$ is a function $d\colon X \times X \to \mathbb{R}$ satisfying the following four properties:*

1. *Positivity: $d(x,y) \geqslant 0$ and $d(x,x) = 0$ for all $x, y \in X$.*
2. *Discrimination: $d(x,y) = 0$ implies $x = y$.*
3. *Symmetry: $d(x,y) = d(y,x)$ for all $x, y \in X$.*




4. *The Triangle Inequality: $d(x,y) \leqslant d(x,z) + d(z,y)$ for all $x, y, z \in X$.*

*A **semimetric** on $X$ is a function $d\colon X \times X \to \mathbb{R}$ satisfying (1), (3), and (4). Obviously, every metric is a semimetric. If $d$ is a metric on a set $X$, then the pair $(X,d)$ is called a **metric space**, and similarly if $d$ is a semimetric, then $(X,d)$ is a **semimetric space**.*

If $d$ is a semimetric, then the binary relation defined by $x \sim y$ if $d(x, y) = 0$ is an equivalence relation, and $d$ defines a metric $\hat{d}$ on the set of equivalence classes by $\hat{d}([x], [y]) = d(x, y)$. For this reason we deal mostly with metric spaces. *Be aware that when we define a concept for metric spaces, there is nearly always a corresponding notion for semimetric spaces, even if we do not explicitly mention it.* The next definition is a good example.

For a nonempty subset $A$ of a metric space $(X, d)$ its **diameter** is defined by

$$\operatorname{diam} A = \sup\{d(x, y) : x, y \in A\}.$$

A set $A$ is **bounded** if $\operatorname{diam} A < \infty$, while $A$ is **unbounded** if $\operatorname{diam} A = \infty$. If $\operatorname{diam} X < \infty$, then $X$ is **bounded** and $d$ is called a **bounded metric**. Similar terminology applies to semimetrics.

In a semimetric space $(X, d)$ the **open ball** centered at a point $x \in X$ with radius $r > 0$ is the subset $B_r(x)$ of $X$ defined by

$$B_r(x) = \{y \in X : d(x, y) < r\}.$$

The **closed ball** centered at a point $x \in X$ with radius $r > 0$ is the subset $C_r(x)$ of $X$ defined by

$$C_r(x) = \{y \in X : d(x, y) \leqslant r\}.$$

**3.2 Definition** *Let $(X, d)$ be a semimetric space. A subset $A$ of $X$ is **d-open** (or simply **open**) if for each $a \in A$ there exists some $r > 0$ (depending on $a$) such that $B_r(a) \subset A$.*

You should verify that the collection of subsets

$$\tau_d = \{A \subset X : A \text{ is } d\text{-open}\}$$

is a topology on $X$, called the **topology generated** or **induced** by $d$. When $d$ is a metric, we call $\tau_d$ the **metric topology** on $(X, d)$. A topological space $(X, \tau)$ is **metrizable** if the topology $\tau$ is generated by some metric. A metric generating a topology is called **compatible** or **consistent** with the topology. Two metrics generating the same topology are **equivalent**.

We have already seen a number of examples of metrizable spaces and compatible metrics in Example 2.2. There are always several metrics on any given set that generate the same topology. Let $(X, d)$ be a metric space. Then $2d$ is also a metric generating the same topology. More interesting is the metric $\hat{d}(x, y) = \min\{d(x, y), 1\}$. It too generates the same open sets as $d$, but $X$ is bounded under $\hat{d}$. In fact, notice that the $\hat{d}$-diameter of $X$ is less than or equal to 1. A potential drawback of $\hat{d}$ is that the families of balls of radius $r$ around $x$ are different for $d$ and $\hat{d}$. (For instance, $\{x \in \mathbb{R} : |x| < 2\}$ is a ball of radius 2 around 0 in the usual metric on $\mathbb{R}$, but in the truncated metric it is not a ball of any finite radius.)

Lemma 3.6 below describes a bounded metric that avoids this criticism. The point of this lemma is that for most anything topological that we want to do with a metric space, it is no restriction to assume that its metric takes on values only in the unit interval $[0, 1]$.

The following lemma summarizes some of the basic properties of metric and semimetric topologies. The proofs are straightforward applications of the definitions. You should be able to do them without looking at the hints.

**3.3 Lemma (Semimetric topology)** *Let $(X, d)$ be a semimetric space. Then:*

1. *The topology $\tau_d$ is Hausdorff if and only if $d$ is a metric.*

2. *A sequence $\{x_n\}$ in $X$ satisfies $x_n \xrightarrow{\tau_d} x$ if and only if $d(x_n, x) \to 0$.*

3. *Every open ball is an open set.*

4. *The topology $\tau_d$ is first countable.*

5. *A point $x$ belongs to the closure $\overline{A}$ of a set $A$ if and only if there exists some sequence $\{x_n\}$ in $A$ with $x_n \to x$.*

6. *A closed ball is a closed set.*

7. *The closure of the open ball $B_r(x)$ is included in the closed ball $C_r(x)$. But the inclusion may also be proper.*

8. *If $(X, d_1)$ and $(Y, d_2)$ are semimetric spaces, the product topology on $X \times Y$ is generated by the semimetric*

$$\rho((x, y), (u, v)) = d_1(x, u) + d_2(y, v).$$

   *It is also generated by* $\max\{d_1(x, u), d_2(y, v)\}$ *and* $(d_1(x, u)^2 + d_2(y, v)^2)^{1/2}$.

9. *For any four points $u, v, x, y$, the semimetric obeys*

$$|d(x, y) - d(u, v)| \leqslant d(x, u) + d(y, v).$$

10. *The real function $d\colon X \times X \to \mathbb{R}$ is jointly continuous.*

*Hints*: The proofs of (1) and (2) are straightforward, and (5) follows from (4).

