LINEAR VECTOR SPACES AND CARTESIAN TENSORS

JAMES K. KNOWLES
California Institute of Technology

New York Oxford
Oxford University Press
1998

Oxford University Press
Oxford New York
Athens Auckland Bangkok Bogotá Bombay Buenos Aires
Calcutta Cape Town Dar es Salaam Delhi Florence Hong Kong
Istanbul Karachi Kuala Lumpur Madras Madrid Melbourne
Mexico City Nairobi Paris Singapore Taipei Tokyo Toronto

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.,
198 Madison Avenue, New York, New York, 10016
http://www.oup-usa.org
1-800-334-4249

Library of Congress Cataloging-in-Publication Data

Knowles, James K. (James Kenyon), 1931–
Linear vector spaces and Cartesian tensors / James K. Knowles.
p. cm.
Includes bibliographical references and index.
ISBN 0-19-511254-7 (cloth)
1. Vector spaces 2. Calculus of tensors. I. Title.
QA186.K657 1997
512′.52—dc21 97-8936
CIP

Printing (last digit): 9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

To Jackie, John, Jeff, and Jamey

TABLE OF CONTENTS

PREFACE

This book represents my approach to some applicable material connected with linear vector spaces that has been part of a course offered in the Division of Engineering and Applied Science at Caltech for many years. The course—called AM125 in the local code—covers a full academic year, which is divided into three quarters, with this material offered in the first quarter. AM125 is usually taught by different instructors in different years, and the material related to vector spaces looks quite different from one instructor to another.

The typical consumer in AM125 is a first-year graduate student, usually from a field in engineering or applied science, but sometimes from geology and occasionally from chemistry or physics. Students in many of the doctoral programs within the Division of Engineering and Applied Science are *required* by their elders to jump through this particular mathematical hoop. Although the backgrounds of the students vary widely, almost all have had something called "linear algebra" in their undergraduate programs, in addition to basic calculus, differential equations, and often complex variables. All of the takers have contemporaneous sustained exposure to serious science or engineering. As a result, most are quite well equipped to appreciate the potential utility of the material presented in AM125.

Usually, a sizable majority of the students who take AM125 are or will be pursuing doctoral research in fields that make heavy use of some aspect of mechanics, often the continuum mechanics of solids or fluids. The slant that I choose to put on the material, especially that covered in the first quarter, is intended to make the subject particularly helpful to this group, though not—I hope—to the detriment of others.

It has been my experience that the weapons brought by the students to AM125 from their earlier experience in linear algebra are for the most part limited to those that help in doing calculations with matrices, especially big ones. Every student can calculate with astonishing accuracy something he or she will call an eigenvalue, but frequently the understanding of what an eigenvalue actually *means* is dim or absent altogether. The geometric content of the theory of vector spaces is often missing from the student's knowledge, and notions of invariance, so important in continuum mechanics, for example, evidently have been given little attention. Most students do not appreciate the fundamental distinctions between vector spaces with real scalars and those whose scalars are complex.

It is my hope that this particular version of the theory and application of finite-dimensional vector spaces will provide remedies for the specific illnesses cited above, while also offering an interesting and useful look at a subject of great beauty and utility. Little is done here with infinite-dimensional spaces, though I have touched upon them from time to time. The basic material on finite-dimensional linear spaces is covered in the first three chapters. Chapter 4 introduces the subject of Cartesian tensors of rank four, partly because of their importance in continuum mechanics, but also as an application of the ideas to the construction of another useful mathematical device. Some applications that lie closer to real physical issues are covered in the final chapter. There are problems of varying degrees of difficulty at the end of each chapter. Solutions to most of these problems are sketched in Appendix 2, though solutions are not given for the very easiest problems, nor for those in which the student is given detailed instructions in the statement of the problem itself. Appendix 1 contains a brief list of some results from elementary algebra that are taken for granted throughout the book. A short list of relevant references is included at the end of each chapter.

I teach slowly. I have never covered literally *all* of the material presented here in one academic quarter, but I have no doubt that others could readily do so. It would be easy to do—even for me—in a semester.

