Real Analysis

Modern Techniques and Their Applications

Second Edition

Gerald B. Folland

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

This book is printed on acid-free paper. ♾

Library of Congress Cataloging-in-Publication Data:

Folland, Gerald B.

Real analysis : modern techniques and their applications / Gerald B. Folland. — 2nd ed.

p. cm. — (Pure and applied mathematics)

"A Wiley-Interscience publication."

Includes bibliographical references and index.

ISBN 0-471-31716-0 (cloth : alk. paper)

1. Mathematical analysis. 2. Functions of real variables. I. Title. II. Series: Pure and applied mathematics (John Wiley & Sons : Unnumbered)

QA300.F67 1999

515—dc21 98-37260

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Contents

0

Prologue

The purpose of this introductory chapter is to establish the notation and terminology that will be used throughout the book and to present a few diverse results from set theory and analysis that will be needed later. The style here is deliberately terse, since this chapter is intended as a reference rather than a systematic exposition.

0.1 THE LANGUAGE OF SET THEORY

It is assumed that the reader is familiar with the basic concepts of set theory; the following discussion is meant mainly to fix our terminology.

Number Systems. Our notation for the fundamental number systems is as follows:

$\mathbb{N}$ = the set of positive integers (not including zero)

$\mathbb{Z}$ = the set of integers

$\mathbb{Q}$ = the set of rational numbers

$\mathbb{R}$ = the set of real numbers

$\mathbb{C}$ = the set of complex numbers

Logic. We shall avoid the use of special symbols from mathematical logic, preferring to remain reasonably close to standard English. We shall, however, use the abbreviation **iff** for "if and only if."

One point of elementary logic that is often insufficiently appreciated by students is the following: If A and B are mathematical assertions and $-A$, $-B$ are their negations, the statement "A implies B" is logically equivalent to the contrapositive statement "$-B$ implies $-A$." Thus one may prove that A implies B by assuming $-B$ and deducing $-A$, and we shall frequently do so. This is not the same as *reductio ad absurdum,* which consists of assuming both A and $-B$ and deriving a contradiction.

Sets. The words "family" and "collection" will be used synonymously with "set," usually to avoid phrases like "set of sets." The empty set is denoted by $\varnothing$, and the family of all subsets of a set X is denoted by $\mathcal{P}(X)$:

$$\mathcal{P}(X) = \{E : E \subset X\}.$$

Here and elsewhere, the inclusion sign $\subset$ is interpreted in the weak sense; that is, the assertion "$E \subset X$" includes the possibility that $E = X$.

If $\mathcal{E}$ is a family of sets, we can form the union and intersection of its members:

$$\bigcup_{E\in\mathcal{E}} E = \{x : x \in E \text{ for some } E \in \mathcal{E}\},$$

$$\bigcap_{E\in\mathcal{E}} E = \{x : x \in E \text{ for all } E \in \mathcal{E}\}.$$

Usually it is more convenient to consider indexed families of sets:

$$\mathcal{E} = \{E_\alpha : \alpha \in A\} = \{E_\alpha\}_{\alpha\in A},$$

in which case the union and intersection are denoted by

$$\bigcup_{\alpha\in A} E_\alpha, \qquad \bigcap_{\alpha\in A} E_\alpha.$$

If $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$, the sets E_α are called **disjoint**. The terms "disjoint collection of sets" and "collection of disjoint sets" are used interchangeably, as are "disjoint union of sets" and "union of disjoint sets."

When considering families of sets indexed by $\mathbb{N}$, our usual notation will be

$$\{E_n\}_{n=1}^\infty \quad \text{or} \quad \{E_n\}_1^\infty,$$

and likewise for unions and intersections. In this situation, the notions of **limit superior** and **limit inferior** are sometimes useful:

$$\limsup E_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty E_n, \qquad \liminf E_n = \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty E_n.$$

The reader may verify that

$$\limsup E_n = \{x : x \in E_n \text{ for infinitely many } n\},$$
$$\liminf E_n = \{x : x \in E_n \text{ for all but finitely many } n\}.$$

If E and F are sets, we denote their **difference** by $E \setminus F$:

$$E \setminus F = \{x : x \in E \text{ and } x \notin F\},$$

and their **symmetric difference** by $E \triangle F$:

$$E \triangle F = (E \setminus F) \cup (F \setminus E).$$

When it is clearly understood that all sets in question are subsets of a fixed set X, we define the **complement** E^c of a set E (in X):

$$E^c = X \setminus E.$$

In this situation we have **deMorgan's laws**:

$$\Big(\bigcup_{\alpha \in A} E_\alpha\Big)^c = \bigcap_{\alpha \in A} E_\alpha^c, \qquad \Big(\bigcap_{\alpha \in A} E_\alpha\Big)^c = \bigcup_{\alpha \in A} E_\alpha^c.$$

If X and Y are sets, their **Cartesian product** $X \times Y$ is the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. A **relation** from X to Y is a subset of $X \times Y$. (If $Y = X$, we speak of a relation **on** X.) If R is a relation from X to Y, we shall sometimes write xRy to mean that $(x, y) \in R$. The most important types of relations are the following:

- Equivalence relations. An **equivalence relation** on X is a relation R on X such that

  $$\begin{aligned} &xRx \text{ for all } x \in X,\\ &xRy \text{ iff } yRx,\\ &xRz \text{ whenever } xRy \text{ and } yRz \text{ for some } y. \end{aligned}$$

  The **equivalence class** of an element x is $\{y \in X : xRy\}$. X is the disjoint union of these equivalence classes.

- Orderings. See §0.2.

- Mappings. A **mapping** $f : X \to Y$ is a relation R from X to Y with the property that for every $x \in X$ there is a unique $y \in Y$ such that xRy, in which case we write $y = f(x)$. Mappings are sometimes called **maps** or **functions**; we shall generally reserve the latter name for the case when Y is $\mathbb{C}$ or some subset thereof.

If $f : X \to Y$ and $g : Y \to Z$ are mappings, we denote by $g \circ f$ their **composition**:

$$g \circ f : X \to Z, \qquad g \circ f(x) = g(f(x)).$$

If $D \subset X$ and $E \subset Y$, we define the **image** of D and the **inverse image** of E under a mapping $f : X \to Y$ by

$$f(D) = \{f(x) : x \in D\}, \qquad f^{-1}(E) = \{x : f(x) \in E\}.$$

It is easily verified that the map $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$ defined by the second formula commutes with union, intersections, and complements:

$$f^{-1}\Big(\bigcup_{\alpha\in A} E_\alpha\Big) = \bigcup_{\alpha\in A} f^{-1}(E_\alpha), \qquad f^{-1}\Big(\bigcap_{\alpha\in A} E_\alpha\Big) = \bigcap_{\alpha\in A} f^{-1}(E_\alpha),$$
$$f^{-1}(E^c) = \big(f^{-1}(E)\big)^c.$$

(The direct image mapping $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ commutes with unions, but in general not with intersections or complements.)