(3) Let $y$ belong to the open ball $B_r(x)$. Put $\varepsilon = r - d(x, y) > 0$. If $z \in B_\varepsilon(y)$, then the triangle inequality implies $d(x, z) \leqslant d(x, y) + d(y, z) < d(x, y) + \varepsilon = r$. So $B_\varepsilon(y) \subset B_r(x)$, which means that $B_r(x)$ is a $\tau_d$-open set.

(4) The countable family of open neighborhoods $\{B_{1/n}(x) : n \in \mathbb{N}\}$ is a base for the neighborhood system at $x$.

(6) Suppose $y \notin C_r(x)$. Then $\varepsilon = d(x, y) - r > 0$, so by the triangle inequality, $B_\varepsilon(y)$ is an open neighborhood of $y$ disjoint from $C_r(x)$. This shows that the complement of $C_r(x)$ is open.

A topological space $X$ is **completely metrizable** if there is a consistent metric $d$ for which $(X, d)$ is complete. A separable topological space that is completely metrizable is called a **Polish space**. Such a topology is called a **Polish topology**.

Here are some important examples of complete metric spaces.

- The space $\mathbb{R}^n$ with the Euclidean metric $d(x, y) = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{1/2}$ is a complete metric space.

- The discrete metric is always complete.

- Let $Y$ be a nonempty subset of a complete metric space $(X, d)$. Then $(Y, d|_Y)$ is a complete metric space if and only if $Y$ is a closed subset of $X$.

- If $X$ is a nonempty set, then the vector space $B(X)$ of all bounded real functions on $X$ is a complete metric space under the **uniform metric** defined by

$$d(f, g) = \sup_{x \in X} |f(x) - g(x)|.$$

It is clear that a sequence $\{f_n\}$ in $B(X)$ is $d$-convergent to $f \in B(X)$ if and only if it converges uniformly to $f$. First let us verify that $d$ is indeed a metric on $B(X)$. Clearly, $d$ satisfies the positivity, discrimination, and symmetry properties of a metric.

To see that $d$ satisfies the triangle inequality, note that if $f, g, h \in B(X)$, then for each $x \in X$ we have

$$|f(x) - g(x)| \leqslant |f(x) - h(x)| + |h(x) - g(x)| \leqslant d(f, h) + d(h, g).$$

Therefore, $d(f, g) = \sup_{x \in X} |f(x) - g(x)| \leqslant d(f, h) + d(h, g)$.

Now we establish that $(B(X), d)$ is complete. To this end, let $\{f_n\}$ be a $d$-Cauchy sequence in $B(X)$. This means that for each $\varepsilon > 0$ there exists some $k$ such that

$$|f_n(x) - f_m(x)| \leqslant d(f_n, f_m) < \varepsilon \qquad (\star)$$

for all $x \in X$ and all $n, m \geqslant k$. In particular, $\{f_n(x)\}$ is a Cauchy sequence of real numbers for each $x \in X$. Let $\lim f_n(x) = f(x) \in \mathbb{R}$ for each $x \in X$. To finish the proof we need to show that $f$ is bounded and so belongs to $B(X)$, and that $d(f_n, f) \to 0$. Pick some $M > 0$ such that $|f_k(x)| \leqslant M$ for each $x \in X$, and then use $(\star)$ to see that

$$|f(x)| \leqslant \lim_{m \to \infty} |f_m(x) - f_k(x)| + |f_k(x)| \leqslant \varepsilon + M$$

for each $x \in X$, so $f$ belongs to $B(X)$. Now another glance at $(\star)$ yields

$$|f_n(x) - f(x)| = \lim_{m \to \infty} |f_n(x) - f_m(x)| \leqslant \varepsilon$$

for all $n \geqslant k$. Hence $d(f_n, f) = \sup_{x \in X} |f_n(x) - f(x)| \leqslant \varepsilon$ for all $n \geqslant k$. This shows that $(B(X), d)$ is a complete metric space.

*Proof*: Since $\{A_n\}$ has vanishing diameter and $\bigcap_{n=1}^{\infty} A_n$ is nonempty, the intersection $\bigcap_{n=1}^{\infty} A_n$ must be some singleton $\{x\}$. Let $\varepsilon > 0$ be given. Since $f$ is continuous, there is some $\delta > 0$ such that $d(z, x) < \delta$ implies $\rho(f(z), f(x)) < \varepsilon$. Also there is some $n_0$ such that for all $n \geqslant n_0$, if $z \in A_n$, then $d(z, x) < \delta$. Thus for $n \geqslant n_0$, the image $f(A_n)$ is included in the ball of $\rho$-radius $\varepsilon$ around $f(x)$, so $\rho$- diam $f(A_n) \leqslant 2\varepsilon$. This shows that $\{f(A_n)\}$ has vanishing $\rho$-diameter—and also that $\bigcap_{n=1}^{\infty} f(A_n) = \{f(x)\}$. ∎

Note that the hypothesis that $\bigcap_{n=1}^{\infty} A_n$ is nonempty is necessary. For instance, consider $X = (0, 1]$ and $Y = \mathbb{R}$ with their usual metrics, let $A_n = (0, \frac{1}{n}]$, and let $f(x) = \sin \frac{1}{x}$. Then for each $n$, the image $f(A_n) = [-1, 1]$, which does not have vanishing diameter.

