INTRODUCTION TO FUNCTIONAL ANALYSIS

SECOND EDITION

Angus E. Taylor

David C. Lay

INTRODUCTION TO FUNCTIONAL ANALYSIS

SECOND EDITION

Angus E. Taylor
University of California,
Berkeley and Los Angeles

David C. Lay
University of Maryland
College Park

JOHN WILEY & SONS
New York Chichester Brisbane Toronto

Library of Congress Cataloging in Publication Data:

Taylor, Angus Ellis, 1911–
Introduction to functional analysis.

Bibliography: p.
Includes index.
1. Functional analysis. I. Lay, David C., joint author. II. Title.
QA320.T3 1980 515'.7 79-13658
ISBN 0-471-84646-5

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To our wives—
PATSY and LILLIAN

PREFACE

The central theme of this book is the theory of normed linear spaces and of linear mappings between such spaces. The text provides the necessary foundation for further study in many areas of analysis, and it strives to generate an appreciation for the unifying power of the abstract linear-space point of view in surveying the problems of linear algebra, classical analysis, and differential and integral equations. While the book is principally addressed to graduate students, it is also intended to be useful to mathematicians and users of mathematics who have need of a simple and direct presentation of the fundamentals of the theory of linear spaces and linear operators.

In many respects, this new edition is similar to the first edition written by Taylor. The prerequisites are the same—the reader should already be acquainted with the fundamentals of real and complex analysis and elementary point set topology. The manner and level of presentation are essentially unchanged, although the scope of the text has been broadened somewhat, and the emphasis on concrete examples and connections with classical mathematics has been retained.

The revision was made in order to incorporate recent developments in functional analysis and to make the selection of topics more appropriate for current courses in functional analysis. Significant additions to this new edition include a chapter on Banach algebras, and material on weak topologies and duality, equicontinuity, the Krein–Milman theorem, and the theory of Fredholm operators. Furthermore, there is greater emphasis on closed unbounded linear operators, with more illustrations drawn from ordinary differential equations. Two background chapters from the first edition (on topology and some topics in integration theory) have been omitted because the material has become part of the standard curriculum. A few facts from those chapters are now reviewed as needed.

The problems in the text, increased from 300 to nearly 500, have been carefully selected—they both illustrate and extend the theory, and they give the reader an opportunity to construct arguments similar to those in the text. The ℓ^p sequence spaces have been chosen for some of the concrete problems and examples, in order to minimize the technical difficulties. In addition, there are problems that relate to differential equations, integral equations, and the theory of analytic functions.

The book begins with a concise review of linear algebra and an introduction to linear problems in analysis, omitting topological considerations. An analytic version of the Hahn–Banach theorem is given in § I.10. A well-prepared reader may start with Chapter II, using Chapter I as a reference when necessary. The first part of Chapter II is on normed linear spaces. However, the basic theory of topological linear spaces is important because it is frequently used today in analysis and because it is relevant in the study of normed linear spaces. This basic theory is developed in parts of Chapters II and III. Certain optional topics are identified in the introductions to these chapters.

The subject of linear operators is begun in detail in Chapter IV, with some of the most important results in Chapters IV and V depending on completeness of the underlying spaces. However, unless needed for effective results, the hypothesis of completeness is not invoked. The exposition is not materially lengthened by this greater generality. The more specialized and distinctive theory of operators on Hilbert space is presented in §§ 3, 11, and 12 of Chapter IV.

Chapter V stresses the importance of complex contour integration and the calculus of residues in the spectral theory of linear operators. The methods apply to all closed linear operators, bounded or not. Although some familiarity with complex analysis is assumed in the text, the main theorems needed for Chapter V are reviewed in § V.1. This chapter also contains the famous Riesz theory of compact operators, as extended and perfected by later research workers, and its application to the classical "determinant-free" theorems for Fredholm integral equations of the second kind. The subject of invariant subspaces is treated briefly in this chapter as well.

Chapter VI presents the standard elementary theory of self-adjoint, normal, and unitary operators on Hilbert space. The discussion of the theory of compact symmetric operators and symmetric operators with compact resolvent is very important for applications to integral and differential equations. The completeness of the inner-product space under consideration is not required here. The spectral analysis of self-adjoint operators is performed with the aid of the Riesz representation theorem for linear functionals on a space of continuous functions. The treatment is deliberately kept as close as possible to classical analysis.