If $f : X \to Y$ is a mapping, X is called the **domain** of f and $f(X)$ is called the **range** of f. f is said to be **injective** if $f(x_1) = f(x_2)$ only when $x_1 = x_2$, **surjective** if $f(X) = Y$, and **bijective** if it is both injective and surjective. If f is bijective, it has an **inverse** $f^{-1} : Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity mappings on X and Y, respectively. If $A \subset X$, we denote by $f|A$ the restriction of f to A:

$$(f|A) : A \to Y, \qquad (f|A)(x) = f(x) \text{ for } x \in A.$$

A **sequence** in a set X is a mapping from $\mathbb{N}$ into X. (We also use the term **finite sequence** to mean a map from $\{1, \ldots, n\}$ into X where $n \in \mathbb{N}$.) If $f : \mathbb{N} \to X$ is a sequence and $g : \mathbb{N} \to \mathbb{N}$ satisfies $g(n) < g(m)$ whenever $n < m$, the composition $f \circ g$ is called a **subsequence** of f. It is common, and often convenient, to be careless about distinguishing between sequences and their ranges, which are subsets of X indexed by $\mathbb{N}$. Thus, if $f(n) = x_n$, we speak of the sequence $\{x_n\}_1^\infty$; whether we mean a mapping from $\mathbb{N}$ to X or a subset of X will be clear from the context.

Earlier we defined the Cartesian product of two sets. Similarly one can define the Cartesian product of n sets in terms of ordered n-tuples. However, this definition becomes awkward for infinite families of sets, so the following approach is used instead. If $\{X_\alpha\}_{\alpha\in A}$ is an indexed family of sets, their **Cartesian product** $\prod_{\alpha\in A} X_\alpha$ is the set of all maps $f : A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for every $\alpha \in A$. (It should be noted, and then promptly forgotten, that when $A = \{1, 2\}$, the previous definition of $X_1 \times X_2$ is set-theoretically different from the present definition of $\prod_1^2 X_j$. Indeed, the latter concept depends on mappings, which are defined in terms of the former one.) If $X = \prod_{\alpha\in A} X_\alpha$ and $\alpha \in A$, we define the αth **projection** or **coordinate map** $\pi_\alpha : X \to X_\alpha$ by $\pi_\alpha(f) = f(\alpha)$. We also frequently write x and x_α instead of f and $f(\alpha)$ and call x_α the αth **coordinate** of x.

If the sets X_α are all equal to some fixed set Y, we denote $\prod_{\alpha\in A} X_\alpha$ by Y^A:

$$Y^A = \text{ the set of all mappings from } A \text{ to } Y.$$

If $A = \{1, \ldots, n\}$, Y^A is denoted by Y^n and may be identified with the set of ordered n-tuples of elements of Y.

0.2 ORDERINGS

A **partial ordering** on a nonempty set X is a relation R on X with the following properties:

- if xRy and yRz, then xRz;
- if xRy and yRx, then $x = y$;
- xRx for all x.

If R also satisfies

- if $x, y \in X$, then either xRy or yRx,

then R is called a **linear** (or **total**) ordering. For example, if E is any set, then $\mathcal{P}(E)$ is partially ordered by inclusion, and $\mathbb{R}$ is linearly ordered by its usual ordering. Taking this last example as a model, we shall usually denote partial orderings by $\leq$, and we write $x < y$ to mean that $x \leq y$ but $x \neq y$. We observe that a partial ordering on X naturally induces a partial ordering on every nonempty subset of X. Two partially ordered sets X and Y are said to be **order isomorphic** if there is a bijection $f : X \to Y$ such that $x_1 \leq x_2$ iff $f(x_1) \leq f(x_2)$.

If X is partially ordered by $\leq$, a **maximal** (resp. **minimal**) **element** of X is an element $x \in X$ such that the only $y \in X$ satisfying $x \leq y$ (resp. $x \geq y$) is x itself. Maximal and minimal elements may or may not exist, and they need not be unique unless the ordering is linear. If $E \subset X$, an **upper** (resp. **lower**) **bound** for E is an element $x \in X$ such that $y \leq x$ (resp. $x \leq y$) for all $y \in E$. An upper bound for E need not be an element of E, and unless E is linearly ordered, a maximal element of E need not be an upper bound for E. (The reader should think up some examples.)

If X is linearly ordered by $\leq$ and every nonempty subset of X has a (necessarily unique) minimal element, X is said to be **well ordered** by $\leq$, and (in defiance of the laws of grammar) $\leq$ is called a **well ordering** on X. For example, $\mathbb{N}$ is well ordered by its natural ordering.

We now state a fundamental principle of set theory and derive some consequences of it.

0.1 The Hausdorff Maximal Principle. *Every partially ordered set has a maximal linearly ordered subset.*

In more detail, this means that if X is partially ordered by $\leq$, there is a set $E \subset X$ that is linearly ordered by $\leq$, such that no subset of X that properly includes E is linearly ordered by $\leq$. Another version of this principle is the following:

0.2 Zorn's Lemma. *If X is a partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.*

Clearly the Hausdorff maximal principle implies Zorn's lemma: An upper bound for a maximal linearly ordered subset of X is a maximal element of X. It is also not difficult to see that Zorn's lemma implies the Hausdorff maximal principle. (Apply Zorn's lemma to the collection of linearly ordered subsets of X, which is partially ordered by inclusion.)

0.3 The Well Ordering Principle. *Every nonempty set X can be well ordered.*

Proof. Let $\mathcal{W}$ be the collection of well orderings of subsets of X, and define a partial ordering on $\mathcal{W}$ as follows. If $\leq_1$ and $\leq_2$ are well orderings on the subsets E_1 and E_2, then $\leq_1$ precedes $\leq_2$ in the partial ordering if (i) $\leq_2$ extends $\leq_1$, i.e., $E_1 \subset E_2$ and $\leq_1$ and $\leq_2$ agree on E_1, and (ii) if $x \in E_2 \setminus E_1$ then $y \leq_2 x$ for all $y \in E_1$. The reader may verify that the hypotheses of Zorn's lemma are satisfied, so that $\mathcal{W}$ has a maximal element. This must be a well ordering on X itself, for if $\leq$ is a well ordering on a proper subset E of X and $x_0 \in X \setminus E$, then $\leq$ can be extended to a well ordering on $E \cup \{x_0\}$ by declaring that $x \leq x_0$ for all $x \in E$. ∎

0.4 The Axiom of Choice. *If $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets, then $\prod_{\alpha \in A} X_\alpha$ is nonempty.*

Proof. Let $X = \bigcup_{\alpha \in A} X_\alpha$. Pick a well ordering on X and, for $\alpha \in A$, let $f(\alpha)$ be the minimal element of X_α. Then $f \in \prod_{\alpha \in A} X_\alpha$. ∎

0.5 Corollary. *If $\{X_\alpha\}_{\alpha \in A}$ is a disjoint collection of nonempty sets, there is a set $Y \subset \bigcup_{\alpha \in A} X_\alpha$ such that $Y \cap X_\alpha$ contains precisely one element for each $\alpha \in A$.*

Proof. Take $Y = f(A)$ where $f \in \prod_{\alpha \in A} X_\alpha$. ∎

We have deduced the axiom of choice from the Hausdorff maximal principle; in fact, it can be shown that the two are logically equivalent.