A word about undergraduates: Although there are usually few undergraduates in AM125 at Caltech, this is by no means because they are unprepared for the mathematical material and its applications. Indeed, this book could serve as the text for an advanced undergraduate course, especially for students with some exposure to courses in science or engineering.

There are, of course, many fine books on the subject treated here. I must especially acknowledge two that I admire greatly and to which I have repeatedly turned for illumination for many years: One is *Lectures on Linear Algebra*, by I.M. Gel'fand, the other *Finite Dimensional Vector Spaces*, by P.R. Halmos. I owe both of these unique resources a great debt of gratitude.

A word of thanks to the many years worth of Caltech students who have taken AM125 either as conscripts or as volunteers: Their efforts, comments and criticisms have led to many revisions of the presentation and the problems. I hope at least some of them have had as much fun confronting the material as I have had in delivering it!

It is pleasant to acknowledge the countless conversations I have had over the years with colleagues at Caltech and elsewhere concerning the subject addressed here and matters closely related to it. I wish particularly to thank Jim Beck, Tom Caughey, and Steve Wiggins of Caltech, Rohan Abeyaratne of M.I.T., Mort Gurtin of Carnegie-Mellon University, Niall Horgan of the

University of Virginia, Phoebus Rosakis of Cornell University, and Dick Shield of Caltech and the University of Illinois for many interactions that contributed enormously to my understanding and appreciation of the issues. Finally, I must express my very special gratitude to the late Eli Sternberg of Caltech for years of exchanges concerning this subject and many others. Although Eli, alas, never taught AM125, his interest in linear spaces was intense, and he was a master of their applications in solid mechanics. I profited greatly from our joint ruminations over twenty-five years.

I am also pleased to acknowledge the efficient help and welcome encouragement that I have received from Mr. Bill Zobrist, who is my editor at Oxford University Press.

C H A P T E R

LINEAR VECTOR SPACES

When we are young, we learn that vectors are arrows. We are told how to add arrows by the parallelogram law, how to multiply arrows by numbers, including negative ones, and how to measure the length of an arrow or the angle between two of them. We are shown how to rotate arrows, stretch them, and reflect them. These operations have to do with the arithmetic of arrows, the metric properties of arrows, and the transformation of arrows into other arrows according to some preset rule, such as rotation through an angle of 45° about a given axis. At this later stage of life, we might observe that the geometric definitions of these operations and the attendant calculations do not involve the use of "coordinate axes" along which to "resolve" the arrows into their "components." The significance of this observation will unfold as we go along.

The collection A_2 of all arrows in a plane with their tails at a fixed point (the origin), together with the real numbers we need to multiply arrows by, represents the prototype of the useful abstract notion of a linear vector space. We can also deal, of course, with the set A_3 of all arrows in ordinary *three-*dimensional space rather than the set of arrows confined to a plane. In any event, it is linear vector spaces that we shall explore in this book. Along the way, we shall encounter some applications of the theory, mainly to that most venerable of physical disciplines, *mechanics*.

A *linear vector space* R is a collection of things called *vectors* (denoted by bold face letters like $\mathbf{x}$, $\mathbf{y}$, $\mathbf{u}$, $\mathbf{v}$, ...), together with some things called scalars (usually designated by symbols like α, β, a, b, ...), that jointly conform to certain rules that have been abstracted from the story of arrows. These rules can be conveniently grouped in two sets:

I. Given any two vectors $\mathbf{x}$ and $\mathbf{y}$ in R, there is a third vector $\mathbf{z}$ in R, called the sum of $\mathbf{x}$ and $\mathbf{y}$ and written $\mathbf{z} = \mathbf{x} + \mathbf{y}$, that obeys the following rules:

(a) addition is commutative ($\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$) and associative ($(\mathbf{x} + \mathbf{y}) + \mathbf{w} = \mathbf{x} + (\mathbf{y} + \mathbf{w})$), etc.;

(b) there is a unique vector $\mathbf{o}$ in R (called the *null* vector) such that $\mathbf{x} + \mathbf{o} = \mathbf{x}$;

(c) for every vector $\mathbf{x}$ in R, there is a unique vector called $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{o}$.