General Banach algebras and the Gelfand theory of commutative Banach algebras are discussed in Chapter VII. Most of the development assumes the existence of a unit, but some examples and problems show how the general theory would proceed for algebras such as $L^1(\boldsymbol{R})$ that lack a unit. The chapter is largely independent of Chapters IV to VI, except for the material in the early sections of these three chapters. The text returns to operator theory in the final section of Chapter VII, where two versions of the spectral theorem for normal operators on Hilbert space are derived from the Gelfand–Naimark theorem for commutative B^*-algebras.

The senior author (Taylor) wishes to include here a personal acknowledgment of thanks to his coauthor: When I left U.C.L.A. to become the academic vice-president of the University of California multicampus system, I soon realized that my busy schedule would not enable me to carry forward the task of revising the book and of tuning it anew to the needs of an oncoming generation of students. Fortunately, David Lay, to whom I had been close in his student days at U.C.L.A., was able and willing to coauthor this book. He has done most of the revision, but we have been in close touch throughout. I am both obligated by what he has done and immensely pleased by his accomplishment and judgment.

One of the pleasures in writing books comes from the opportunities afforded authors to improve the breadth and depth of their own understanding of a subject in relation to its origins and applications. We hope that some of the same pleasure will accrue to readers of this text.

We are especially grateful to Professor Steven Lay who worked closely with the junior author one summer and who made valuable contributions during the early stages of the revision. We are also indebted to Professor Denny Gulick whose suggestions and critical analysis of the manuscript led to substantial improvements in the text. Our appreciation is extended to Professors John Brace, Robert Ellis, and John Horváth, who class-tested portions of the text material and made helpful comments, and to Professor Seymour Goldberg for many stimulating discussions. We also wish to thank Professor Ronald Douglas for his review of the manuscript and useful remarks.

Finally, we wish to thank our many students and colleagues who have, through communications or conversations, influenced us in specific or perhaps even unrealized ways and thereby, we hope, helped make a better book.

Berkeley, California
College Park, Maryland
March 1979

Angus E. Taylor
David C. Lay

CONTENTS

INTRODUCTION

It is the purpose of this introduction to explain certain terminology used throughout the book and to list some inequalities for easy reference.

FUNCTIONS

Let X and Y be arbitrary nonempty sets and let $\mathcal{D}$ be a nonempty subset of X. A *function* f from $\mathcal{D}$ into Y is a rule that to each $x \in \mathcal{D}$ assigns a unique element $f(x)$ in Y. We sometimes denote the function by the expression $x \mapsto f(x)$. The *domain* of f is the set $\mathcal{D}$, often written as $\mathcal{D}(f)$, and the *range* of f is the set $\mathcal{R}(f) = \{f(x): x \in \mathcal{D}\}$. The *graph* of f (sometimes identified with f itself) is the set of ordered pairs $\{(x, f(x)): x \in \mathcal{D}\}$. This is a particular kind of nonempty subset of the Cartesian product $X \times Y$ of all ordered pairs (x, y), where $x \in X$, $y \in Y$. A function g is said to be a *restriction* of f, and f an *extension* of g, if $\mathcal{D}(g) \subset \mathcal{D}(f)$ and $g(x) = f(x)$ for $x \in \mathcal{D}(g)$.

We say that a function f is *injective*, or *one-to-one*, if for each y in the range $\mathcal{R}(f)$ there exists only one x in the domain $\mathcal{D}(f)$ such that $f(x) = y$, and we denote this unique x by $f^{-1}(y)$. When f is injective, the correspondence $y \mapsto f^{-1}(y)$ is a function f^{-1} called the *inverse* of f, whose domain is $\mathcal{R}(f)$ and range is $\mathcal{D}(f)$. We sometimes say that f^{-1} exists, in place of saying that f is injective.

If each y in Y is in the range of f, we say that f is *surjective*, or that f maps its domain *onto* Y. If f is both injective and surjective, we say that f is *bijective*. In this case f^{-1} maps Y onto the domain of f.

Given $A \subset X$, $B \subset Y$ and f as above, we use the following notation.

$$f(A) = \{f(x): x \in A \cap \mathcal{D}(f)\},$$

$$f^{-1}(B) = \{x \in \mathcal{D}(f): f(x) \in B\}.$$

We call $f(A)$ the image of A under f. Note that $f(A) = \varnothing$, where $\varnothing$ denotes the empty set, if $A \cap \mathcal{D}(f) = \varnothing$. We call $f^{-1}(B)$ the inverse image of B under f. We write $f^{-1}(B)$ even though f^{-1} may not exist as a function. [In fact, f^{-1} exists if and only if $f^{-1}(B)$ is a set consisting of a single element of $\mathcal{D}(f)$ whenever B is a set consisting of just one element of $\mathcal{R}(f)$.]

