FINITE DIMENSIONAL VECTOR SPACES

PAUL R. HALMOS

FINITE DIMENSIONAL VECTOR SPACES

BY

PAUL R. HALMOS

PRINCETON, NEW JERSEY
PRINCETON UNIVERSITY PRESS

First Printing 1942
Second Edition, Second Printing 1947
Third Printing 1948
Fourth Printing 1949
Fifth Printing 1951
Sixth Printing 1953
Seventh Printing 1955

Printed in the United States of America

PREFACE

That Hilbert space theory and elementary matrix theory are intimately associated came as a surprise to me and to many colleagues of my generation only after studying the two subjects separately. This is deplorable: it took us as much time to discover for ourselves that there is a connection as it took to learn the two seemingly separate disciplines. I present this little book in an attempt to remedy the situation. Addressing the advanced undergraduate or beginning graduate student, I treat linear transformations on finite dimensional vector spaces by the methods of more general theories. My purpose is to emphasize the simple geometric notions common to many parts of mathematics and its applications, and to do this in a language which gives away the trade secrets and tells the student what is in the back of the minds of people proving theorems about integral equations and Banach spaces. The reader does not, however, have to share my prejudiced motivation. Except for an occasional reference to undergraduate mathematics the book is self contained and may be read by anyone who is trying to get a feeling for the linear problems usually discussed in courses on "matrix theory" or "higher algebra". The algebraic, coordinate - free, methods do not lose power and elegance by specialization to a finite number of dimensions, and are, in my belief, as elementary as the classical coordinatized treatment.

I originally intended this book to contain a theorem if and only if an infinite dimensional generalization of it already exists. Barring a few concessions to the tempting easiness of some essentially finite dimensional notions and results, I have followed this plan. My emphasis, however, is more on method than on results. The reader may sometimes see some obvious way of shortening the proofs I give. (He is, for example, very likely to do this in connection with the representation of a linear functional by an inner product or the treatment of direct products of unitary spaces.) The chances are that the infinite dimensional analog of the shorter proof is either much longer or else non existent.

To supplement the hints in the body of the book concerning the various directions in which a student may proceed, I have appended a bibliography. This very short list makes no pretense to completeness; it consists merely of the books which have helped me the most. Their perusal should give the student an idea of most of the important extensions of the subjects I treat.

In conclusion I want to express my really sincere thanks to virtually every mathematician in Princeton. Most of them have read parts of the manuscript, discussed the project with me, and were very kind in giving encouragement and criticism. I am particularly grateful to two men: John von Neumann, who is one of the originators of the modern spirit and methods which I have tried to present and whose teaching was the inspiration for this book, and J. L. Doob, who read the entire manuscript and made many valuable suggestions.

Paul R. Halmos

The Institute for Advanced Study,
Princeton, New Jersey

TABLE OF CONTENTS

Page

Chapter I

SPACES

§1.

In what follows we shall have occasion to use different classes of numbers (such as the class of all real numbers or the class of all complex numbers). Because we don't want, at this early stage, to commit ourselves to any specific class we shall adopt the dodge of referring to numbers as <u>scalars</u>. The reader will not lose anything essential if he consistently interprets scalars as real numbers or as complex numbers: in the examples that we shall give both classes will occur.

DEFINITION. A <u>vector space</u>, $\mathfrak{V}$, is a set of elements x, y, z, etc., called <u>vectors</u>, satisfying the following axioms.

A.

To every pair x and y, of vectors in $\mathfrak{V}$ there corresponds a vector z, called the <u>sum</u> of x and y, $z = x+y$, in such a way that

(1) addition is commutative, $x+y = y+x$;

(2) addition is associative, $x+(y+z) = (x+y) + z$;

(3) there exists in $\mathfrak{V}$ a unique vector, 0, (called the <u>origin</u>) such that for all x in $\mathfrak{V}$, $x+0 = x$; and

(4) to every x in $\mathfrak{V}$ there corresponds a unique vector, denoted by $-x$, with the proper-

ty $x + (-x) = 0$.

B.

To every pair, α and x, where α is a scalar and x is a vector in $\mathfrak{V}$, there corresponds a vector y in $\mathfrak{V}$, called the product of α and x, $y = \alpha x$, such that

(1) multiplication is distributive with respect to vector addition, $\alpha(x+y) = \alpha x + \alpha y$;

(2) multiplication is distributive with respect to scalar addition, $(\alpha + \beta)x = \alpha x + \beta x$

(3) multiplication is associative, $\alpha(\beta x) = (\alpha\beta)x$; and

(4) $0x = 0$, $1x = x$.