0.3 CARDINALITY

If X and Y are nonempty sets, we define the expressions

$$\text{card}(X) \leq \text{card}(Y), \qquad \text{card}(X) = \text{card}(Y), \qquad \text{card}(X) \geq \text{card}(Y)$$

to mean that there exists $f : X \to Y$ which is injective, bijective, or surjective, respectively. We also define

$$\text{card}(X) < \text{card}(Y), \qquad \text{card}(X) > \text{card}(Y)$$

to mean that there is an injection but no bijection, or a surjection but no bijection, from X to Y. Observe that we attach no meaning to the expression "card(X)" when it stands alone; there are various ways of doing so, but they are irrelevant for our purposes (except when X is finite — see below). These relationships can be extended to the empty set by declaring that

$$\text{card}(\varnothing) < \text{card}(X) \text{ and } \text{card}(X) > \text{card}(\varnothing) \text{ for all } X \neq \varnothing.$$

For the remainder of this section we assume implicitly that all sets in question are nonempty in order to avoid special arguments for $\varnothing$. Our first task is to prove that the relationships defined above enjoy the properties that the notation suggests.

0.6 Proposition. $\operatorname{card}(X) \leq \operatorname{card}(Y)$ *iff* $\operatorname{card}(Y) \geq \operatorname{card}(X)$.

Proof. If $f : X \to Y$ is injective, pick $x_0 \in X$ and define $g : Y \to X$ by $g(y) = f^{-1}(y)$ if $y \in f(X)$, $g(y) = x_0$ otherwise. Then g is surjective. Conversely, if $g : Y \to X$ is surjective, the sets $g^{-1}(\{x\})$ $(x \in X)$ are nonempty and disjoint, so any $f \in \prod_{x \in X} g^{-1}(\{x\})$ is an injection from X to Y. ∎

0.7 Proposition. *For any sets X and Y, either* $\operatorname{card}(X) \leq \operatorname{card}(Y)$ *or* $\operatorname{card}(Y) \leq \operatorname{card}(X)$.

Proof. Consider the set $\mathcal{I}$ of all injections from subsets of X to Y. The members of $\mathcal{I}$ can be regarded as subsets of $X \times Y$, so $\mathcal{I}$ is partially ordered by inclusion. It is easily verified that Zorn's lemma applies, so $\mathcal{I}$ has a maximal element f, with (say) domain A and range B. If $x_0 \in X \setminus A$ and $y_0 \in Y \setminus B$, then f can be extended to an injection from $A \cup \{x_0\}$ to $Y \cup \{y_0\}$ by setting $f(x_0) = y_0$, contradicting maximality. Hence either $A = X$, in which case $\operatorname{card}(X) \leq \operatorname{card}(Y)$, or $B = Y$, in which case f^{-1} is an injection from Y to X and $\operatorname{card}(Y) \leq \operatorname{card}(X)$. ∎

0.8 The Schröder-Bernstein Theorem. *If* $\operatorname{card}(X) \leq \operatorname{card}(Y)$ *and* $\operatorname{card}(Y) \leq \operatorname{card}(X)$ *then* $\operatorname{card}(X) = \operatorname{card}(Y)$.

Proof. Let $f : X \to Y$ and $g : Y \to X$ be injections. Consider a point $x \in X$: If $x \in g(Y)$, we form $g^{-1}(x) \in Y$; if $g^{-1}(x) \in f(X)$, we form $f^{-1}(g^{-1}(x))$; and so forth. Either this process can be continued indefinitely, or it terminates with an element of $X \setminus g(Y)$ (perhaps x itself), or it terminates with an element of $Y \setminus f(X)$. In these three cases we say that x is in X_∞, X_X, or X_Y; thus X is the disjoint union of X_∞, X_X, and X_Y. In the same way, Y is the disjoint union of three sets Y_∞, Y_X, and Y_Y. Clearly f maps X_∞ onto Y_∞ and X_X onto Y_X, whereas g maps Y_Y onto X_Y. Therefore, if we define $h : X \to Y$ by $h(x) = f(x)$ if $X \in X_\infty \cup X_X$ and $h(x) = g^{-1}(x)$ if $x \in X_Y$, then h is bijective. ∎

0.9 Proposition. *For any set X,* $\operatorname{card}(X) < \operatorname{card}(\mathcal{P}(X))$.

Proof. On the one hand, the map $f(x) = \{x\}$ is an injection from X to $\mathcal{P}(X)$. On the other, if $g : X \to \mathcal{P}(X)$, let $Y = \{x \in X : x \notin g(x)\}$. Then $Y \notin g(X)$, for if $Y = g(x_0)$ for some $x_0 \in X$, any attempt to answer the question "Is $x_0 \in Y$?" quickly leads to an absurdity. Hence g cannot be surjective. ∎

A set X is called **countable** (or **denumerable**) if $\operatorname{card}(X) \leq \operatorname{card}(\mathbb{N})$. In particular, all finite sets are countable, and for these it is convenient to interpret "$\operatorname{card}(X)$" as the number of elements in X:

$$\operatorname{card}(X) = n \text{ iff } \operatorname{card}(X) = \operatorname{card}(\{1, \ldots, n\}).$$

If X is countable but not finite, we say that X is **countably infinite**.

0.10 Proposition.

a. *If X and Y are countable, so is $X \times Y$.*

b. *If A is countable and X_α is countable for every $\alpha \in A$, then $\bigcup_{\alpha \in A} X_\alpha$ is countable.*

c. *If X is countably infinite, then* $\mathrm{card}(X) = \mathrm{card}(\mathbb{N})$.

Proof. To prove (a) it suffices to prove that $\mathbb{N}^2$ is countable. But we can define a bijection from $\mathbb{N}$ to $\mathbb{N}^2$ by listing, for n successively equal to $2, 3, 4, \ldots$, those elements $(j, k) \in \mathbb{N}^2$ such that $j + k = n$ in order of increasing j, thus:

$$(1,1),\ (1,2),\ (2,1),\ (1,3),\ (2,2),\ (3,1),\ (1,4),\ (2,3),\ (3,2),\ (4,1), \ldots$$

As for (b), for each $\alpha \in A$ there is a surjective $f_\alpha : \mathbb{N} \to X_\alpha$, and then the map $f : \mathbb{N} \times A \to \bigcup_{\alpha \in A} X_\alpha$ defined by $f(n, \alpha) = f_\alpha(n)$ is surjective; the result therefore follows from (a). Finally, for (c) it suffices to assume that X is an infinite subset of $\mathbb{N}$. Let $f(1)$ be the smallest element of X, and define $f(n)$ inductively to be the smallest element of $E \setminus \{f(1), \ldots, f(n-1)\}$. Then f is easily seen to be a bijection from $\mathbb{N}$ to X. ∎

0.11 Corollary. *$\mathbb{Z}$ and $\mathbb{Q}$ are countable.*

Proof. $\mathbb{Z}$ is the union of the countable sets $\mathbb{N}$, $\{-n : n \in \mathbb{N}\}$, and $\{0\}$, and one can define a surjection $f : \mathbb{Z}^2 \to \mathbb{Q}$ by $f(m, n) = m/n$ if $n \neq 0$ and $f(m, 0) = 0$. ∎