## 3.3 Uniformly continuous functions

Some aspects of metric spaces are not topological, but depend on the particular compatible metric. These properties include its uniformly continuous functions and Cauchy sequences. A function $f: (X, d) \to (Y, \rho)$ between two metric spaces is **uniformly continuous** if for each $\varepsilon > 0$ there exists some $\delta > 0$ (depending only on $\varepsilon$) such that $d(x, y) < \delta$ implies $\rho(f(x), f(y)) < \varepsilon$. Any uniformly continuous function is obviously continuous. An important property of uniformly continuous functions is that they map Cauchy sequences into Cauchy sequences. (The proof of this is a simple exercise.)

A function $f: (X, d) \to (Y, \rho)$ between metric spaces is **Lipschitz continuous** if there is some real number $c$ such that for every $x$ and $y$ in $X$,

$$\rho(f(x), f(y)) \leqslant c d(x, y).$$

The number $c$ is called a **Lipschitz constant** for $f$. Clearly every Lipschitz continuous function is uniformly continuous.

The set $X \times X$ has a natural metric $\rho$ given by $\rho((x, y), (u, v)) = d(x, u) + d(y, v)$. The metric $d$ can be viewed as a function from the metric space $(X \times X, \rho)$ to $\mathbb{R}$. Viewed this way, $d$ is Lipschitz continuous with Lipschitz constant 1 (and hence it is also a uniformly continuous function). This fact, which follows immediately from Property (9) of Lemma 3.3, may be used throughout this book without any specific reference.

An **isometry** between metric spaces $(X, d)$ and $(Y, \rho)$ is a one-to-one function $\varphi$ mapping $X$ into $Y$ satisfying

$$d(x, y) = \rho(\varphi(x), \varphi(y))$$

for all $x, y \in X$. If in addition $\varphi$ is surjective, then $(X, d)$ and $(Y, \rho)$ are **isometric**. If two metric spaces are isometric, then any property expressible in terms of metrics holds in one if and only if it holds in the other. Notice that isometries are uniformly continuous, indeed Lipschitz continuous.

Given a metric space $(X, d)$, denote by $U_d(X)$ or more simply, $U_d$, the collection of all bounded $d$-uniformly continuous real-valued functions on $X$. The set $U_d$ is a function space (recall Definition 1.1) that includes the constant functions.

In general, two different equivalent metrics determine different classes of uniformly continuous functions. For example, $x \mapsto \frac{1}{x}$ is not uniformly continuous on $(0, 1)$ under the usual metric, but it is uniformly continuous under the equivalent metric $d$ defined by $d(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|$.

The example just given is a particular instance of the following lemma on creating new metric spaces out of old ones. The proof of the lemma is a straightforward application of the definitions and is left as an exercise.

**3.9 Lemma** *Let $\varphi\colon (X, d) \to Y$ be one-to-one and onto. Then $\varphi$ induces a metric $\rho$ on $Y$ by $\rho(x, y) = d(\varphi^{-1}(x), \varphi^{-1}(y))$. Furthermore, $\varphi\colon (X, d) \to (Y, \rho)$ is an isometry. The metric $\rho$ is also known as $d \circ \varphi^{-1}$.*

*On the other hand, if $\varphi\colon Y \to (X, d)$, then $\varphi$ induces a semimetric $\rho$ on $Y$ by $\rho(x, y) = d(\varphi(x), \varphi(y))$. If $\varphi$ is one-to-one, then it is an isometry onto its range.*

The bounded uniformly continuous functions form a complete subspace of the space of bounded continuous functions.

**3.10 Lemma** *If $X$ is metrizable and $\rho$ is a compatible metric on $X$, then the vector space $U_\rho(X)$ of all bounded $\rho$-uniformly continuous real functions on $X$ is a closed subspace of $C_b(X)$. Thus $U_\rho(X)$ equipped with the uniform metric is a complete metric space in its own right.*[1]

The next theorem asserts that every uniformly continuous partial function can be uniquely extended to a uniformly continuous function on the closure of its domain simply by taking limits. The range space is assumed to be complete.

**3.11 Lemma (Uniformly continuous extensions)** *Let $A$ be a nonempty subset of $(X, d)$, and let $\varphi\colon (A, d) \to (Y, \rho)$ be uniformly continuous. Assume that $(Y, \rho)$ is complete. Then $\varphi$ has a unique uniformly continuous extension $\hat{\varphi}$ to the closure $\overline{A}$ of $A$. Moreover, the extension $\hat{\varphi}\colon \overline{A} \to Y$ is given by*

$$\hat{\varphi}(x) = \lim_{n\to\infty} \varphi(x_n)$$

*for any $\{x_n\} \subset A$ satisfying $x_n \to x$.*

*In particular, if $Y = \mathbb{R}$, then $\|\varphi\|_\infty = \|\hat{\varphi}\|_\infty$.*

[1] In the terminology of Section 9.5, $U_\rho(X)$ is a closed Riesz subspace of $C_b(X)$, and is also an AM-space with unit the constant function one.

*Proof*: Let $x \in \overline{A}$ and pick a sequence $\{x_n\}$ in $A$ converging to $x$. Since $\{x_n\}$ converges, it is $d$-Cauchy. Since $\varphi$ is uniformly continuous, $\{\varphi(x_n)\}$ is $\rho$-Cauchy. Since $Y$ is $\rho$-complete, there is some $y \in Y$ such that $\varphi(x_n) \to y$.

This $y$ is independent of the particular sequence $\{x_n\}$. To see this, let $\{z_n\}$ be another sequence in $A$ converging to $x$. Interlace the terms of $\{z_n\}$ and $\{x_n\}$ to form the sequence $\{z_1, x_1, z_2, x_2, \dots\}$ converging to $x$. Then $\{\varphi(z_1), \varphi(x_1), \varphi(z_2), \varphi(x_2), \dots\}$ is again $\rho$-Cauchy and since $\{\varphi(x_n)\}$ is a subsequence, the limit is again $y$. The latter implies that $\varphi(z_n) \to y$. Thus, setting $\hat{\varphi}(x) = y$ is well defined.