II. Given any vector $\mathbf{x}$ and any scalar α, there is a vector $\mathbf{z}$ called the product of $\mathbf{x}$ with α and written $\mathbf{z} = \alpha\mathbf{x}$, that obeys the following rules:

(a) multiplication by a scalar is distributive in the sense that $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ and $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$;

(b) there are scalars called 0 and 1 such that $0\mathbf{x} = \mathbf{o}$, $1\mathbf{x} = \mathbf{x}$.

Although more general choices are possible, the objects referred to as *scalars* above are usually either real numbers or complex numbers. Our interest will ultimately lie in *real* vector spaces, i.e., spaces with real numbers as scalars, but we will also need *complex* vector spaces (spaces whose scalars are complex numbers) from time to time, especially for purposes of contrast.

The arrows in A_2 or A_3 comprise a real vector space; if $\mathbf{x}$ is an arrow, $3\mathbf{x}$ is the arrow three times as long as $\mathbf{x}$ whose direction is that of $\mathbf{x}$, and $(-1)\mathbf{x} \equiv -\mathbf{x}$ is the arrow directed exactly opposite to $\mathbf{x}$ whose length coincides with that of $\mathbf{x}$. You may find it helpful to verify that the spaces of arrows A_2 and A_3 conform to the rules for a linear vector space laid out above.

Here are some other examples of vector spaces.

Example 1.1. *The space of columns of real numbers.* A much-loved vector space is the collection $R = R_n$ of all columns $\mathbf{x}$ of n real numbers, where n is a positive integer; the scalars are real numbers. Thus a typical element of R_n is of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}, \tag{1.1}$$

where $x_1, x_2, \ldots, x_n$ are real numbers. If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in R_n, their sum is defined to be the column whose entries are the sums of corresponding entries in $\mathbf{x}$ and $\mathbf{y}$ (i.e., $x_1 + y_1$, $x_2 + y_2$, etc.). Similarly, $\alpha\mathbf{x}$ is the column with entries $\alpha x_1, \alpha x_2, \ldots, \alpha x_n$. The null vector $\mathbf{o}$ is the column with every entry zero, and $-\mathbf{x}$ is the column in which the generic entry is $-x_k$. Under these operations, the collection R_n is readily shown to obey the rules for a vector space set out above. When we use R_n in the future, we shall refer to the operations of addition of vectors and multiplication of a vector by a scalar as specified here as the "natural" operations on columns.

Because Descartes explained in the seventeenth century that a plane can be usefully described in terms of the "coordinates" of each of its points—say x_1, x_2—with respect to a set of rectilinear axes, the special case R_2 of the vector space in Example 1.1 is particularly valuable. So of course is R_3, which describes the space we live in through the coordinates x_1, x_2, x_3.

Example 1.2. *The space of columns of complex numbers.* An equally good example of a vector space is the set C_n of all columns of n *complex* numbers $z_1, z_2, ..., z_n$, with the complex numbers serving as scalars. The definitions of the operations of addition and multiplication by a scalar are analogous to those used in the invention of R_n in Example 1.1.

Example 1.3. *The space of continuous real-valued functions.* Let C be the collection of all real-valued functions that are defined and continuous on the closed interval $[0, \pi]$. If f and g are two such functions, we say their sum $h = f + g$ is the real valued function defined by $h(t) = f(t) + g(t)$ for every t such that $0 \leq t \leq \pi$, and if α is a real number, $h = \alpha f$ is defined by $h(t) = \alpha f(t)$, $t \epsilon [0, \pi]$. C is a vector space under these "natural" operations.