A set X is said to have the **cardinality of the continuum** if $\mathrm{card}(X) = \mathrm{card}(\mathbb{R})$. We shall use the letter $\mathfrak{c}$ as an abbreviation for $\mathrm{card}(\mathbb{R})$:

$$\mathrm{card}(X) = \mathfrak{c} \text{ iff } \mathrm{card}(X) = \mathrm{card}(\mathbb{R}).$$

0.12 Proposition. $\mathrm{card}(\mathcal{P}(\mathbb{N})) = \mathfrak{c}$.

Proof. If $A \subset \mathbb{N}$, define $f(A) \in \mathbb{R}$ to be $\sum_{n \in A} 2^{-n}$ if $\mathbb{N} \setminus A$ is infinite and $1 + \sum_{n \in A} 2^{-n}$ if $\mathbb{N} \setminus A$ is finite. (In the two cases, $f(A)$ is the number whose base-2 decimal expansion is $0.a_1a_2 \cdots$ or $1.a_1a_2 \cdots$, where $a_n = 1$ if $n \in A$ and $a_n = 0$ otherwise.) Then $f : \mathcal{P}(\mathbb{N}) \to \mathbb{R}$ is injective. On the other hand, define $g : \mathcal{P}(\mathbb{Z}) \to \mathbb{R}$ by $g(A) = \log(\sum_{n \in A} 2^{-n})$ if A is bounded below and $g(A) = 0$ otherwise. Then g is surjective since every positive real number has a base-2 decimal expansion. Since $\mathrm{card}(\mathcal{P}(\mathbb{Z})) = \mathrm{card}(\mathcal{P}(\mathbb{N}))$, the result follows from the Schröder-Bernstein theorem. ∎

0.13 Corollary. *If* $\mathrm{card}(X) \geq \mathfrak{c}$, *then X is uncountable.*

Proof. Apply Proposition 0.9. ∎

The converse of this corollary is the so-called continuum hypothesis, whose validity is one of the famous undecidable problems of set theory; see §0.7.

0.14 Proposition.

a. *If* $\operatorname{card}(X) \leq \mathfrak{c}$ *and* $\operatorname{card}(Y) \leq \mathfrak{c}$, *then* $\operatorname{card}(X \times Y) \leq \mathfrak{c}$.

b. *If* $\operatorname{card}(A) \leq \mathfrak{c}$ *and* $\operatorname{card}(X_\alpha) \leq \mathfrak{c}$ *for all* $\alpha \in A$, *then* $\operatorname{card}(\bigcup_{\alpha \in A} X_\alpha) \leq \mathfrak{c}$.

Proof. For (a) it suffices to take $X = Y = \mathcal{P}(\mathbb{N})$. Define $\phi, \psi : \mathbb{N} \to \mathbb{N}$ by $\phi(n) = 2n$ and $\psi(n) = 2n - 1$. It is then easy to check that the map $f : \mathcal{P}(\mathbb{N})^2 \to \mathcal{P}(\mathbb{N})$ defined by $f(A, B) = \phi(A) \cup \psi(B)$ is bijective. (b) follows from (a) as in the proof of Proposition 0.10. ∎

0.4 MORE ABOUT WELL ORDERED SETS

The material in this section is optional; it is used only in a few exercises and in some notes at the ends of chapters.

Let X be a well ordered set. If $A \subset X$ is nonempty, A has a minimal element, which is its maximal lower bound or **infimum**; we shall denote it by $\inf A$. If A is bounded above, it also has a minimal upper bound or **supremum**, denoted by $\sup A$. If $x \in X$, we define the **initial segment** of x to be

$$I_x = \{y \in X : y < x\}.$$

The elements of I_x are called **predecessors** of x.

The principle of mathematical induction is equivalent to the fact that $\mathbb{N}$ is well ordered. It can be extended to arbitrary well ordered sets as follows:

0.15 The Principle of Transfinite Induction. *Let X be a well ordered set. If A is a subset of X such that $x \in A$ whenever $I_x \subset A$, then $A = X$.*

Proof. If $X \neq A$, let $x = \inf(X \setminus A)$. Then $I_x \subset A$ but $x \notin A$. ∎

0.16 Proposition. *If X is well ordered and $A \subset X$, then $\bigcup_{x \in A} I_x$ is either an initial segment or X itself.*

Proof. Let $J = \bigcup_{x \in A} I_x$. If $J \neq X$, let $b = \inf(X \setminus J)$. If there existed $y \in J$ with $y > b$, we would have $y \in I_x$ for some $x \in A$ and hence $b \in I_x$, contrary to construction. Hence $J \subset I_b$, and it is obvious that $I_b \subset J$. ∎

0.17 Proposition. *If X and Y are well ordered, then either X is order isomorphic to Y, or X is order isomorphic to an initial segment in Y, or Y is order isomorphic to an initial segment in X.*

Proof. Consider the set $\mathcal{F}$ of order isomorphisms whose domains are initial segments in X or X itself and whose ranges are initial segments in Y or Y itself. $\mathcal{F}$ is nonempty since the unique $f : \{\inf X\} \to \{\inf Y\}$ belongs to $\mathcal{F}$, and $\mathcal{F}$ is partially ordered by inclusion (its members being regarded as subsets of $X \times Y$).

An application of Zorn's lemma shows that $\mathcal{F}$ has a maximal element f, with (say) domain A and range B. If $A = I_x$ and $B = I_y$, then $A \cup \{x\}$ and $B \cup \{y\}$ are again initial segments of X and Y, and f could be extended by setting $f(x) = y$, contradicting maximality. Hence either $A = X$ or $B = Y$ (or both), and the result follows. ∎

0.18 Proposition. *There is an uncountable well ordered set Ω such that I_x is countable for each $x \in \Omega$. If Ω' is another set with the same properties, then Ω and Ω' are order isomorphic.*

Proof. Uncountable well ordered sets exist by the well ordering principle; let X be one. Either X has the desired property or there is a minimal element x_0 such that I_{x_0} is uncountable, in which case we can take $\Omega = I_{x_0}$. If Ω' is another such set, Ω' cannot be order isomorphic to an initial segment of Ω or vice versa, because Ω and Ω' are uncountable while their initial segments are countable, so Ω and Ω' are order isomorphic by Proposition 0.17. ∎

The set Ω in Proposition 0.18, which is essentially unique *qua* well ordered set, is called the **set of countable ordinals**. It has the following remarkable property:

0.19 Proposition. *Every countable subset of Ω has an upper bound.*

Proof. If $A \subset \Omega$ is countable, $\bigcup_{x \in A} I_x$ is countable and hence is not all of Ω. By Proposition 0.16, there exists $y \in \Omega$ such that $\bigcup_{x \in A} I_x = I_y$, and y is thus an upper bound for A. ∎

The set $\mathbb{N}$ of positive integers may be identified with a subset of Ω as follows. Set $f(1) = \inf \Omega$, and proceeding inductively, set $f(n) = \inf(\Omega \setminus \{f(1), \ldots, f(n-1)\})$. The reader may verify that f is an order isomorphism from $\mathbb{N}$ to I_ω, where ω is the minimal element of Ω such that I_ω is infinite.