To see that $\hat{\varphi}$ is uniformly continuous on $\overline{A}$, let $\varepsilon > 0$ be given and pick $\delta > 0$ so that if $x, y \in A$ and $d(x, y) < \delta$, then $\rho(\varphi(x), \varphi(y)) < \varepsilon$. Now suppose $x, y \in \overline{A}$ and $d(x, y) < \delta$. Pick sequences $\{x_n\}$ and $\{y_n\}$ in $A$ converging to $x$ and $y$ respectively. From $|d(x_n, y_n) - d(x, y)| \leqslant d(x_n, x) + d(y_n, y)$, we see that $d(x_n, y_n) \to d(x, y)$, so eventually $d(x_n, y_n) < \delta$. Thus $\rho(\varphi(x_n), \varphi(y_n)) < \varepsilon$ eventually, so

$$\rho(\hat{\varphi}(x), \hat{\varphi}(y)) = \lim_{n\to\infty} \rho(\varphi(x_n), \varphi(y_n)) \leqslant \varepsilon.$$

The uniqueness of the extension is obvious. ∎

It is interesting to note that with an appropriate change of the metric of the domain of a continuous function between metric spaces the function becomes Lipschitz continuous.

**3.12 Lemma** *If $f\colon (X, d) \to (Y, \rho)$ is a continuous function between metric spaces, then there exists an equivalent metric $d_1$ on $X$ such that $f\colon (X, d_1) \to (Y, \rho)$ is Lipschitz (and hence uniformly) continuous.*

*More generally, if $\mathfrak{F}$ is a countable family of continuous functions from $(X, d)$ to $(Y, \rho)$, then there exists an equivalent metric $d_2$ on $X$ and an equivalent metric $\rho_1$ on $Y$ such that for each $f \in \mathfrak{F}$ the function $f\colon (X, d_2) \to (Y, \rho_1)$ is Lipschitz (and hence uniformly) continuous.*

*Proof*: The metric $d_1$ is defined by $d_1(x, y) = d(x, y) + \rho(f(x), f(y))$. The reader should verify that $d_1$ is indeed a metric on $X$ such that $d_1(x_n, x) \to 0$ holds in $X$ if and only if $d(x_n, x) \to 0$. This shows that the metric $d_1$ is equivalent to $d$. Now notice that the inequality $\rho(f(x), f(y)) \leqslant d_1(x, y)$ guarantees that the function $f\colon (X, d_1) \to (Y, \rho)$ is Lipschitz continuous.

The general case can be established in a similar manner. To see this, consider a countable set $\mathfrak{F} = \{f_1, f_2, \dots\}$ of continuous functions from $(X, d)$ to $(Y, \rho)$. Next, introduce the equivalent metric $\rho_1$ on $Y$ by $\rho_1(u, v) = \frac{\rho(u,v)}{1+\rho(u,v)}$. Subsequently, define the function $d_2\colon X \times X \to \mathbb{R}$ by

$$d_2(x, y) = d(x, y) + \sum_{n=1}^{\infty} \frac{1}{2^n} \rho_1(f_n(x), f_n(y)),$$

and note that $d_2$ is a metric on $X$ that is equivalent to $d$. In addition, for each $n$ we have the inequality $\rho_1(f_n(x), f_n(y)) \leqslant 2^n d_2(x, y)$. This shows that each function $f_n\colon (X, d_2) \to (Y, \rho_1)$ is Lipschitz continuous. ∎

## 3.4 Semicontinuous functions on metric spaces

On metric spaces, upper and lower semicontinuous functions are pointwise limits of monotone sequences of Lipschitz continuous functions.

**3.13 Theorem** *Let $f\colon (X,d) \to \mathbb{R}$ be bounded below. Then $f$ is lower semicontinuous if and only if it is the pointwise limit of an increasing sequence of Lipschitz continuous functions.*

*Similarly, if $g\colon (X,d) \to \mathbb{R}$ is bounded above, then $g$ is upper semicontinuous if and only if it is the pointwise limit of a decreasing sequence of Lipschitz continuous functions.*

*Proof*: We give a constructive proof of the first part. The second part follows from the first applied to $-f$. Let $f\colon X \to \mathbb{R}$ be lower semicontinuous and bounded from below. For each $n$, define $f_n\colon X \to \mathbb{R}$ by

$$f_n(x) = \inf\{f(y) + nd(x,y) : y \in X\}.$$

Clearly, $f_n(x) \leqslant f_{n+1}(x) \leqslant f(x)$ for each $x$. Moreover, observe that

$$|f_n(x) - f_n(z)| \leqslant nd(x,z),$$

which shows that each $f_n$ is Lipschitz continuous.

Let $f_n(x) \uparrow h(x) \leqslant f(x)$ for each $x$. Now fix $x$ and let $\varepsilon > 0$. For each $n$ pick some $y_n \in X$ with

$$f(y_n) \leqslant f(y_n) + nd(x,y_n) \leqslant f_n(x) + \varepsilon. \tag{$\star$}$$

If $f(u) \geqslant M > -\infty$ for all $u \in X$, then it follows from ($\star$) that

$$0 \leqslant d(x,y_n) \leqslant \frac{f_n(x) + \varepsilon - f(y_n)}{n} \leqslant \frac{f(x) + \varepsilon - M}{n}$$

for each $n$, and this shows that $y_n \to x$. Using the lower semicontinuity of $f$ and the inequality $f(y_n) \leqslant f_n(x) + \varepsilon$, we see that

$$f(x) \leqslant \liminf_{n\to\infty} f(y_n) \leqslant \lim_{n\to\infty} [f_n(x) + \varepsilon] = h(x) + \varepsilon$$

for each $\varepsilon > 0$. So $f(x) \leqslant h(x)$, and hence $f(x) = h(x) = \lim_{n\to\infty} f_n(x)$.