(These axioms are not logically independent: they are merely a convenient characterization of the objects we wish to study.) According as scalars are interpreted as real or complex numbers we shall refer to real or complex vector spaces.

§2. EXAMPLES OF VECTOR SPACES

Before discussing the implications of these axioms we give some examples. We shall refer to these examples over and over again and we shall use the notation established here throughout the rest of our work.

(1) Let $\mathfrak{C}_1$ be the set of all complex numbers; if we interpret $x+y$ and αx as ordinary complex numerical addition and multiplication, $\mathfrak{C}_1$ becomes a complex vector space.

(2) Let $\mathfrak{P}$ be the set of all polynomials with complex coefficients in a real variable t. (There is no deep reason for this arbitrary choice: it is merely a matter of convenience for the purpose of giving examples later). To make $\mathfrak{P}$ into a complex vector space we interpret addition and scalar multiplication as the ordina-

ry addition of two polynomials and multiplication of a polynomial by a complex number, respectively; the origin in $\mathfrak{P}$ is the polynomial identically zero.

Example (1) is too simple and example (2) too complicated to be typical of the main contents of this book. We give now another example of complex vector spaces which (as we shall see later) is general enough for all our purposes.

(3) Let $\mathfrak{C}_n$, $n = 1, 2, \ldots$, be the set of all n-tuples of complex numbers, $x = \{\xi_1, \ldots, \xi_n\}$; if $y = \{\eta_1, \ldots, \eta_n\}$ we write, by definition,

$$x+y = \{\xi_1 + \eta_1, \ldots, \xi_n + \eta_n\},$$
$$\alpha x = \{\alpha\xi_1, \ldots, \alpha\xi_n\},$$
$$0 = \{0, \ldots, 0\}$$
$$-x = \{-\xi_1, \ldots, -\xi_n\}.$$

It is easy to verify that all parts of our two axioms (A) and (B), §1, are satisfied, so that $\mathfrak{C}_n$ is a complex vector space; it is usually called n-dimensional complex Euclidean space.

(4) For any positive integer n let $\mathfrak{P}_n$ be the set of all polynomials (with the same restrictions as in (2)) of degree $\leq$ n-1, together with the polynomial identically zero. (In the usual discussion of degree the degree of this polynomial is not defined, so that we cannot say that it has degree $\leq$ n-1.) With the same interpretation of the linear operations (addition and scalar multiplication) as in (2), $\mathfrak{P}_n$ is a complex vector space.

(5) A close relative of $\mathfrak{C}_n$ is the set $\mathfrak{R}_n$ of all n-tuples of real numbers, $x = \{\xi_1, \ldots, \xi_n\}$. With the same formal definitions of addition and scalar multiplication as for $\mathfrak{C}_n$, excepting that we consider only real scalars α, the space $\mathfrak{R}_n$, the ordinary or real n-dimensional Euclidean space, is a real vector space.

§3. COMMENTS ON NOTATION AND TERMINOLOGY

A few comments on our axioms and notation. Those familiar with algebraic terminology will have recognized the axioms (A), §1, as the defining conditions of an abelian (commutative) group; the axioms (B) express the fact that the group admits scalars as operators. We use the "scalar" terminology to emphasize the fact that we are not even necessarily dealing with real or complex numbers. Ninety percent of the theory remains valid if we interpret scalars as elements of any field. If scalars are elements of a ring a vector space is sometimes called a <u>modul</u>.

Special real vector spaces (namely $\mathfrak{R}_n$) are familiar in geometry. There seems at this stage to be no excuse for the apparently uninteresting insistence on complex numbers. We hope that reader is willing to take it on faith that we shall have to make use of deep properties of complex numbers later, (conjugation, algebraic closure), and that in both the applications of vector spaces to modern (quantum mechanical) physics and in the mathematical generalization of our results to Hilbert space, complex numbers play an important role. Their one great disadvantage is the difficulty of drawing pictures: the ordinary picture (Argand diagram) of $\mathfrak{C}_1$ is indistinguishable from that of $\mathfrak{R}_2$, and a graphic representation of $\mathfrak{C}_2$ seems to be out of human reach. On occasions when we have to use pictorial language we shall therefore use the terminology of $\mathfrak{R}_n$ in $\mathfrak{C}_n$, and speak of $\mathfrak{C}_2$, for example, as a plane.