It is sometimes convenient to add an extra element ω_1 to Ω to form a set $\Omega^* = \Omega \cup \{\omega_1\}$ and to extend the ordering on Ω to Ω^* by declaring that $x < \omega_1$ for all $x \in \Omega$. ω_1 is called the **first uncountable ordinal**. (The usual notation for ω_1 is Ω, since ω_1 is generally taken to be the set of countable ordinals itself.)

0.5 THE EXTENDED REAL NUMBER SYSTEM

It is frequently useful to adjoin two extra points ∞ $(= +\infty)$ and $-\infty$ to $\mathbb{R}$ to form the **extended real number system** $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, and to extend the usual ordering on $\mathbb{R}$ by declaring that $-\infty < x < \infty$ for all $x \in \mathbb{R}$. The completeness of $\mathbb{R}$ can then be stated as follows: Every subset A of $\overline{\mathbb{R}}$ has a least upper bound, or **supremum**, and a greatest lower bound, or **infimum**, which are denoted by $\sup A$ and $\inf A$. If $A = \{a_1, \ldots a_n\}$, we also write

$$\max(a_1, \ldots, a_n) = \sup A, \qquad \min(a_1, \ldots, a_n) = \inf A.$$

From completeness it follows that every sequence $\{x_n\}$ in $\overline{\mathbb{R}}$ has a **limit superior** and a **limit inferior**:

$$\limsup x_n = \inf_{k\geq 1}\Big(\sup_{n\geq k} x_n\Big), \qquad \liminf x_n = \sup_{k\geq 1}\Big(\inf_{n\geq k} x_n\Big).$$

The sequence $\{x_n\}$ converges (in $\mathbb{R}$) iff these two numbers are equal (and finite), in which case its limit is their common value. One can also define lim sup and lim inf for functions $f : \mathbb{R} \to \overline{\mathbb{R}}$, for instance:

$$\limsup_{x\to a} f(x) = \inf_{\delta>0}\Big(\sup_{0<|x-a|<\delta} f(x)\Big).$$

The arithmetical operations on $\mathbb{R}$ can be partially extended to $\overline{\mathbb{R}}$:

$$x \pm \infty = \pm\infty\ (x \in \mathbb{R}), \qquad \infty + \infty = \infty, \qquad -\infty - \infty = -\infty,$$
$$x\cdot(\pm\infty) = \pm\infty\ (x > 0), \qquad x\cdot(\pm\infty) = \mp\infty\ (x < 0).$$

We make no attempt to define $\infty - \infty$, but we abide by the convention that, unless otherwise stated,

$$0\cdot(\pm\infty) = 0.$$

(The expression $0\cdot\infty$ turns up now and then in measure theory, and for various reasons its proper interpretation is almost always 0.)

We employ the following notation for intervals in $\overline{\mathbb{R}}$: if $-\infty \leq a < b \leq \infty$,

$$(a,b) = \{x : a < x < b\}, \qquad [a,b] = \{x : a \leq x \leq b\},$$
$$(a,b] = \{x : a < x \leq b\}, \qquad [a,b) = \{x : a \leq x < b\}.$$

We shall occasionally encounter uncountable sums of nonnegative numbers. If X is an arbitrary set and $f : X \to [0,\infty]$, we define $\sum_{x\in X} f(x)$ to be the supremum of its finite partial sums:

$$\sum_{x\in X} f(x) = \sup\Big\{\sum_{x\in F} f(x) : F \subset X,\ F \text{ finite}\Big\}.$$

(Later we shall recognize this as the integral of f with respect to counting measure on X.)

0.20 Proposition. *Given $f : X \to [0,\infty]$, let $A = \{x : f(x) > 0\}$. If A is uncountable, then $\sum_{x\in X} f(x) = \infty$. If A is countably infinite, then $\sum_{x\in X} f(x) = \sum_1^\infty f(g(n))$ where $g : \mathbb{N} \to A$ is any bijection and the sum on the right is an ordinary infinite series.*

Proof. We have $A = \bigcup_1^\infty A_n$ where $A_n = \{x : f(x) > 1/n\}$. If A is uncountable, then some A_n must be uncountable, and $\sum_{x\in F} f(x) > \operatorname{card}(F)/n$ for F a finite subset of A_n; it follows that $\sum_{x\in X} f(x) = \infty$. If A is countably infinite,

$g : \mathbb{N} \to A$ is a bijection, and $B_N = g(\{1, \ldots, N\})$, then every finite subset F of A is contained in some B_N. Hence

$$\sum_{x \in F} f(x) \leq \sum_1^N f(g(n)) \leq \sum_{x \in X} f(x).$$

Taking the supremum over N, we find

$$\sum_{x \in F} f(x) \leq \sum_1^\infty f(g(n)) \leq \sum_{x \in X} f(x),$$

and then taking the supremum over F, we obtain the desired result. ∎

Some terminology concerning (extended) real-valued functions: A relation between numbers that is applied to functions is understood to hold pointwise. Thus $f \leq g$ means that $f(x) \leq g(x)$ for every x, and $\max(f, g)$ is the function whose value at x is $\max(f(x), g(x))$. If $X \subset \overline{\mathbb{R}}$ and $f : X \to \overline{\mathbb{R}}$, f is called **increasing** if $f(x) \leq f(y)$ whenever $x \leq y$ and **strictly increasing** if $f(x) < f(y)$ whenever $x < y$; similarly for **decreasing.** A function that is either increasing or decreasing is called **monotone.**

If $f : \mathbb{R} \to \mathbb{R}$ is an increasing function, then f has right- and left-hand limits at each point:

$$f(a+) = \lim_{x \searrow a} f(x) = \inf_{x > a} f(x), \qquad f(a-) = \lim_{x \nearrow a} f(x) = \sup_{x < a} f(x).$$

Moreover, the limiting values $f(\infty) = \sup_{a \in \mathbb{R}} f(x)$ and $f(-\infty) = \inf_{a \in \mathbb{R}} f(x)$ exist (possibly equal to $\pm\infty$). f is called **right continuous** if $f(a) = f(a+)$ for all $a \in \mathbb{R}$ and **left continuous** if $f(a) = f(a-)$ for all $a \in \mathbb{R}$.

For points x in $\mathbb{R}$ or $\mathbb{C}$, $|x|$ denotes the ordinary absolute value or modulus of x, $|a + ib| = \sqrt{a^2 + b^2}$. For points x in $\mathbb{R}^n$ or $\mathbb{C}^n$, $|x|$ denotes the Euclidean norm:

$$|x| = \left[\sum_1^n |x_j|^2\right]^{1/2}.$$

We recall that a set $U \subset \mathbb{R}$ is **open** if, for every $x \in U$, U includes an interval centered at x.