Example 1.4. *Solutions of a linear ordinary differential equation.* Consider the collection of all twice continuously differentiable, real-valued functions φ on the interval $[0, \pi]$ that satisfy the linear ordinary differential equation

$$\varphi'' + \varphi = 0 \quad \text{on } [0, \pi]. \tag{1.2}$$

Let the scalars be real numbers, and define the operations of addition and multiplication by a scalar exactly as in the preceding example. Because the sum of two twice continuously differentiable functions is also twice continuously differentiable, because the sum of two solutions of (1.2) is also a solution, and because any real multiple of a solution is a solution, this collection is also a linear vector space. Observe that every vector in this vector space is also in the space C of the preceding example. Problem 1.1 explores how things would change if the zero on the right side of (1.2) were replaced by one.

Example 1.5. *A space of polynomials.* If N is a non-negative integer and $n = N + 1$, denote by P_n the set containing every polynomial with real coefficients whose degree is not greater than N. If $\mathbf{p}$ and $\mathbf{q}$ are two polynomials in P_n, define $\mathbf{p} + \mathbf{q}$ and $\alpha\mathbf{p}$, where α is real, in the natural way. P_n then becomes a real vector space. Note that P_n is contained in the space C of Example 1.3.

The first notions of importance in the study of vector spaces are the complementary ones of linear dependence and linear independence. Let R be a linear vector space, and let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ be a set of k vectors in R, where k is a positive integer. This set is said to be *linearly dependent* if there are k scalars $\alpha_1, \alpha_2, ..., \alpha_k$, *not all zero*, such that

$$\alpha_1\mathbf{x}_1 + \alpha_2\,\mathbf{x}_2 + ... + \alpha_k\,\mathbf{x}_k = \mathbf{o}. \tag{1.3}$$

If the only set of scalars for which (1.3) holds is the set $\alpha_1 = \alpha_2 = ... = \alpha_k = 0$, the set of k vectors $\mathbf{x}_1, ..., \mathbf{x}_k$ is said to be *linearly independent*. Note that linear dependence and independence are properties of *sets* of vectors, not of individual vectors themselves. The set $\{\mathbf{x}\}$ containing the single vector $\mathbf{x}$ is a linearly independent set unless $\mathbf{x} = \mathbf{o}$, in which case it is linearly dependent. Any set of k vectors that includes $\mathbf{o}$ is a linearly dependent set; see Problem 1.3.

In the vector space A_2 of arrows in a plane, any set of two arrows that are parallel (or anti-parallel) is a linearly dependent set. Any set of two arrows that are both of non-zero length and neither parallel nor anti-parallel is linearly independent. Any set of three or more arrows is linearly dependent; see Problem 1.4.

In the vector space of Example 1.4 above, $\varphi_1(t) = \sin t$ and $\varphi_2(t) = \cos t$ are vectors in the space, and together they clearly constitute a linearly independent set. The theory of linear ordinary differential equations tells us that *every* solution of (1.2) is expressible as a linear combination of these two special solutions. This means that any set of *three* elements of the vector space of solutions of (1.2) that contains both φ_1 and φ_2 is a linearly *dependent* set. Indeed, as discussed in Problem 1.5, the theory of differential equations tells us more than this: *any* set of three solutions of (1.2) is a linearly dependent set.

The vector space of Example 1.3 is richer. Let k be *any* positive integer. For every positive integer k, the set of continuous functions $\varphi_1(t) = 1$, $\varphi_2(t) = t$, $\varphi_3(t) = t^2$, ..., $\varphi_{k+1}(t) = t^k$ is a linearly independent set, as will be shown in Problem 1.6. Thus in this vector space, linearly independent sets can be as big as one likes, in the sense that such a set can contain an arbitrarily large number of vectors.

If a vector space R contains a linearly independent set of n vectors but contains *no* linearly independent set of n + 1 vectors, where n is a positive integer, then R is said to have dimension n. The set $\{\mathbf{o}\}$ consisting of the null vector alone is trivially a vector space; it fails to contain a linearly independent set of n vectors for *any* positive integer n, and is therefore said to have dimension zero. In contrast, there are vector spaces that contain linearly independent sets of n vectors for *every* positive integer n; such spaces are said to be infinite dimensional. For the most part, we shall be concerned with finite-dimensional spaces.

