REAL ANALYSIS

Modern Techniques and Their Applications

GERALD B. FOLLAND

REAL ANALYSIS

Modern Techniques and Their Applications

GERALD B. FOLLAND
University of Washington

A WILEY-INTERSCIENCE PUBLICA'' 'N

JOHN WILEY & SONS

New York . Chichester . Brisbane . Toronto . Singapore

Library of Congress Cataloging in Publication Data:

Folland, G. B.
Real analysis.

"A Wiley-Interscience publication."
Bibliography: p.
Includes index.
1. Mathematical analysis. 2. Functions of real variables. I. Title

QA300.F67 1984 515 84-10435
ISBN 0-471-80958-6

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

The name "real analysis" is something of an anachronism. Originally applied to the theory of functions of one and several real variables, it has come to encompass several subjects of a more general and abstract nature that underlie much of modern analysis. These general theories and their applications are the subject of this book, which is intended primarily as a text for a graduate-level analysis course. Chapters 1 through 7 are devoted to the core material from measure and integration theory, point set topology, and functional analysis which is a part of most graduate curricula in mathematics, together with a few related but less standard items with which I think all analysts should be acquainted. The last three chapters contain a variety of topics that are meant to introduce some of the other branches of analysis and to illustrate the uses of the preceding material. I believe these topics are all interesting and important, but their selection in preference to others is largely a matter of personal predilection.

The things one needs to know in order to read this book are as follows. (1) First and foremost, the classical theory of functions of real variable: limits and continuity, differentiation and (Riemann) integration, infinite series, uniform convergence, the notion of a metric space. (2) The arithmetic of complex numbers and the basic properties of the complex exponential function $e^{x+iy} = e^x(\cos y + i \sin y)$. (Results from more advanced complex function theory are used only in the proof of the Riesz–Thorin theorem and in a few exercises and remarks.) (3) Some elementary set theory. (4) A bit of linear algebra—actually, not much beyond the definitions of vector spaces, linear mappings, and determinants. All of the necessary material in (1) and (2) can be found, for example, in W. Rudin's *Principles of Mathematical Analysis* (3rd ed., McGraw-Hill, New York, 1976), or T. M. Apostol's *Mathematical Analysis* (2nd ed., Addison-Wesley, Reading, Mass., 1974). A summary of the relevant facts about sets and metric spaces is provided here in the Prologue. The reader should begin this book by examining Sections 1 and 5 of the Prologue to become familiar with my notation and terminology; the rest of the Prologue can then be referred to as needed.

Each chapter concludes with a section entitled "Notes and References." These sections contain miscellaneous remarks, acknowledgments of sources, indications of results not discussed in the text, references for further reading, and historical notes. The latter are quite sketchy, although references to more detailed sources are provided; they are intended mainly to give an idea of how the subject grew out of its classical origins. I found it entertaining and instructive to read some of the original papers and monographs, and I hope to encourage others to do the same.

A sizable portion of this book is devoted to exercises, which are mostly in the form of assertions to be proved and which range from trivial to difficult. Every reader should peruse them, although only the most ambitious will try to work them all out. They serve several purposes: amplification of results and completion of proofs in the text, discussion of examples and counterexamples, applications of theorems, and development of further ideas. Instructors will probably wish to do some of the exercises in class; to maximize flexibility and minimize verbosity, I have followed the principle of "when in doubt, leave it as an exercise." Exercises occur at the end of each section, but they are numbered consecutively within each chapter. In referring to them, "Exercise n" means the nth exercise in the present chapter, while "Exercise $m.n$" means the nth exercise in Chapter m. (The same numbering system is used for sections of chapters).

The topics in the book are arranged so as to allow some flexibility of presentation. For example, Chapters 4 and 5 do not depend on Chapters 1–3 except for a few examples and exercises. On the other hand, if one wishes to proceed quickly to L^p theory, one can skip from Section 3.3 to Sections 5.1 and 5.2 and thence to Chapter 6. The last three chapters are independent of each other, except that the material in Section 8.8 is used in Chapter 9.

The writer of a text on such a well-developed subject as real analysis must necessarily be indebted to his predecessors. I kept a large supply of books on hand while writing this one; they are too numerous to list here, but most of them can be found in the bibliography. I am also happy to acknowledge the influence of two of my teachers: Lynn Loomis, from whose lectures I first learned this subject, and Elias Stein, who has done much to shape my point of view. Finally, I wish to thank my colleagues Edwin Hewitt, Isaac Namioka, Scott Osborne, and Garth Warner for many helpful discussions.