0.21 Proposition. *Every open set in $\mathbb{R}$ is a countable disjoint union of open intervals.*

Proof. If U is open, for each $x \in U$ consider the collection $\mathcal{I}_x$ of all open intervals I such that $x \in I \subset U$. It is easy to check that the union of any family of open intervals containing a point in common is again an open interval, and hence $J_x = \bigcup_{I \in \mathcal{I}_x} I$ is an open interval; it is the largest element of $\mathcal{I}_x$. If $x, y \in U$ then either $J_x = J_y$ or $J_x \cap J_y = \varnothing$, for otherwise $J_x \cup J_y$ would be a larger open interval than J_x in $\mathcal{I}_x$. Thus if $\mathfrak{J} = \{J_x : x \in U\}$, the (distinct) members of $\mathfrak{J}$ are disjoint, and $U = \bigcup_{J \in \mathfrak{J}} J$. For each $J \in \mathfrak{J}$, pick a rational number $f(J) \in J$. The map $f : \mathfrak{J} \to \mathbb{Q}$ thus defined is injective, for if $J \neq J'$ then $J \cap J' = \varnothing$; therefore $\mathfrak{J}$ is countable. ∎

0.6 METRIC SPACES

A **metric** on a set X is a function $\rho : X \times X \to [0, \infty)$ such that

- $\rho(x, y) = 0$ iff $x = y$;
- $\rho(x, y) = \rho(y, x)$ for all $x, y \in X$;
- $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ for all $x, y, z \in X$.

(Intuitively, $\rho(x, y)$ is to be interpreted as the distance from x to y.) A set equipped with a metric is called a **metric space**. Some examples:

i. The Euclidean distance $\rho(x, y) = |x - y|$ is a metric on $\mathbb{R}^n$.

ii. $\rho_1(f, g) = \int_0^1 |f(x) - g(x)|\, dx$ and $\rho_\infty(f, g) = \sup_{0 \leq x \leq 1} |f(x) - g(x)|$ are metrics on the space of continuous functions on $[0, 1]$.

iii. If ρ is a metric on X and $A \subset X$, then $\rho|(A \times A)$ is a metric on A.

iv. If (X_1, ρ_1) and (X_2, ρ_2) are metric spaces, the **product metric** ρ on $X_1 \times X_2$ is given by

$$\rho\big((x_1, x_2), (y_1, y_2)\big) = \max\big(\rho_1(x_1, y_1), \rho_2(x_2, y_2)\big).$$

Other metrics are sometimes used on $X_1 \times X_2$, for instance,

$$\rho_1(x_1, y_1) + \rho_2(x_2, y_2) \quad \text{or} \quad \big[\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2\big]^{1/2}.$$

These, however, are equivalent to the product metric in the sense that we shall define at the end of this section.

Let (X, ρ) be a metric space. If $x \in X$ and $r > 0$, the (open) **ball** of radius r about x is

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

A set $E \subset X$ is **open** if for every $x \in E$ there exists $r > 0$ such that $B(r, x) \subset E$, and **closed** if its complement is open. For example, every ball $B(r, x)$ is open, for if $y \in B(r, x)$ and $\rho(x, y) = s$ then $B(r - s, y) \subset B(r, x)$. Also, X and $\varnothing$ are both open and closed. Clearly the union of any family of open sets is open, and hence the intersection of any family of closed sets is closed. Also, the intersection (resp. union) of any finite family of open (resp. closed) sets is open (resp. closed). Indeed, if $U_1, \ldots U_n$ are open and $x \in \bigcap_1^n U_j$, for each j there exists $r_j > 0$ such that $B(r_j, x) \subset U_j$, and then $B(r, x) \subset \bigcap_1^n U_j$ where $r = \min(r_1, \ldots, r_n)$, so $\bigcap_1^n U_j$ is open.

If $E \subset X$, the union of all open sets $U \subset E$ is the largest open set contained in E; it is called the **interior** of E and is denoted by E^o. Likewise, the intersection of all closed sets $F \supset E$ is the smallest closed set containing E; it is called the **closure** of E and is denoted by $\overline{E}$. E is said to be **dense** in X if $\overline{E} = X$, and **nowhere dense** if

$\overline{E}$ has empty interior. X is called **separable** if it has a countable dense subset. (For example, $\mathbb{Q}^n$ is a countable dense subset of $\mathbb{R}^n$.) A sequence $\{x_n\}$ in X **converges** to $x \in X$ (symbolically: $x_n \to x$ or $\lim x_n = x$) if $\lim_{n\to\infty} \rho(x_n, x) = 0$.

0.22 Proposition. *If X is a metric space, $E \subset X$, and $x \in X$, the following are equivalent:*

a. $x \in \overline{E}$.

b. $B(r, x) \cap E \neq \varnothing$ *for all* $r > 0$.

c. There is a sequence $\{x_n\}$ *in* E *that converges to* x.

Proof. If $B(r,x) \cap E = \varnothing$, then $B(r,x)^c$ is a closed set containing E but not x, so $x \notin \overline{E}$. Conversely, if $x \notin \overline{E}$, since $(\overline{E})^c$ is open there exists $r > 0$ such that $B(r,x) \subset (\overline{E})^c \subset E^c$. Thus (a) is equivalent to (b). If (b) holds, for each $n \in \mathbb{N}$ there exists $x_n \in B(n^{-1}, x) \cap E$, so that $x_n \to x$. On the other hand, if $B(r,x) \cap E = \varnothing$, then $\rho(y,x) \geq r$ for all $y \in E$, so no sequence of E can converge to x. Thus (b) is equivalent to (c). ∎

If (X_1, ρ_1) and (X_2, ρ_2) are metric spaces, a map $f : X_1 \to X_2$ is called **continuous** at $x \in X$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\rho_2(f(y), f(x)) < \epsilon$ whenever $\rho_1(x,y) < \delta$ — in other words, such that $f^{-1}(B(\epsilon, f(x))) \supset B(\delta, x)$. The map f is called **continuous** if it is continuous at each $x \in X_1$ and **uniformly continuous** if, in addition, the δ in the definition of continuity can be chosen independent of x.

0.23 Proposition. $f : X_1 \to X_2$ *is continuous iff* $f^{-1}(U)$ *is open in* X_1 *for every open* $U \subset X_2$.

Proof. If the latter condition holds, then for every $x \in X_1$ and $\epsilon > 0$, the set $f^{-1}(B(\epsilon, f(x)))$ is open and contains x, so it contains some ball about x; this means that f is continuous at x. Conversely, suppose that f is continuous and U is open in X_2. For each $y \in U$ there exists $\epsilon_y > 0$ such that $B(\epsilon_y, y) \subset U$, and for each $x \in f^{-1}(\{y\})$ there exists $\delta_x > 0$ such that $B(\delta_x, x) \subset f^{-1}(B(\epsilon_y, y)) \subset f^{-1}(U)$. Thus $f^{-1}(U) = \bigcup_{x \in f^{-1}(U)} B(\delta_x, x)$ is open. ∎

A sequence $\{x_n\}$ in a metric space (X, ρ) is called **Cauchy** if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$. A subset E of X is called **complete** if every Cauchy sequence in E converges and its limit is in E. For example, $\mathbb{R}^n$ (with the Euclidean metric) is complete, whereas $\mathbb{Q}^n$ is not.