Let R be a vector space of finite dimension $n \geq 1$. A linearly independent set of n vectors is called a *basis* for R. Let $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ be a basis for R, and let $\mathbf{x}$ be an arbitrary vector in R. Then, by the definition of dimension, the set $\{\mathbf{x}, \mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ is linearly dependent, so there are n + 1 scalars $\alpha_0, \alpha_1, ..., \alpha_n$, not all zero, such that $\alpha_0\mathbf{x} + \alpha_1\mathbf{e}_1 + \alpha_2\,\mathbf{e}_2 + ... + \alpha_n\,\mathbf{e}_n = \mathbf{o}$. Clearly α_0 cannot vanish, or the **e**s would comprise a linearly dependent set. For k = 1, ..., n, put $\xi_k = -\alpha_k/\alpha_0$. Then

$$\mathbf{x} = \xi_1\,\mathbf{e}_1 + \xi_2\,\mathbf{e}_2 + ... + \xi_n\,\mathbf{e}_n. \tag{1.4}$$

This shows that, for every $\mathbf{x}$ in R, there are scalars $\xi_1, \xi_2, ..., \xi_n$ such that $\mathbf{x}$ may be represented in terms of the basis $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ by (1.4). The linear independence of the set of basis vectors assures that, for the given $\mathbf{x}$, the scalars ξ_i in (1.4) are unique, as will be shown in Problem 1.7. They are called the *components* of $\mathbf{x}$ *in the basis* $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$, or *in the basis* $\mathbf{e}$, for brevity. The components of a vector are both useful and dangerous. Their utility stems from the fact that, as we shall see often enough, using components is frequently helpful in making calculations. The potential danger in appealing to components arises because they depend not only on the vector being represented, but also on the choice of basis. Thus when a result is established by a calculation involving components, one may question whether the result is basis-dependent. Because basis-*in*dependent results are crucial in many physical applications of our subject, one must keep this issue in mind as we proceed.

One further depressing note: the discussion in the preceding paragraph fails to explain how to *find* the components of a given vector. We shall address this question later.

Back to the space A_2 of arrows in a plane. Since according to Problem 1.4, any three arrows in A_2 are linearly dependent, the dimension of A_2 cannot exceed 2. But if $\mathbf{e}_1$ and $\mathbf{e}_2$ are two perpendicular arrows of non-zero length, neither is a scalar multiple of the other, so together they comprise a linearly independent set, and—since there cannot be three arrows with this property—they form a basis. Suppose in addition that $\mathbf{e}_1$ and $\mathbf{e}_2$ are perpendicular arrows, each with unit length, and let $\mathbf{x}$ be any arrow in A_2. If θ is the angle between $\mathbf{x}$ and $\mathbf{e}_1$ and $|\mathbf{x}|$ is the length of the arrow $\mathbf{x}$, some trigonometry shows that $\mathbf{x} = |\mathbf{x}| \cos \theta \, \mathbf{e}_1 + |\mathbf{x}| \sin \theta \, \mathbf{e}_2$, so that $\xi_1 = |\mathbf{x}| \cos \theta$ and $\xi_2 = |\mathbf{x}| \sin \theta$ are the components of the arrow $\mathbf{x}$ in this particular basis.

As a second illustration, consider the space R_2 of columns of two real numbers. It is again easy to show that any three vectors in R_2 are linearly dependent. Moreover, if

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{1.5}$$

is any vector in R_2, and

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.5a}$$

clearly $\mathbf{e}_1$ and $\mathbf{e}_2$ form a linearly independent set and therefore a basis (often called the *natural* basis for R_2). Furthermore, $\mathbf{x} = x_1 \, \mathbf{e}_1 + x_2 \, \mathbf{e}_2$, so that $\xi_1 = x_1$ and $\xi_2 = x_2$ are the components of $\mathbf{x}$ in this basis. This discussion has an easy generalization to R_n, for any positive integer n.

Our earlier remarks lead us to conclude that the dimension of the vector space of solutions of the differential equation (1.2) in Example 1.4 is 2, and that cos t and sin t form a basis.