The final draft of this book was written on a VAX 11/750 with the EQN|TROFF text-formatting programs. I am grateful to the people at Bell Laboratories for developing this fine software and to David Ragozin and Douglas Lind for helping me learn how to use it.

GERALD B. FOLLAND

Seattle, Washington
August 1984

GUIDE TO NOTATION

Abstract vector spaces: $L(\mathscr{X}, \mathscr{Y})$ (bounded linear maps), 145. $\mathscr{X}^*$ (dual space), 148.

Analysis on Euclidean space: $f(x\pm)$ (one-sided limits), 32. $x'y$ (dot product), 226. $\partial^\alpha, x^\alpha, \alpha!, |\alpha|$ (multiindex notation), 227. $\mathbf{T}^n$ (n-torus), 229.

Functions and operations on functions: f^+, f^- (positive and negative parts), 45. sgn, 45. χ_E (characteristic function), 45. f_x, f^y (sections), 63. Γ, 76. supp(f) (support), 125. λ_f (distribution function), 189. f^y (translation), 228. $f * g$ (convolution), 230. f_t (dilation), 233. $\hat{f}, \mathscr{F}f$ (Fourier transform), 240. $\check{f}$ (inverse Fourier transform), 243. $\langle f, \varphi \rangle$, 258. $\tilde{f}$ (reflection), 259.

Integrals: The basic notation is developed in Section 2.2. $\int f(x)\,dx$ (Lebesgue integral), 56, 68. $\int\int f\,d\mu\,d\nu$ (interated integral), 66. $\int g\,dF$ (Stieltjes integral), 99.

Measures: μ_F, 34. m, m^n (Lebesgue measure), 37, 68. $\mu \times \nu$ (product), 62. σ (surface measure on sphere), 74. ν^+, ν^- (positive and negative variations), 82. $|\nu|$ (total variation), 82, 88. $\mu \perp \nu$ (mutual singularity), 82. $\mu \ll \nu$ (absolute continuity), 83. $f\,d\mu$, 84. $d\mu/d\nu$ (Radon-Nikodym derivative), 85. $\mu \hat{\times} \nu$ (Radon product), 221. $\mu * \nu$ (convolution), 281.

Norms and seminorms: $\|f\|_u$ (uniform norm), 115. $\|T\|$ (operator norm), 145. $\|f\|_p$ (L^p norm), 173, 176. $[f]_p$ (weak L^p quasi norm), 191. $\|\mu\|$ (measure norm), 216. $\|f\|_{(s)}$ (Sobolev norm), 268.

Probability theory: $E(X)$ (expectation), 287. $\sigma(X)$ (standard deviation), 287. $\sigma^2(X)$ (variance), 287. P_φ (image measure, distribution), 287. $\nu_\mu^{\sigma^2}$ (normal distribution), 298.

Sets: card(X), 6. $\mathfrak{c}$ (cardinality of the continuum), 8. $F_\sigma, F_{\sigma\delta}, G_\delta, G_{\delta\sigma}$, 21. E_x, E^y (sections), 63.

σ-algebras: $\mathscr{M}(\mathscr{E})$ (σ-algebra generated by $\mathscr{E}$), 21. $\mathscr{B}_X$ (Borel sets), 21. $\otimes_{\alpha \in A} \mathscr{M}_\alpha$, $\mathscr{M} \otimes \mathscr{N}$ (products), 22. $\mathscr{L}, \mathscr{L}^n$ (Lebesgue measurable sets), 37, 68. $\mathscr{B}_X^0$ (Baire sets), 208.

Spaces of functions, measures, and distributions: L^+, 47. L^1, 52. L^1_{loc}, 90. BV, 97. NBV, 98. $C(X, Y)$, 113. $B(X, \mathbf{R})$, 114. $BC(X, \mathbf{R})$, 115. $B(X)$, 115. $C(X)$, 115. $BC(X)$, 115. $C_c(X)$, 125. $C_0(X)$, 126. L^2, 164, 173. l^2, 169, 173. L^p, 173,

176. l^p, 173. weak L^p, 191. $M(X)$, 216. C^k, 226. C^∞, 226. C_c^∞, 226. $\mathscr{S}$, 227. $\mathscr{S}'$, 258. $\mathscr{D}'$, 262. $\mathscr{E}'$, 263. H_s, 268. H_s^{loc}, 271.