Finally we comment on notation. We observe that the symbol 0 has been used in two meanings: once as a number and once as a vector. To make the situation worse we shall later, when we introduce linear functionals and linear transformations, give it still other meanings. Fortunately the relation between the various interpretations of 0 is such that after this word of warning no confusion should arise from this practice. Another notationally happy circumstance is that $-x$ (defined in §1, (A)(4)) and $(-1)x$ are the same thing. This is true since

$$x + (-1)x = 1x + (-1)x = (1 + (-1))x = 0x = 0.$$

§4. DEFINITION OF LINEAR DEPENDENCE

Now that we have described the spaces we shall work with we must specify the relations among the elements of these spaces which will be of interest to us. Vector spaces are used to study linear problems: the general form of a linear relation is described in the following definition.

> DEFINITION. A *finite* set of vectors, $x_1, \ldots, x_n$, is <u>linearly dependent</u> if there exist scalars $\alpha_1, \ldots, \alpha_n$, not all zero, such that
> $\sum_i \alpha_i x_i = \alpha_1 x_1 + \ldots + \alpha_n x_n = 0.$
> If, on the other hand, $\sum_i \alpha_i x_i = 0$ implies that $\alpha_1 = \cdots = \alpha_n = 0$, the vectors $x_1, \ldots, x_n$ are <u>linearly independent</u>.

Linear dependence or independence are properties of <u>sets</u> of vectors; it is customary however to apply these adjectives to vectors themselves and thus we shall sometimes say "a set of linearly independent vectors" instead of "a linearly independent set of vectors". It will be convenient also to speak of the linear dependence or independence of a not necessarily finite set, $\mathfrak{X}$, of vectors. We shall say that $\mathfrak{X}$ is linearly independent if every finite subset of $\mathfrak{X}$ is such; otherwise $\mathfrak{X}$ is linearly dependent.

To gain insight into the meaning of linear dependence let us study the examples of vector spaces that we already have.

(1) If x and y are any two vectors in $\mathfrak{C}_1$, then x and y form a linearly dependent set. If $x = y = 0$ this is trivial; if not we have, for example, the relation $yx + (-x)y = 0$. Since it is clear that any set containing a linearly dependent subset is itself linearly dependent, this shows that any set containing more than one element is a linearly dependent set.

(2) More interesting is the situation in the space $\mathfrak{P}$. The vectors $x = x(t) = 1 - t$, $y = y(t) = t(1-t)$, and $z = z(t) = 1-t^2$ are, for example, linearly dependent, since $x + y - z = 0$. However the infinite set of vectors $x_0, x_1, x_2, \ldots$ defined by

$x_0(t) = 1$, $x_1(t) = t$, $x_2(t) = t^2$, $x_3(t) = t^3$, ...,

is a linearly independent set, for if we had any relation of the form

$$\alpha_0 x_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n = 0,$$

then we should have a polynomial relation

$$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n \equiv 0,$$

whence $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$

(3) As we mentioned before, the spaces $\mathfrak{C}_n$ are the prototype of what we want to study: let us examine, for example, the case $n = 3$. To those familiar with higher dimensional geometry the notion of linear dependence in this space (or, more properly speaking, in its real analog $\mathfrak{R}_3$) has a concrete geometric meaning which we shall only mention. In geometrical language: two vectors are linearly dependent if and only if they are collinear with the origin, and three vectors are linearly dependent if and only if they are coplanar with the origin. (If one thinks of a vector not as a point in a space but as an arrow pointing from the origin to some given point, the preceding sentence should be modified by crossing out the phrase "with the origin" both times that it occurs). We shall presently introduce the notion

of linear manifolds (or vector subspaces) in a vector space and use the geometrical language thereby suggested.

For a concrete example consider the vectors $x = \{1, 0, 0\}$, $y = \{0, 1, 0\}$ $z = \{0, 0, 1\}$, and $u = \{1, 1, 1\}$. These four vectors form a linearly dependent set, since $x + y + z - u = 0$; it is easy to verify however that any three of these vectors form a linearly independent set.

§5. CHARACTERIZATION OF LINEAR DEPENDENCE

Returning to the general considerations we shall say, whenever $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$, that x is a <u>linear combination</u> of $x_1, \ldots, x_n$; we shall use without any further explanation all the simple grammatical implications of our terminology. Thus we shall say, in case x is a linear combination of $x_1, \ldots, x_n$, that x is linearly dependent on $x_1, \ldots, x_n$; we shall leave to the reader the proof of the fact that if $x_1, \ldots, x_n$ are linearly independent then x is a linear combination of them if and only if the vectors $x, x_1, \ldots, x_n$ are linearly dependent.