The converse follows immediately from Lemma 2.41. ∎

**3.14 Corollary** *Let $(X,d)$ be a metric space, and let $F$ be a closed subset of $X$. Then there is a sequence $\{f_n\}$ of Lipschitz continuous functions taking values in $[0,1]$ satisfying $f_n(x) \downarrow \chi_F(x)$ for all $x \in X$.*

*Proof*: Clearly $N_\varepsilon(A) \subset N_\varepsilon(\overline{A})$. For the reverse inclusion, let $y \in N_\varepsilon(\overline{A})$. Then there is some $x \in \overline{A}$ (so $d(x, A) = 0$) satisfying $d(x, y) < \varepsilon$. By equation ($\star$) in the proof of Theorem 3.16, we have $d(y, A) < \varepsilon$, or in other words $y \in N_\varepsilon(A)$. ∎

**3.19 Corollary** *In a metrizable space, every closed set is a $\mathcal{G}_\delta$, and every open set is an $\mathcal{F}_\sigma$.*

*Proof*: Let $F$ be a closed subset of $(X, d)$, and put $G_n = \{x \in X : d(x, F) < 1/n\}$. Since the distance function is continuous, $G_n$ is open, and clearly $F = \bigcap_{n=1}^{\infty} G_n$. Thus $F$ is a $\mathcal{G}_\delta$. Since the complement of an open set is closed, de Morgan's laws imply that every open set is an $\mathcal{F}_\sigma$. ∎

We can now show that a metric space is perfectly normal.

**3.20 Lemma** *If $(X, d)$ is a metric space and $A$ and $B$ are disjoint nonempty closed sets, then the continuous function $f\colon X \to [0, 1]$, defined by*

$$f(x) = \frac{d(x, A)}{d(x, A) + d(x, B)},$$

*satisfies $f^{-1}(0) = A$ and $f^{-1}(1) = B$.*

*Moreover, if* $\inf\{d(x, y) : x \in A \text{ and } y \in B\} > 0$*, then the function $f$ is Lipschitz continuous, and hence $d$-uniformly continuous.*

*Proof*: The first assertion is obvious. For the second, assume that there exists some $\delta > 0$ such that $d(x, y) \geqslant \delta$ for all $x \in A$ and all $y \in B$. Then, for any $z \in X$, $a \in A$, and $b \in B$, $\delta \leqslant d(a, b) \leqslant d(a, z) + d(z, b)$, so $d(z, A) + d(z, B) \geqslant \delta > 0$ for each $z \in X$. Now use the inequalities

$$\begin{aligned}
|f(x) - f(y)| &= \left| \frac{d(x, A)}{d(x, A) + d(x, B)} - \frac{d(y, A)}{d(y, A) + d(y, B)} \right| \\
&= \frac{|[d(y, A) + d(y, B)]d(x, A) - [d(x, A) + d(x, B)]d(y, A)|}{[d(x, A) + d(x, B)][d(y, A) + d(y, B)]} \\
&= \frac{|[d(x, A) - d(y, A)]d(x, B) + [d(y, B) - d(x, B)]d(x, A)|}{[d(x, A) + d(x, B)][d(y, A) + d(y, B)]} \\
&\leqslant \frac{[d(x, B) + d(x, A)]d(x, y)}{[d(x, A) + d(x, B)][d(y, A) + d(y, B)]} \leqslant \frac{d(x, y)}{\delta},
\end{aligned}$$

to see that $f$ is indeed Lipschitz continuous. ∎

**3.21 Corollary** *Every metrizable space is perfectly normal.*

Using distance functions we can establish the following useful result.

**3.22 Theorem** *Every metrizable space is paracompact.*

Now let $j > i$, $x \in U_j^n$ and $y \in U_i^n$. Pick $u \in S_j^n$ and $v \in S_i^n$ so that $d(x,u) < 1/2^{n+2}$ and $d(y,v) < 1/2^{n+2}$ and note that from (5) we get

$$\frac{1}{2^n} \leqslant d(u,v) \leqslant d(u,x) + d(x,y) + d(y,v) < d(x,y) + \frac{1}{2^{n+1}}.$$

This implies:

$$\text{If } i \neq j,\ x \in U_j^n \text{ and } y \in U_i^n, \text{ then } d(x,y) > 1/2^{n+1}. \tag{6}$$

Next, for each fixed $n$ consider the family of closed sets $\{C_i^n\}_{i \in I}$. We claim that for each $x \in X$ the open ball $B = B_{\frac{1}{2^{n+2}}}(x)$ intersects at most one of the sets $\{C_i^n\}_{i \in I}$. To see this, assume that for $i \neq j$ we have $y \in B \cap C_i^n$ and $z \in B \cap C_j^n$. Now a glance at (6) yields

$$\frac{1}{2^{n+1}} < d(y,z) \leqslant d(y,x) + d(x,z) < \frac{1}{2^{n+2}} + \frac{1}{2^{n+2}} = \frac{1}{2^{n+1}},$$

a contradiction. This implies (how?) that for each $n$ the set $C_n = \bigcup_{i \in I} C_i^n$ is closed.

Finally, for each $n$ and $i \in I$ define the sets:

$$W_i^1 = U_i^1 \quad \text{and} \quad W_i^n = U_i^n \setminus \bigcup_{k=1}^{n-1} C_k \quad \text{if } n > 1.$$

Clearly, each $W_i^n$ is an open set. We claim that the family of open sets $\{W_i^n\}_{(n,i) \in \mathbb{N} \times I}$ is an open locally finite refinement cover of $\{V_i\}_{i \in I}$. We establish this claim by steps.