Note: For the basic notation used throughout the book for sets, numbers, and metric and topological spaces see Sections 1, 5, and 6 of the Prologue, and Section 4.1. Notation used only in the section in which it is introduced is, for the most part, not listed here.

CONTENTS

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.

1. THE LANGUAGE OF SET THEORY

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

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 "$-B$ 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 deducing an absurdity.

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

$\mathbf{N}$ = the set of positive integers (not including zero)
$\mathbf{Z}$ = the set of integers
$\mathbf{Q}$ = the set of rational numbers
$\mathbf{R}$ = the set of real numbers
$\mathbf{C}$ = the set of complex numbers

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 $\mathscr{P}(X)$:

$$\mathscr{P}(X) = \{E\colon E \subset X\},$$

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

$$\bigcup_{E\in\mathscr{E}} E = \{x\colon x \in E \text{ for some } E \in \mathscr{E}\},$$

$$\bigcap_{E\in\mathscr{E}} E = \{x\colon x \in E \text{ for every } E \in \mathscr{E}\}.$$

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

$$\mathscr{E} = \{E_\alpha\colon \alpha \in A\} = \{E_\alpha\}_{\alpha\in A},$$

in which case the union and intersection are denoted by

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

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 **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

$$\limsup E_n = \{x\colon x \in E_n \text{ for infinitely many } n\},$$

$$\liminf E_n = \{x\colon x \in E_n \text{ for all but finitely many } n\}.$$

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

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

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

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

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:**

$$\left(\bigcup_{\alpha \in A} E_\alpha\right)^c = \bigcap_{\alpha \in A} E_\alpha^c, \quad \left(\bigcap_{\alpha \in A} E_\alpha\right)^c = \bigcup_{\alpha \in A} E_\alpha^c.$$

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:

(i) Equivalence relations. An **equivalence relation** on X is a relation R on X such that xRx for all x, xRy iff yRx, and xRz whenever xRy and yRz for some y. The **equivalence class** of $x \in X$ is $\{y \in X \colon xRy\}$; X is the disjoint union of these equivalence classes.

(ii) Orderings. See Section 2.

(iii) Mappings. A **mapping** $f \colon X \to Y$ 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 $\mathbf{C}$ or some subset thereof.

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

$$g \circ f \colon X \to Z, \quad g \circ f(x) = g(f(x)).$$

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

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

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

$$f^{-1}\Big(\bigcup_{\alpha\in A} E_\alpha\Big) = \bigcup_{\alpha\in A} f^{-1}(E_\alpha),$$

$$f^{-1}\Big(\bigcap_{\alpha\in A} E_\alpha\Big) = \bigcap_{\alpha\in A} f^{-1}(E_\alpha),$$

$$f^{-1}(E^c) = \big(f^{-1}(E)\big)^c.$$

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

If $f\colon X \to Y$ 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** $f^{-1}\colon Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity mappings on X and Y. If $A \subset X$, we denote by $f|A$ the restriction of f to A:

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

A **sequence** in a set X is a mapping from $\mathbf{N}$ into X. (We also use the term **finite sequence** to mean a map from $\{1, \ldots, n\}$ into X where $n \in \mathbf{N}$.) If $f\colon \mathbf{N} \to X$ is a sequence and $g\colon \mathbf{N} \to \mathbf{N}$ satisfies $g(n) < g(m)$ whenever $n < m$, the composition $f \circ g$ is called a **subsequence** of f. It is common, and usually convenient, to be careless about distinguishing between sequences and their ranges, which are subsets of X indexed by $\mathbf{N}$. Thus, if $f(n) = x_n$, we speak of the sequence $\{x_n\}_1^\infty$; whether we mean a mapping from $\mathbf{N}$ 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 $f\colon A \to \bigcup_{\alpha\in A} X_\alpha$ 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** $\pi_\alpha\colon X \to X_\alpha$ 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, then $\prod_{\alpha\in A} X_\alpha$ is simply the set of all mappings from A into Y, and it is denoted by Y^A. 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.