0.24 Proposition. *A closed subset of a complete metric space is complete, and a complete subset of an arbitrary metric space is closed.*

Proof. If X is complete, $E \subset X$ is closed, and $\{x_n\}$ is a Cauchy sequence in E, $\{x_n\}$ has a limit in X. By Proposition 0.22, $x \in \overline{E} = E$. If $E \subset X$ is complete and $x \in \overline{E}$, by Proposition (0.22) there is a sequence $\{x_n\}$ in E converging to x. $\{x_n\}$ is Cauchy, so its limit lies in E; thus $E = \overline{E}$. ∎

In a metric space (X, ρ) we can define the distance from a point to a set and the distance between two sets. Namely, if $x \in X$ and $E, F \subset X$,

$$\rho(x, E) = \inf\{\rho(x, y) : y \in E\},$$
$$\rho(E, F) = \inf\{\rho(x, y) : x \in E,\ y \in F\} = \inf\{\rho(x, F) : x \in E\}.$$

Observe that, by Proposition 0.22, $\rho(x, E) = 0$ iff $x \in \overline{E}$. We also define the **diameter** of $E \subset X$ to be

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

E is called **bounded** if $\operatorname{diam} E < \infty$.

If $E \subset X$ and $\{V_\alpha\}_{\alpha \in A}$ is a family of sets such that $E \subset \bigcup_{\alpha \in A} V_\alpha$, $\{V_\alpha\}_{\alpha \in A}$ is called a **cover** of E, and E is said to be **covered** by the V_α's. E is called totally bounded if, for every $\epsilon > 0$, E can be covered by finitely many balls of radius ϵ. Every totally bounded set is bounded, for if $x, y \in \bigcup_1^n B(\epsilon, z_j)$, say $x \in B(\epsilon, z_1)$ and $y \in B(\epsilon, z_2)$, then

$$\rho(x, y) \leq \rho(x, z_1) + \rho(z_1, z_2) + \rho(z_2, y) \leq 2\epsilon + \max\{\rho(z_j, z_k) : 1 \leq j, k \leq n\}.$$

(The converse is false in general.) If E is totally bounded, so is $\overline{E}$, for it is easily seen that if $E \subset \bigcup_1^n B(\epsilon, z_j)$, then $\overline{E} \subset \bigcup_1^n B(2\epsilon, z_j)$.

0.25 Theorem. *If E is a subset of the metric space (X, ρ), the following are equivalent:*

a. *E is complete and totally bounded.*
b. **(The Bolzano-Weierstrass Property)** *Every sequence in E has a subsequence that converges to a point of E.*
c. **(The Heine-Borel Property)** *If $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there is a finite set $F \subset A$ such that $\{V_\alpha\}_{\alpha \in F}$ covers E.*

Proof. We shall show that (a) and (b) are equivalent, that (a) and (b) together imply (c), and finally that (c) implies (b).

(a) implies (b): Suppose that (a) holds and $\{x_n\}$ is a sequence in E. E can be covered by finitely many balls of radius 2^{-1}, and at least one of them must contain x_n for infinitely many n: say, $x_n \in B_1$ for $n \in N_1$. $E \cap B_1$ can be covered by finitely many balls of radius 2^{-2}, and at least one of them must contain x_n for infinitely many $n \in N_1$: say, $x_n \in B_2$ for $n \in N_2$. Continuing inductively, we obtain a sequence of balls B_j of radius 2^{-j} and a decreasing sequence of subsets N_j of $\mathbb{N}$ such that $x_n \in B_j$ for $n \in N_j$. Pick $n_1 \in N_1$, $n_2 \in N_2, \ldots$ such that $n_1 < n_2 < \cdots$. Then $\{x_{n_j}\}$ is a Cauchy sequence, for $\rho(x_{n_j}, x_{n_k}) < 2^{1-j}$ if $k > j$, and since E is complete, it has a limit in E.

(b) implies (a): We show that if either condition in (a) fails, then so does (b). If E is not complete, there is a Cauchy sequence $\{x_n\}$ in E with no limit in E. No subsequence of $\{x_n\}$ can converge in E, for otherwise the whole squence would converge to the same limit. On the other hand, if E is not totally bounded, let $\epsilon > 0$

be such that E cannot be covered by finitely many balls of radius ϵ. Choose $x_n \in E$ inductively as follows. Begin with any $x_1 \in E$, and having chosen $x_1, \ldots, x_n$, pick $x_{n+1} \in E \setminus \bigcup_1^n B(\epsilon, x_j)$. Then $\rho(x_n, x_m) > \epsilon$ for all n, m, so $\{x_n\}$ has no convergent subsequence.

(a) and (b) imply (c): It suffices to show that if (b) holds and $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there exists $\epsilon > 0$ such that every ball of radius ϵ that intersects E is contained in some V_α, for E can be covered by finitely many such balls by (a). Suppose to the contrary that for each $n \in \mathbb{N}$ there is a ball B_n of radius 2^{-n} such that $B_n \cap E \neq \varnothing$ and B_n is contained in no V_α. Pick $x_n \in B_n \cap E$; by passing to a subsequence we may assume that $\{x_n\}$ converges to some $x \in E$. We have $x \in V_\alpha$ for some α, and since V_α is open, there exists $\epsilon > 0$ such that $B(\epsilon, x) \subset V_\alpha$. But if n is large enough so that $\rho(x_n, x) < \epsilon/3$ and $2^{-n} < \epsilon/3$, then $B_n \subset B(\epsilon, x) \subset V_\alpha$, contradicting the assumption on B_n.

(c) implies (b): If $\{x_n\}$ is a sequence in E with no convergent subsequence, for each $x \in E$ there is a ball B_x centered at x that contains x_n for only finitely many n (otherwise some subsequence would converge to x). Then $\{B_x\}_{x \in E}$ is a cover of E by open sets with no finite subcover. ∎

A set E that possesses the properties (a)–(c) of Theorem 0.25 is called **compact**. Every compact set is closed (by Proposition 0.24) and bounded; the converse is false in general but true in $\mathbb{R}^n$.

0.26 Proposition. *Every closed and bounded subset of $\mathbb{R}^n$ is compact.*

Proof. Since closed subsets of $\mathbb{R}^n$ are complete, it suffices to show that bounded subsets of $\mathbb{R}^n$ are totally bounded. Since every bounded set is contained in some cube

$$Q = [-R, R]^n = \{x \in \mathbb{R}^n : \max(|x_1|, \ldots, |x_n|) \leq R\},$$

it is enough to show that Q is totally bounded. Given $\epsilon > 0$, pick an integer $k > R\sqrt{n}/\epsilon$, and express Q as the union of k^n congruent subcubes by dividing the interval $[-R, R]$ into k equal pieces. The side length of these subcubes is $2R/k$ and hence their diameter is $\sqrt{n}(2R/k) < 2\epsilon$, so they are contained in the balls of radius ϵ about their centers. ∎

Two metrics ρ_1 and ρ_2 on a set X are called **equivalent** if

$$C\rho_1 \leq \rho_2 \leq C'\rho_1 \text{ for some } C, C' > 0.$$

It is easily verified that equivalent metrics define the same open, closed, and compact sets, the same convergent and Cauchy sequences, and the same continuous and uniformly continuous mappings. Consequently, most results concerning metric spaces depend not on the particular metric chosen but only on its equivalence class.