Step I: *$\{W_i^n\}_{(n,i) \in \mathbb{N} \times I}$ is a refinement of $\{V_i\}_{i \in I}$.*

To see this, note that $W_i^n \subset U_i^n = N_{n+2}(S_i^n) \subset N_n(S_i^n) \subset N_n(E_n(V_i)) \subset V_i$.

Step II: *$\{W_i^n\}_{(n,i) \in \mathbb{N} \times I}$ covers $X$, that is, $X = \bigcup_{n=1}^{\infty} \bigcup_{i \in I} W_i^n$.*

Fix $x \in X$. From $S_i^n \subset C_i^n$ and (4), we see that the family $\{C_i^n\}_{(n,i) \in \mathbb{N} \times I}$ covers $X$. Put $k = \min\{n \in \mathbb{N} : x \in C_i^n \text{ for some } i\}$. Assume that $x \notin W_i^1$. If $x \in C_i^k \subset U_i^k$, then $k > 1$ and $x \notin C_n$ for each $n < k$. Hence $x \in W_i^k$.

Step III: *$\{W_i^n\}_{(n,i) \in \mathbb{N} \times I}$ is locally finite.*

Fix $x \in X$. According to (4) there exists some $n$ and $i_0 \in I$ such that $x \in S_{i_0}^n$. Now note that

$$B_{1/2^{n+3}}(x) \subset N_{n+3}(S_{i_0}^n) \subset \overline{N_{n+3}(S_{i_0}^n)} = C_{i_0}^n \subset C_n.$$

This implies $B_{1/2^{n+3}}(x) \cap W_i^k = \varnothing$ for all $k > n$ and all $i \in I$.

Next, fix $1 \leqslant k \leqslant n$ and assume that $B_{1/2^{n+3}}(x) \cap U_i^k \neq \varnothing$ for some $i \in I$. Then $B_{1/2^{n+3}}(x) \cap U_i^k = \varnothing$ for all $j \neq i$. To see this, assume that for $i \neq j$ there exist $y \in B_{1/2^{n+3}}(x) \cap U_i^k$ and $z \in B_{1/2^{n+3}}(x) \cap U_j^k$. But then from (6) we get $1/2^{n+1} \leqslant d(y,z) \leqslant d(y,x) + d(x,z) < 1/2^{n+3}$, which is impossible. This shows that $B_{1/2^{n+3}}(x)$ intersects at most $n$ of the $\{U_i^k : 1 \leqslant k \leqslant n \text{ and } i \in I\}$. It follows that $B_{1/2^{n+3}}(x)$ intersects at most $n$ of the sets $W_i^k$. ∎

## 3.6 Embeddings and completions

An **isometric embedding** of the metric space $(X, d)$ in the metric space $(Y, \rho)$ is simply an isometry $f: X \to Y$.

**3.23 Embedding Lemma** *Every metric space can be isometrically embedded in its space of bounded uniformly continuous real functions.*

*Proof*: Let $(X, d)$ be a metric space. Fix an arbitrary point $a \in X$ as a reference, and for each $x$ define the function $\theta_x$ by

$$\theta_x(y) = d(x, y) - d(a, y).$$

For the uniform continuity of $\theta_x$ note that

$$|\theta_x(y) - \theta_x(z)| \leqslant |d(x, y) - d(x, z)| + |d(a, y) - d(a, z)| \leqslant 2d(y, z).$$

To see that $\theta_x$ is bounded, use the inequality $d(x, y) \leqslant d(x, a) + d(a, y)$ and the definition of the function $\theta_x$ to see that $\theta_x(y) \leqslant d(x, a)$. Likewise the inequality $d(a, y) \leqslant d(a, x) + d(x, y)$ implies $-\theta_x(y) = d(a, y) - d(x, y) \leqslant d(x, a)$. Furthermore, these inequalities hold exactly for $y = a$ and $y = x$ respectively. Consequently we have $\|\theta_x\|_\infty = \sup_y |\theta_x(y)| = d(x, a)$.

Next, observe that

$$\begin{aligned} |\theta_x(y) - \theta_z(y)| &= |d(x, y) - d(a, y) - [d(z, y) - d(a, y)]| \\ &= |d(x, y) - d(z, y)| \leqslant d(x, z) \end{aligned}$$

for all $y \in X$. Also $|\theta_x(z) - \theta_z(z)| = d(x, z)$. Thus,

$$\|\theta_x - \theta_z\|_\infty = \sup_{y \in X} |\theta_x(y) - \theta_z(y)| = d(x, z)$$

for all $x, z \in X$. That is, $\theta$ is an isometry. ∎

Note that for the special case when $d$ is a bounded metric on $X$, the mapping $x \mapsto d(x, \cdot)$ is an isometry from $X$ into $C_b(X)$.

A complete metric space $(Y, \rho)$ is the **completion** of the metric space $(X, d)$ if there exists an isometry $\varphi: (X, d) \to (Y, \rho)$ satisfying $\overline{\varphi(X)} = Y$. It is customary to identify $X$ with $\varphi(X)$ and consider $X$ to be a dense subset of $Y$. The next result justifies calling $Y$ *the* completion of $X$ rather than *a* completion of $X$.

**3.24 Theorem** *Every metric space has a completion. It is unique up to isometry, that is, any two completions are isometric.*

*Proof*: Since $C_b(X)$ is a complete metric space in the metric induced by its norm, Lemma 3.23 shows that a completion exists, namely $\overline{\theta(X)}$.