REAL AND COMPLEX NUMBERS

The real number system is denoted by $\boldsymbol{R}$, the complex numbers by C. The *real* and *imaginary parts* of a complex number λ are written as $\operatorname{Re}\lambda$ and $\operatorname{Im}\lambda$, respectively. On occasion it will be necessary to work within the extended real number system $\boldsymbol{R} \cup \{+\infty\} \cup \{-\infty\}$. Algebraic operations in this system are discussed in Taylor [5, pages 178–180].* The *least upper bound* (*supremum*) and *greatest lower bound* (*infimum*) of a nonempty set S of real numbers always exist in the extended real number system; they are denoted by $\sup S$ and $\inf S$, respectively.

TOPOLOGY

We assume the reader is familiar with the basic definitions and theorems of topology, such as those presented in Taylor [5, pages 89–140]. The *closure* of a set A in a topological space X is denoted by $\bar{A}$, the *interior* by $\operatorname{int}(A)$. We say that a set V in X is a *neighborhood* of a point $x \in X$ if there exists an open set U such that $x \in U$ and $U \subset V$. (This usage follows the Bourbaki tradition and differs from Taylor [5], where a neighborhood of x means an open set containing x.) The space X is called a T_1 space if every set consisting of a single point is closed. The space is a *Hausdorff* space (or T_2 space) if for each pair of distinct points $x_1, x_2 \in X$ there exist disjoint neighborhoods of x_1 and x_2, respectively. If τ_1 and τ_2 are topologies for the same set X, then τ_1 is called *weaker* than τ_2 (equivalently, τ_2 is *stronger* than τ_1) if every τ_1-open set is τ_2-open.

THE KRONECKER DELTA

The symbol δ_{ij} denotes the number 1 if $i = j$ and the number 0 if $i \neq j$. Here i and j are positive integers.

INEQUALITIES

At a number of places in this book we use some of the standard inequalities concerning sums and integrals. We list the most commonly used ones here. The standard reference work on this subject is the book, *Inequalities*, by Hardy, Littlewood, and Pólya [1]. In what follows we refer to this book as H, L, and P, and cite by number the section in which the stated inequality is discussed. Most of the inequalities are given as exercises, with hints for

* References to the bibliography are made by listing the author's name and a number in brackets, identifying a book or article by that author.

solutions, in Taylor [5, pages 119–120, 278]. In all inequalities the quantities involved may be either real or complex. Sums are either all from 1 to n or from 1 to ∞, and in the latter case certain evident assumptions and implications of convergence are involved. For simplicity the inequalities for integrals are written for the case in which the functions are defined on a finite or infinite interval of the real axis. The inequalities are valid with more general interpretations of the set over which integration is extended.

Hölder's inequality for sums (H, L, and P, § 2.8): If $1 < p < \infty$ and $p' = p/(p-1)$, then

$$\sum |a_i b_i| \leq \left(\sum |a_i|^p\right)^{1/p} \left(\sum |b_i|^{p'}\right)^{1/p'}.$$

The special case when $p = p' = 2$ is called Cauchy's inequality (H, L, and P, § 2.4).

Minkowski's inequality for sums (H, L, and P, § 2.11): If $1 \leq p < \infty$, then

$$\left(\sum |a_i + b_i|^p\right)^{1/p} \leq \left(\sum |a_i|^p\right)^{1/p} + \left(\sum |b_i|^p\right)^{1/p}.$$

Jensen's inequality (H, L, and P, § 2.10): If $0 < p < q$, then

$$\left(\sum |a_i|^q\right)^{1/q} \leq \left(\sum |a_i|^p\right)^{1/p}.$$

Hölder's inequality for integrals (H, L, and P, § 6.9): If $1 < p < \infty$ and $p' = p/(p-1)$, then

$$\int |f(x)g(x)|\,dx \leq \left(\int |f(x)|^p\,dx\right)^{1/p} \left(\int |g(x)|^{p'}\,dx\right)^{1/p'}.$$

The special case when $p = p' = 2$ is called the Schwarz inequality (H, L, and P, § 6.5).