The fundamental theorem concerning linear dependence is the following.

THEOREM. The set of non zero vectors, $x_1, \ldots, x_n$, is linearly dependent if and only if some x_k, $2 \leq k \leq n$, is a linear combination of the preceding ones.

PROOF. Let us suppose that the set $x_1, \ldots, x_n$ is linearly dependent and let k be the first integer between 2 and n for which $x_1, \ldots, x_k$ is linearly dependent. (If worse comes to worst the hypothesis of the theorem assures us that $k = n$ will do). Then

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = 0$$

for a suitable set of α's; moreover, whatever the α's,

$\alpha_k = 0$ is impossible, for then we should have a linear dependence relation among $x_1, \ldots, x_{k-1}$, contrary to the definition of k. Hence

$$x_k = \left[\frac{-\alpha_1}{\alpha_k}\right] x_1 + \cdots + \left[\frac{-\alpha_{k-1}}{\alpha_k}\right] x_{k-1},$$

as was to be proved. This proves the necessity of our condition; sufficiency is clear since, as we remarked before, any set containing a linearly dependent set is itself such.

§6. DEFINITION AND CONSTRUCTION OF BASES

DEFINITION. A (linear) basis (or a coordinate system) in a vector space $\mathfrak{V}$ is a set $\mathfrak{X}$ of linearly independent vectors such that every vector in $\mathfrak{V}$ is a linear combination of elements of $\mathfrak{X}$. A vector space $\mathfrak{V}$ is finite dimensional if it has a finite basis.

Except for the occasional consideration of examples we shall restrict our attention, throughout this book, to finite dimensional vector spaces.

For examples of bases and finite and non finite dimensional spaces we turn again to the spaces $\mathfrak{C}_n$ and $\mathfrak{P}$. In $\mathfrak{P}$ the set $x_n = x_n(t) = t^n$, $n = 0, 1, 2, \ldots$ is a basis: every polynomial is, by definition, a linear combination of a finite number of x_n. Moreover $\mathfrak{P}$ has no finite basis, for given any finite set of polynomials we may find a polynomial of higher degree than any of them: this latter polynomial is obviously not a linear combination of the former ones.

An example of a basis in $\mathfrak{C}_n$ is the set of vectors x_i, $i = 1, \ldots, n$, defined by the condition that the j-th coordinate of x_i is δ_{ij}. (Here we use for the first time the popular Kronecker δ; it is defined by $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$). Thus we assert that in $\mathfrak{C}_3$ the vectors $x_1 = \{1, 0, 0\}$, $x_2 = \{0, 1, 0\}$, and $x_3 = \{0, 0, 1\}$ are a basis. We have seen before that they are linearly independent; the formula

$$x = \{\xi_1, \xi_2, \xi_3\} = \xi_1 x_1 + \xi_2 x_2 + \xi_3 x_3$$

proves that every x in $\mathfrak{C}_3$ is a linear combination of them.

In a general finite dimensional vector space $\mathfrak{V}$, with basis $x_1, \ldots, x_n$, we know that every x may be written in the form

$$x = \sum_i \xi_i x_i;$$

we assert that the ξ's are uniquely determined by x. The proof of this assertion is an argument often used in the theory of linear dependence. If we had $x = \sum_i \eta_i x_i$, then we should have, by subtraction,

$$\sum_i (\xi_i - \eta_i) x_i = 0.$$

Since the x's are linearly independent, this implies that $\xi_i - \eta_i = 0$ for $i = 1, \ldots, n$: in other words the η's are the same as the ξ's.

THEOREM. If $\mathfrak{V}$ is a finite dimensional vector space and $y_1, \ldots, y_m$ is any set of linearly independent vectors in $\mathfrak{V}$, then, unless the y's already form a basis, we can find vectors $y_{m+1}, \ldots, y_{m+p}$ so that the totality of y's, $y_1, y_2, \ldots, y_m, y_{m+1}, \ldots, y_{m+p}$, forms a basis. In other words: every linearly independent set can be extended to a basis.

PROOF. Since $\mathfrak{V}$ is finite dimensional it has a finite basis, say $x_1, \ldots, x_n$. We consider the set $\mathfrak{S}$ of vectors

$$y_1, \ldots, y_m, x_1, \ldots, x_n,$$