To prove the uniqueness of the completion up to isometry, let both $(Y_1, \rho_1)$ and $(Y_2, \rho_2)$ be completions of $(X, d)$ with isometries $\varphi_i \colon (X, d) \to (Y_i, \rho_i)$. Then the function $\varphi = \varphi_1 \circ \varphi_2^{-1} \colon (\varphi_2(X), \rho_2) \to (\varphi_1(X), \rho_1)$ is an isometry and hence is uniformly continuous. By Lemma 3.11, $\varphi$ has a uniformly continuous extension $\hat{\varphi}$ to the closure $Y_2$ of $\varphi_2(X)$. Routine arguments show that $\hat{\varphi} \colon (Y_2, \rho_2) \to (Y_1, \rho_1)$ is a surjective isometry. That is, $(Y_2, \rho_2)$ and $(Y_1, \rho_1)$ are isometric. ∎

**3.25 Theorem** *The completion of a separable metric space is separable.*

*Proof*: Let $Y$ be the completion of a metric space $X$ and let $\varphi \colon X \to Y$ be an isometry such that $\overline{\varphi(X)} = Y$. If $A$ is a countable dense subset of $X$, then (in view of Theorem 2.27 (5)) the countable subset $\varphi(A)$ of $Y$ satisfies $\varphi(X) = \varphi(\overline{A}) \subset \overline{\varphi(A)}$, so $Y = \overline{\varphi(X)} = \overline{\varphi(A)}$. ∎

## 3.7 Compactness and completeness

A subset $A$ of a metric space $X$ is **totally bounded** if for each $\varepsilon > 0$ there exists a finite subset $\{x_1, \ldots, x_n\} \subset X$ that is **$\varepsilon$-dense** in $A$, meaning that the collection of $\varepsilon$-balls $B_\varepsilon(x_i)$ covers $A$. Note that if a set is totally bounded, then so are its closure and any subset. Any metric for which the space $X$ is totally bounded is also called a **totally bounded metric**.

Every compact metric space is obviously totally bounded. It is easy to see that a totally bounded metric space is separable.

**3.26 Lemma** *Every totally bounded metric space is separable.*

*Proof*: If $(X, d)$ is totally bounded, then for each $n$ pick a finite subset $F_n$ of $X$ such that $X = \bigcup_{x \in F_n} B_{1/n}(x)$, and then note that the set $F = \bigcup_{n=1}^{\infty} F_n$ is countable and dense. ∎

This implies that every compact metric space is separable, but that is not necessarily true of nonmetrizable compact topological spaces. (Can you think of a nonseparable compact topological space?) For the next result, recall that a topological space is sequentially compact if every sequence has a convergent subsequence.

**3.27 Lemma** *Let $(X, d)$ be a sequentially compact metric space, and let $\{V_i\}_{i \in I}$ be an open cover of $X$. Then there exists some $\delta > 0$, called the **Lebesgue number** of the cover, such that for each $x \in X$ we have $B_\delta(x) \subset V_i$ for at least one $i$.*

*Proof*: Assume by way of contradiction that no such $\delta$ exists. Then for each $n$ there exists some $x_n \in X$ satisfying $B_{1/n}(x_n) \cap V_i^c \neq \varnothing$ for each $i \in I$. If $x$ is the limit point of some subsequence of $\{x_n\}$, then it is easy to see (how?) that $x \in \bigcap_{i\in I} V_i^c = \left(\bigcup_{i\in I} V_i\right)^c = \varnothing$, a contradiction. ∎

The next two results sharpen the relationship between compactness and total boundedness.

**3.28 Theorem (Compactness of metric spaces)** *For a metric space the following are equivalent:*

1. *The space is compact.*
2. *The space is complete and totally bounded.*
3. *The space is sequentially compact. That is, every sequence has a convergent subsequence.*

*Proof*: Let $(X, d)$ be a metric space.

(1) $\Longrightarrow$ (2) Since $X = \bigcup_{x\in X} B_\varepsilon(x)$, there exist $x_1, \ldots, x_k$ in $X$ such that $X = \bigcup_{i=1}^k B_\varepsilon(x_i)$. That is, $X$ is totally bounded. To see that $X$ is also complete, let $\{x_n\}$ be a Cauchy sequence in $X$, and let $\varepsilon > 0$ be given. Pick $n_0$ so that $d(x_n, x_m) < \varepsilon$ whenever $n, m \geqslant n_0$. By Theorem 2.31, the sequence $\{x_n\}$ has a limit point, say $x$. We claim that $x_n \to x$. Indeed, if we choose $k \geqslant n_0$ such that $d(x_k, x) < \varepsilon$, then for each $n \geqslant n_0$, we have

$$d(x_n, x) \leqslant d(x_n, x_k) + d(x_k, x) < \varepsilon + \varepsilon = 2\varepsilon,$$

proving $x_n \to x$. That is, $X$ is also complete.

(2) $\Longrightarrow$ (3) Fix a sequence $\{x_n\}$ in $X$. Since $X$ is totally bounded, there must be infinitely many terms of the sequence in a closed ball of radius $1/2$. (Why?) This ball is totally bounded too, so it must also include a closed set of diameter less than $\frac{1}{4}$ that contains infinitely many terms of the sequence. By induction, construct a decreasing sequence of closed sets with vanishing diameter, each of which contains infinitely many terms of the sequence. Use this and the Cantor Intersection Theorem 3.7 to construct a convergent subsequence.