For the vector space P_n of Example 1.5, the special polynomials 1, t, t^2, ..., t^N are linearly independent and form a basis; the dimension of P_n is $n = N + 1$. Problem 1.8 addresses the question of how to find the components in this basis of a polynomial $\mathbf{p}$ in P_n.

Example 1.3 provides an example of an infinite-dimensional vector space; since the definition of basis given above applies only to finite-dimensional spaces, we cannot speak of this notion for the space of Example 1.3 without extending the idea in some suitable way. This task presents some troublesome technical problems that would take us well off our intended path, so we avoid this issue and suggest that the reader who is especially interested in infinite-dimensional spaces should consult the references by Halmos [1.4], Kolmogorov and Fomin [1.5], Debnath and Mikusinski [1.1], and Naylor and Sell [1.6] listed at the end of this chapter.

A question that will arise from time to time in what follows concerns the relationship between the components of a vector in two bases. Let $\mathbf{e}_1, ..., \mathbf{e}_n$ and $\mathbf{f}_1, ..., \mathbf{f}_n$ be two bases for a vector space R of finite dimension n, and let $\mathbf{x}$ be a given vector in R. Then we may write

$$\mathbf{x} = \sum_{k=1}^{n} \xi_k \, \mathbf{e}_k = \sum_{k=1}^{n} \eta_k \, \mathbf{f}_k, \tag{1.6}$$

in terms of the two sets of components ξ_k and η_k. But since the vectors $\mathbf{f}_k$

form a basis, each of the vectors $\mathbf{e}_k$ may be expressed as follows:

$$\mathbf{e}_k = \sum_{j=1}^{n} \rho_{jk}\, \mathbf{f}_j \tag{1.7}$$

for some scalars ρ_{jk}; here, for each k = 1, 2, ..., n, ρ_{1k}, ρ_{2k}, ..., ρ_{nk} are the components of the vector $\mathbf{e}_k$ in the basis **f**. Expressing $\mathbf{e}_k$ in (1.6) in terms of the $\mathbf{f}_j$s by means of (1.7) gives

$$\sum_{k=1}^{n} \xi_k \left(\sum_{j=1}^{n} \rho_{jk}\mathbf{f}_j \right) = \sum_{k=1}^{n} \eta_k\, \mathbf{f}_k = \sum_{j=1}^{n} \eta_j\, \mathbf{f}_j; \tag{1.8}$$

the last equality comes about because of the purely cosmetic step of replacing the summation index k by j in the central expression in (1.8). Since the **f**s are linearly independent, the coefficients of $\mathbf{f}_j$ in the extreme members of (1.8) must coincide, yielding

$$\eta_j = \sum_{k=1}^{n} \rho_{jk}\, \xi_k,\ j = 1, ..., n. \tag{1.9}$$

This "change-of-basis" formula expresses the components η_j of **x** in the basis **f** in terms of the components ξ_k of **x** in the basis **e** with the help of the n^2 scalars ρ_{jk} that determine the relation (1.7) between the two bases.

Calculations with components of vectors often lead to repulsive expressions like those appearing in (1.8), or worse. There is a notational abbreviation that helps. Inspection of (1.6)–(1.9) reveals that the subscript subject to summation always occurs twice in the summand. This suggests that we *delete* the summation sign Σ from the equations, and agree to sum *automatically* over the repeated index. With this *summation convention* in force, (1.9) would be replaced by

$$\eta_j = \rho_{jk}\, \xi_k,\ j = 1, ..., n. \tag{1.10}$$

To make this foolproof, we must understand that the summation automatically extends over the *range* 1, 2, ..., n of possible values of the subscripts. In both (1.9) and (1.10), the subscripts j and k are different in character: the index k that is repeated on one side of the equation is "summed out" and could be replaced by any other symbol (m or p, for example), while the subscript j—called the "free index"—takes successively the values indicated at the end of the equation. We can extend our notational agreement by adopting the *range convention*: the free index is *assumed* to take the values 1, ..., n