Minkowski's inequality for integrals (H, L, and P, § 6.13): If $1 \leq p < \infty$, then

$$\left(\int |f(x) + g(x)|^p\,dx\right)^{1/p} \leq \left(\int |f(x)|^p\,dx\right)^{1/p} + \left(\int |g(x)|^p\,dx\right)^{1/p}.$$

I THE ABSTRACT APPROACH TO LINEAR PROBLEMS

The modern treatment of many topics in pure and applied mathematics is characterized by the effort that is made to strip away nonessential details and to show clearly the fundamental assumptions and the structure of the reasoning. This effort often leads to some degree of abstraction, with the concrete nature of the originally contemplated problem being temporarily put aside and the aspects of the problem that are of greatest significance being cast into axiomatic form. It is found that in this way there is a considerable gain in transparency and that diverse problems exhibit common characteristics that enable them all to be at least partially solved by the methods of a single general theory.

In this chapter we consider the algebraic aspects of such an abstract approach to linear problems. In essence, all linear problems are viewed in some measure as analogous to the linear problems exhibited in elementary algebra by the theory of systems of linear equations. The linear problems of analysis usually require topological as well as algebraic considerations. However, in this chapter, we exclude all concern with topology; the topological aspects of the abstract approach to linear problems will be taken up in later chapters.

The most profound results of the chapter are the extension theorems in § 10 (Theorems 10.1 and 10.4) and Theorem 11.2 on the existence of a complementary subspace. They all depend on Zorn's lemma (§ 9). Theorem 10.4 is one version of the important Hahn–Banach theorem. Other versions will be discussed in § III.2 and § III.3.

Chapter I culminates in § 13 with two theorems relating the range and null space of a linear operator to the null space and range of the transpose of the operator (Theorems 13.4 and 13.5). These theorems furnish information on existence and uniqueness theorems for certain kinds of linear problems. For the finite-dimensional case these theorems include the standard results concerning algebraic systems of linear equations. In the infinite-dimensional case the results are not as useful as results that can be obtained with the aid of metric or topological tools. Nevertheless, the material of § 13 points the way to more incisive results, some of which are given in § IV.8.

I.1 ABSTRACT LINEAR SPACES

We have as yet made no formal definition of what is meant by the adjective *linear* in the phrase "linear problems." We can cite various particular kinds of linear problems: the problems of homogeneous and inhomogeneous systems of linear equations in n "unknowns" in elementary algebra; the problems of the theory of linear ordinary differential equations (existence theorems, particular and general solutions, problems of finding solutions satisfying given conditions at one or two end points); boundary or initial-value problems in the theory of linear partial differential equations; problems in the theory of linear integral equations; linear "transform" problems, for example, problems related to Fourier and Laplace transforms. This is by no means an exhaustive list of the types of mathematical situations in which linear problems arise.

At the bottom of every linear problem is a mathematical structure called a *linear space.* We shall, therefore, begin with an axiomatic treatment of abstract linear spaces.

Definition. Let X be a set of elements, hereafter sometimes called *points*, and denoted by small italic letters: $x, y, \ldots$. We assume that each pair of elements x, y can be combined by a process called addition to yield another element z denoted by $z = x + y$. We also assume that each real number α and each element x can be combined by a process called multiplication to yield another element y denoted by $y = \alpha x$. The set X with these two processes is called a *linear space* if the following axioms are satisfied:

1. $x + y = y + x$.
2. $x + (y + z) = (x + y) + z$.
3. There is in X a unique element, denoted by 0 and called the zero element, such that $x + 0 = x$ for each x.
4. To each x in X corresponds a unique element, denoted by $-x$, such that $x + (-x) = 0$.
5. $\alpha(x + y) = \alpha x + \alpha y$.
6. $(\alpha + \beta)x = \alpha x + \beta x$.
7. $\alpha(\beta x) = (\alpha\beta)x$.
8. $1 \cdot x = x$.
9. $0 \cdot x = 0$.

Anyone who is familiar with the algebra of vectors in ordinary three-dimensional Euclidean space will see at once that the set of all such vectors forms a linear space. An abstract linear space embodies so many of the features of ordinary vector algebra that the word *vector* has been taken over into a more general context. A linear space is often called a *vector space*, and the elements of the space are called *vectors*.

In the foregoing list of axioms it was assumed that the multiplication operation was performed with *real* numbers, α, β. To emphasize this, if necessary, we call the space a *real* linear space, or a *real* vector space. An alternative notion of a linear space is obtained if it is assumed that any *complex* number α and any element x can be multiplied, yielding another element αx. The axioms are the same as before. The space is then called a *complex* linear space.

The notion of a vector space is defined even more generally in abstract algebra, by allowing the multipliers $\alpha, \beta, \ldots$ to be elements of an arbitrary commutative field. In this book, however, we confine ourselves to the two fields of real and complex numbers, respectively. The elements of the field are called *scalars*, to contrast with the *vector* elements of the linear space.