(3) $\Longrightarrow$ (1) By Lemma 3.27, there is some $\delta > 0$ such that for each $x \in X$ we have $B_\delta(x) \subset V_i$ for at least one $i$. We claim that there exist $x_1, \ldots, x_k \in X$ such that $X = \bigcup_{i=1}^k B_\delta(x_i)$. To see this, assume by way of contradiction that this is not the case. Fix $y_1 \in X$. Since the claim is false, there exists some $y_2 \in X$ such that $d(y_1, y_2) \geqslant \delta$. Similarly, since $X \neq B_\delta(y_1) \cup B_\delta(y_2)$, there exists some $y_3 \in X$ such that $d(y_1, y_3) \geqslant \delta$ and $d(y_2, y_3) \geqslant \delta$. So by an inductive argument, there exists a sequence $\{y_n\}$ in $X$ satisfying $d(y_n, y_m) \geqslant \delta$ for $n \neq m$. However, any such sequence $\{y_n\}$ cannot have any convergent subsequence, contrary to our hypothesis. Hence there exist $x_1, \ldots, x_k \in X$ such that $X = \bigcup_{i=1}^k B_\delta(x_i)$.

The next three results deal with subsets of metric spaces that are completely metrizable given their induced topologies.

**3.33 Lemma** *If the relative topology of a subset of a metric space is completely metrizable, then the subset is a* $\mathcal{G}_\delta$.

*Proof*: Let $X$ be a subset of a metric space $(Y, d)$ such that $X$ admits a metric $\rho$ that is consistent with the relative topology on $X$ and for which $(X, \rho)$ is complete.

Heuristically, $X$ is $\bigcap_{n=1}^{\infty}\{y \in Y : d(y, X) < 1/n\} \cap \{y \in Y : \rho(y, X) < 1/n\}$. But this makes no sense, since $\rho(y, x)$ is not defined for $y \in Y \setminus X$. So what we need is a way to include points in $Y$ that would be both $d$-close and $\rho$-close to $X$ if $\rho$ were defined on $Y$. Recall that any open set $U$ in $X$ is the intersection of $X$ with an open subset $V$ of $Y$. The idea is to consider open sets $V$ where $V \cap X$ is $\rho$-small. To this end, for each $n$ let

$$Y_n = \{y \in Y : \text{there is an open set } V \text{ in } Y \text{ with } y \in V \text{ and } \rho\text{-diam}\,(X \cap V) < 1/n\},$$

and put

$$G_n = \{y \in Y : d(y, X) < 1/n\} \bigcap Y_n.$$

First, we claim that each $G_n$ is an open subset of $Y$. Indeed, if $y \in G_n$, then pick the open subset $V$ of $Y$ with $y \in V$ and $\rho$-diam $(X \cap V) < \frac{1}{n}$ and note that the open neighborhood $W = V \cap \{z \in Y : d(z, X) < \frac{1}{n}\}$ of $y$ in $Y$ satisfies $W \subset G_n$. To complete the proof, we shall show that $X = \bigcap_{n=1}^{\infty} G_n$.

First let $x$ belong to $X$ and fix $n$. Then $U = \{y \in X : \rho(y, x) < 1/3n\}$ is an open subset of $X$. So there exists an open subset $V$ of $Y$ with $U = X \cap V$. It follows that $x \in V$ and $\rho$-diam $(X \cap V) < 1/n$, so $x \in G_n$. Since $n$ is arbitrary, $X \subset \bigcap_{n=1}^{\infty} G_n$.

For the reverse inclusion, let $y \in \bigcap_{n=1}^{\infty} G_n$. Then $d(y, X) = 0$, so $y \in \overline{X}$. In particular, there exists a sequence $\{x_n\}$ in $X$ such that $x_n \to y$. For each $n$ pick an open subset $V_n$ of $Y$ with $y \in V_n$ and $\rho$-diam $(X \cap V_n) < 1/n$. Since $X \cap V_n$ is an open subset of $X$, it follows that for each $n$ there exists some $k_n$ such that $x_m \in V_n$ for all $m \geqslant k_n$. From $\rho$-diam $(X \cap V_n) < 1/n$, we see that $\{x_n\}$ is a $\rho$-Cauchy sequence, and since $(X, \rho)$ is complete, $\{x_n\}$ is $\rho$-convergent to some $z \in X$. It follows that $y = z \in X$, so $X = \bigcap_{n=1}^{\infty} G_n$, as desired. ∎

For complete metric spaces the converse of Lemma 3.33 is also true.

**3.34 Alexandroff's Lemma** *Every* $\mathcal{G}_\delta$ *in a complete metric space is completely metrizable.*

*Proof*: Let $(Y, d)$ be a complete metric space, and assume that $X \neq Y$ is a $\mathcal{G}_\delta$. (The case $X = Y$ is trivial.) Then there exists a sequence $\{G_n\}$ of open sets satisfying $G_n \neq Y$ for each $n$ and $X = \bigcap_{n=1}^{\infty} G_n$. (We want $G_n \neq Y$ so that $G_n^c = Y \setminus G_n$ is nonempty, so $0 < d(x, G_n^c) < \infty$ for all $x \in X$.) Next, define the metric $\rho$ on $X$ by

$$\rho(x, y) = d(x, y) + \sum_{n=1}^{\infty} \min\left\{\frac{1}{2^n}, \left|\frac{1}{d(x, G_n^c)} - \frac{1}{d(y, G_n^c)}\right|\right\}.$$

ALIPRANTIS · BORDER

Infinite Dimensional Analysis ▪ 3rd Edition

This monograph presents a complete and rigorous study of modern functional analysis. It is intended for the student or researcher who could benefit from functional analytic methods, but does not have an extensive background and does not plan to make a career as a functional analyst. It develops the topological structures in connection with measure theory, convexity, Banach lattices, integration, correspondences (multifunctions), and the analytic approach to Markov processes. Many of the results were previously available only in works scattered throughout the literature. The choice of material was motivated from problems in control theory and economics, although the material is more applicable than applied.

ISBN 3-540-32696-0

9 783540 326960

› springer.com