0.7 NOTES AND REFERENCES

§§0.1–0.4: The best exposition of set theory for beginners is Halmos [63], and Smullyan and Fitting [135] is a good text on a more advanced level. Kelley [83] also contains a concise account of of basic axiomatic set theory. All of these books present a deduction of the Hausdorff maximal principle from the axiom of choice, as does Hewitt and Stromberg [76].

The axiom of choice (or one of the propositions equivalent to it) is generally taken as one of the basic postulates in the axiomatic formulations of set theory. Some mathematicians of the intuitionist or constructivist persuasion reject it on the grounds that one has not proved the existence of a mathematical object until one has shown how to construct it in some reasonably explicit fashion, whereas the whole point of the axiom of choice is to provide existence theorems when constructive methods fail (or are too cumbersome for comfort). People who are seriously bothered by such objections belong to a minority that does not include the present writer; in this book the axiom of choice is used sparingly but freely.

The **continuum hypothesis** is the assertion that if $\text{card}(X) < \mathfrak{c}$, then X is countable. (Since it follows easily from the construction of Ω, the set of countable ordinals, that $\text{card}(\Omega) \leq \text{card}(X)$ for any uncountable X, an equivalent assertion is that $\text{card}(\Omega) = \mathfrak{c}$.) It is known, thanks to Gödel and Cohen, that the continuum hypothesis and its negation are both consistent with the standard axioms of set theory including the axiom of choice, assuming that those axioms are themselves consistent. (An exposition of the consistency and independence theorems for the axiom of choice and the continuum hypothesis can be found in Smullyan and Fitting [135].) Some mathematicians are willing to accept the continuum hypothesis as true, seemingly as a matter of convenience, but Gödel [56] and Cohen [26, p. 151] have both expressed suspicions that it should be false, and as of this writing no one has found any really compelling evidence on one side or the other. My own feeling, subject to revision in the event of a major breakthrough in set theory, is that if the answer to one's question turns out to depend on the continuum hypothesis, one should give up and ask a different question.

§0.6: A more detailed discussion of metric spaces can be found in Loomis and Sternberg [95] and DePree and Swartz [32].

1

Measures

In this chapter we set forth the basic concepts of measure theory, develop a general procedure for constructing nontrivial examples of measures, and apply this procedure to construct measures on the real line.

1.1 INTRODUCTION

One of the most venerable problems in geometry is to determine the area or volume of a region in the plane or in 3-space. The techniques of integral calculus provide a satisfactory solution to this problem for regions that are bounded by "nice" curves or surfaces but are inadequate to handle more complicated sets, even in dimension one. Ideally, for $n \in \mathbb{N}$ we would like to have a function μ that assigns to each $E \subset \mathbb{R}^n$ a number $\mu(E) \in [0, \infty]$, the n-dimensional measure of E, such that $\mu(E)$ is given by the usual integral formulas when the latter apply. Such a function μ should surely possess the following properties:

i. If $E_1, E_2, \ldots$ is a finite or infinite sequence of disjoint sets, then

$$\mu(E_1 \cup E_2 \cup \cdots) = \mu(E_1) + \mu(E_2) + \cdots.$$

ii. If E is congruent to F (that is, if E can be transformed into F by translations, rotations, and reflections), then $\mu(E) = \mu(F)$.

iii. $\mu(Q) = 1$, where Q is the unit cube

$$Q = \{x \in \mathbb{R}^n : 0 \leq x_j < 1 \text{ for } j = 1, \ldots, n\}.$$

Unfortunately, these conditions are mutually inconsistent. Let us see why this is true for $n = 1$. (The argument can easily be adapted to higher dimensions.) To begin with, we define an equivalence relation on $[0, 1)$ by declaring that $x \sim y$ iff $x - y$ is rational. Let N be a subset of $[0, 1)$ that contains precisely one member of each equivalence class. (To find such an N, one must invoke the axiom of choice.) Next, let $R = \mathbb{Q} \cap [0, 1)$, and for each $r \in R$ let

$$N_r = \{x + r : x \in N \cap [0, 1 - r)\} \cup \{x + r - 1 : x \in N \cap [1 - r, 1)\}.$$

That is, to obtain N_r, shift N to the right by r units and then shift the part that sticks out beyond $[0, 1)$ one unit to the left. Then $N_r \subset [0, 1)$, and every $x \in [0, 1)$ belongs to precisely one N_r. Indeed, if y is the element of N that belongs to the equivalence class of x, then $x \in N_r$ where $r = x - y$ if $x \geq y$ or $r = x - y + 1$ if $x < y$; on the other hand, if $x \in N_r \cap N_s$, then $x - r$ (or $x - r + 1$) and $x - s$ (or $x - s + 1$) would be distinct elements of N belonging to the same equivalence class, which is impossible.

Suppose now that $\mu : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ satisfies (i), (ii), and (iii). By (i) and (ii),

$$\mu(N) = \mu(N \cap [0, 1 - r)) + \mu(N \cap [1 - r, 1)) = \mu(N_r)$$

for any $r \in R$. Also, since R is countable and $[0, 1)$ is the disjoint union of the N_r's,

$$\mu([0, 1)) = \sum_{r \in R} \mu(N_r)$$

by (i) again. But $\mu([0, 1)) = 1$ by (iii), and since $\mu(N_r) = \mu(N)$, the sum on the right is either 0 (if $\mu(N) = 0$) or ∞ (if $\mu(N) > 0$). Hence no such μ can exist.

Faced with this discouraging situation, one might consider weakening (i) so that additivity is required to hold only for finite sequences. This is not a very good idea, as we shall see: The additivity for countable sequences is what makes all the limit and continuity results of the theory work smoothly. Moreover, in dimensions $n \geq 3$, even this weak form of (i) is inconsistent with (ii) and (iii). Indeed, in 1924 Banach and Tarski proved the following amazing result:

> Let U and V be arbitrary bounded open sets in $\mathbb{R}^n$, $n \geq 3$. There exist $k \in \mathbb{N}$ and subsets $E_1, \ldots, E_k, F_1, \ldots, F_k$ of $\mathbb{R}^n$ such that
>
> - the E_j's are disjoint and their union is U;
> - the F_j's are disjoint and their union is V;
> - E_j is congruent to F_j for $j = 1, \ldots, k$.

Thus one can cut up a ball the size of a pea into a finite number of pieces and rearrange them to form a ball the size of the earth! Needless to say, the sets E_j and F_j are *very* bizarre. They cannot be visualized accurately, and their construction depends on the axiom of choice. But their existence clearly precludes the construction of any $\mu : \mathcal{P}(\mathbb{R}^n) \to [0, \infty]$ that assigns positive, finite values to bounded open sets and satisfies (i) for finite sequences as well as (ii).