It is easy to see that $-1 \cdot x = -x$ and that $\alpha \cdot 0 = 0$. We write $x - y$ for convenience in place of $x + (-y)$. The following "cancellation" rules are also easily deduced from the axioms:

(1-1) $$x + y = x + z \text{ implies } y = z;$$

(1-2) $$\alpha x = \alpha y \text{ and } \alpha \neq 0 \text{ imply } x = y;$$

(1-3) $$\alpha x = \beta x \text{ and } x \neq 0 \text{ imply } \alpha = \beta.$$

Definition. A nonempty subset M of a linear space X is called a *linear manifold* in X if $x + y$ is in M whenever x and y are both in M and if also αx is in M whenever x is in M and α is any scalar.

In this definition and generally throughout the book, statements made about linear spaces, without qualification as to whether the space is real or complex, will be intended to apply equally to real spaces and complex spaces.

It will be seen at once that, if M is a linear manifold in X, it may be regarded as a linear space by itself. For, if x is in M, then $-1 \cdot x = -x$ is also in M, and $x - x = 0$ is also in M. The nine axioms for a linear space are now found to be satisfied in M. Another term for a linear manifold in X is *subspace* of X. A subspace of X is called *proper* if it is not all of X.

The set consisting of 0 alone is a subspace. We denote it by (0).

Suppose S is any nonempty subset of X. Consider the set M of all finite linear combinations of elements of S, that is, elements of the form $\alpha_1 x_1 + \cdots + \alpha_n x_n$, where n is any positive integer (not fixed), $x_1, \ldots, x_n$ are any elements of S, and $\alpha_1, \ldots, \alpha_n$ are any scalars. This set M is a linear manifold. It is called the linear manifold generated, or determined, by S. Sometimes we speak of M as the linear manifold *spanned* by S. It is easy to verify the truth of the following statements: (1) M consists of those vectors that belong to every linear manifold that contains S; that is, M is the intersection of all such manifolds. (2) M is the smallest linear manifold that contains S; that is, if N is a linear manifold that contains S, then M is contained in N.

One of the most important concepts in a vector space is that of linear dependence.

Definition. A finite set of vectors $x_1, \ldots, x_n$ in the space X is *linearly dependent* if there exists scalars $\alpha_1, \ldots, \alpha_n$, not all zero, such that $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$. If the finite set $x_1, \ldots, x_n$ is not linearly dependent, it is called *linearly independent*. In that case, a relation $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$ implies that $\alpha_1 = \cdots = \alpha_n = 0$. An infinite set S of vectors is called linearly independent if every finite subset of S is linearly independent; otherwise S is called linearly dependent.

Observe that if a set of vectors contains a linearly dependent subset, the whole set is linearly dependent. Also note that a linearly independent set cannot contain the vector 0.

We note the following simple but important theorem, of which use will be made in later arguments:

Theorem 1.1. *Suppose $x_1, \ldots, x_n$ is a set of vectors with $x_1 \neq 0$. The set is linearly dependent if and only if some one of the vectors $x_2, \ldots, x_n$, say x_k, is in the linear manifold generated by $x_1, \ldots, x_{k-1}$.*

Proof. Suppose the set is linearly dependent. There is a smallest integer k, with $2 \leqslant k \leqslant n$, such that the set $x_1, \ldots, x_k$ is linearly dependent. This dependence is expressed by an equation $\alpha_1 x_1 + \cdots + \alpha_k x_k = 0$, with not all the α's equal to zero. Necessarily, then, $\alpha_k \neq 0$, for otherwise $x_1, \ldots, x_{k-1}$ would form a linearly dependent set. Consequently, $x_k = \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1}$, where $\beta_i = -\alpha_i/\alpha_k$. This shows that x_k is in the manifold spanned by $x_1, \ldots, x_{k-1}$. On the other hand, if we assume that some x_k is in the linear manifold spanned by $x_1, \ldots, x_{k-1}$, then an equation of the form $x_k = \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1}$ shows that the set $x_1, \ldots, x_k$ is linearly dependent, whence the same is true of the set $x_1, \ldots, x_n$. □

It is convenient to say that x is a *linear combination* of $x_1, \ldots, x_n$ if it is in the linear manifold spanned by these vectors.

Using the notion of linear dependence, we can define the concept of a finite-dimensional vector space.

Definition. Let X be a vector space. Suppose there is some positive integer n such that X contains a set of n vectors that are linearly independent, while every set of $n+1$ vectors in X is linearly dependent. Then X is called