2. ORDERINGS

A **partial ordering** on a nonempty set X is a relation R on X such that (i) if xRy and yRz then xRz; (ii) if xRy and yRx then $x = y$; and (iii) xRx for all x. If R also satisfies (iv) 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 $\mathscr{P}(E)$ is partially ordered by inclusion, and $\mathbf{R}$ 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 $x \leq y$ and $x \neq y$. 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 (minimal) element** of X is an element $x \in X$ such that the only $y \in X$ satisfying $x \leq y$ ($y \leq x$) 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 (lower) bound** for E is an element $x \in X$ such that $y \leq x$ ($x \leq y$) for all $y \in E$. An upper bound for E need not be an element of E, and the reader should verify that unless E is linearly ordered, a maximal element of E need not be an upper bound for E.

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, $\mathbf{N}$ is well ordered by its natural ordering.

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

(P.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 $E \subset X$ which is linearly ordered by $\leq$, such that no subset of X which properly includes E is linearly ordered by $\leq$. Another version of this principle is the following:

(P.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 (P.1) implies (P.2): 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 (P.2) implies (P.1). [Apply (P.2) to the collection of linearly ordered subsets of X, which is partially ordered by inclusion.]

(P.3) The Well Ordering Principle. Every nonempty set X can be well ordered.

Proof. Consider the collection $\mathscr{W}$ of well orderings of subsets of X. Such well orderings may be regarded as subsets of $X \times X$, so $\mathscr{W}$ is partially ordered by inclusion. It is easy to verify that the hypotheses of Zorn's lemma are satisfied, so $\mathscr{W}$ has a maximal element. This must be a well ordering of X itself, for if $\leq$ is a well ordering on a proper subset E of X and $x_0 \in X \setminus E$, $\leq$ can be extended to a well ordering on $E \cup \{x_0\}$ by declaring that $x_0 \leq x$ for all $x \in E$. □

(P.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 $f \in \prod_{\alpha \in A} X_\alpha$. □

(P.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 every $\alpha \in A$.

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, the two are known to be logically equivalent.

3. CARDINALITY

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

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

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

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

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

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

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 which the notation suggests.

(P.6) Proposition. card(X) ≤ card(Y) iff card(Y) ≥ card(X).

Proof. If $f\colon X \to Y$ is injective, pick $x_0 \in X$ and define $g\colon Y \to X$ by $g(y) = f^{-1}(y)$ if $y \in f(X)$, $g(y) = x_0$ otherwise. Then g is surjective. Conversely, if $g\colon Y \to X$ 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. □

(P.7) Proposition. For any sets X and Y, either card(X) ≤ card(Y) or card(Y) ≤ card(X).

Proof. Consider the set $\mathscr{I}$ of all injections from subsets of X into Y. The members of $\mathscr{I}$ can be regarded as subsets of $X \times Y$, so $\mathscr{I}$ is partially ordered by inclusion. It is easily verified that Zorn's lemma applies, so $\mathscr{I}$ 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 $B \cup \{y_0\}$ by setting $f(x_0) = y_0$, contradicting maximality. Hence either $A = X$, in which case card(X) ≤ card(Y), or $B = Y$, in which case f^{-1} is an injection from Y to X and card(Y) ≤ card(X). □

(P.8) The Schröder–Bernstein Theorem. If card(X) ≤ card(Y) and card(Y) ≤ card(X), then card(X) = card(Y).

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ be injections. Consider a point $x \in X$: if $x \in \text{range}(g)$ we form $g^{-1}(x) \in Y$; if $g^{-1}(x) \in \text{range}(f)$ we form $f^{-1}(g^{-1}(x)) \in X$; and so forth. Either this process can be continued indefinitely, or it terminates with an element of $X \setminus \text{range}(g)$ (perhaps x itself), or it terminates with an element of $Y \setminus \text{range}(f)$. 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 $h\colon X \to Y$ 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. □

(P.9) Proposition. For any set X, card(X) < card($\mathscr{P}(X)$).

Proof. Given $f\colon X \to \mathscr{P}(X)$, let $Y = \{x \in X\colon\ x \notin f(x)\}$. Then $Y \notin \text{range}(f)$, for if $Y = f(x_0)$ for some $x_0 \in X$, any attempt to answer the question "is $x_0 \in Y$?" quickly leads to an absurdity. Hence no map from X to $\mathscr{P}(X)$ can be surjective. □

A set X is called **countable** if card(X) ≤ card(N). In particular, all finite sets are countable, and for these it is convenient to interpret "card(X)" as the number of elements in X:

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