LINEAR ALGEBRA

LINEAR ALGEBRA

Burton W. Jones

The Mathematics Department
The University of Colorado

HOLDEN-DAY, INC.

San Francisco

Düsseldorf Johannesburg London
Panama Singapore Sydney Toronto

LINEAR ALGEBRA

500 Sansome Street, San Francisco, California 94111

Library of Congress Catalog Card Number: 72-83244
ISBN: 0-8162-4544-4

Printed in the United States of America

2 3 4 5 6 7 8 9 0 M P 8 0 7 9 8 7 6 5 4 3

PREFACE

This is an introductory text in linear algebra for students who have had experience with beginning college mathematics and have some acquaintance with proof, including a little mathematical induction. No previous experience with linear algebra or matrices is assumed, but it would be helpful if the student had some familiarity with the ideas of a field and a group. However, for those without such information, appendixes are provided which review, with some proofs, the salient ideas needed. The teacher who wishes to confine attention to the fields of real and complex numbers can do so by exercising a little care. However, I feel that the student should gradually progress toward thinking in terms of a general field, if only from the point of view that the field postulates really are a formulation of the rules of manipulation.

In the process of developing the important definitions and results of linear algebra, I have tried to demonstrate to the student what more abstract mathematics is like. Thus motivation is of fundamental importance. A set of postulates should arise from experience. The student should be told where we are going and why. Interrelationships should be stressed. Terminology should be introduced only when it is about to be used. It is important that the student know the pedigree of a concept so that he may become a partner in the development.

The exercises are, of course, a very important part of the book. A student can test his knowledge first by working through the routine numerical ones. He is then given the opportunity of proving some of the theorems and extending the theory so that he may experience the joy of being a mathematician and have an important part in the development of the subject. On occasion, exercises lead into material which is to follow. However, important results whose proofs are left as exercises are stated in the body of the text for continuity and easy reference.

The first three chapters constitute the basic material on vector spaces, matrices, and linear transformations. We begin with a short preview of where we are going and why. Then we trace the evolution of the idea of a vector from that of a physical "arrow" to an n-tuple of numbers. We next codify the properties of an n-tuple and, on the basis of this experience, set up an abstract definition of a vector space; at the same time showing the advantages of this point of view. The abstract definition is basis free. In fact, polynomials in a single variable are shown to constitute an infinite-dimensional vector space. However, though the abstract definition is extensively used, almost all the spaces dealt with in this book are finite dimensional.

The second chapter introduces matrices through their use in connection with sets of linear equations. The development of the echelon form is shown to be a natural abbreviation of the usual method of solution by elimination. Cues for the definitions of the operations on matrices are provided by our experience with vectors and by applications we should like to make.

The first nine sections of Chapter 3 deal with linear transformations without the use of matrices or determinants; this is to stress their inherent properties without being matrix bound. However, we reach a point, in Section 3.10, where matrices are needed to give us information about transformations. (Certainly it would be difficult to define the trace of a transformation without a matrix.) Then, too, from this point on, the interplay between matrices and linear transformations adds enrichment to both developments. Very fundamental, too, is the idea of certain invariants of a matrix or transformation under change of basis. In this connection, I have found that one gains conceptually by postponing the use of determinants. So without determinants it is shown that the minimum polynomial of a transformation of a vector space into itself has degree not greater than the dimension of the space. The proof is a constructive one and paves the way for the general theorems on the Jordan normal form in Chapter 6.

In the fourth chapter, determinants are developed to introduce the characteristic equation, prove the Cayley-Hamilton theorem, and give a more efficient means of calculating characteristic roots. No previous knowledge of determinants is assumed. Indeed the approach is essentially that of Weierstrass; that is, we decide what we want a determinant to do and how we wish it to behave, and then we define it so that it does and behaves as we want. This not only has the advantage of motivation, but also results in some economy of development.

The fifth chapter opens with the idea of a dot product of vectors and, after some development of its properties, leads into the more abstract concept of a vector product function. Here progress is from the general bilinear form, through the idea of a nonsingular form, to the inner product space and quadratic forms. Connections with and applications to Euclidean transformations are stressed to the point where it is shown that any Euclidean transformation can be represented as a product of symmetries. Dual spaces are not introduced as such, though they are inherent in the development of the adjoint transformation. It seemed that without the prospect of dealing more systematically with duals, there was not much point in introducing them explicitly.

In the sixth chapter, the Jordan form is developed. It begins with a review of some results already obtained and, for motivation purposes, a problem in stochastic matrices of order 2. Then, after the general form is developed, applications are given to a method of computing the characteristic polynomial, to differential equations for those with some experience in that subject, and to positive stochastic matrices.

The first three appendixes constitute material on groups, fields, and polynomials which is used in the book but which some students may not already know. Also, in the case of permutation groups, relegation to an appendix avoids breaking the continuity of the development of determinants. The fourth appendix is a curious and interesting result from Artin's *Geometric Algebra.*

From time to time throughout the book some ideas of projective geometry are introduced not only for historical reasons but because the applications are especially close. However, no prior knowledge of projective geometry is assumed.

For students without previous experience with linear algebra, the first three chapters should provide enough for a semester's course, meeting three times a week. Not much of this material could be omitted except perhaps for affine transformations and projective geometry. On the other hand, for students with some previous knowledge of the subject, one should be able to cover much of the material in Chapter 5 in a semester's course, by going lightly over some of the chapter on determinants.

The level of sophistication rises throughout the book as the student acquires experience. In a number of cases a difficult proof is approached by a preview of the method, and afterward it is illustrated by a numerical example. In particular, special cases of the Jordan normal form are proved in previous chapters by essentially the same method as that used for the general case in Chapter 6.

I wish to give special thanks to Professor Andrew M. Gleason, without whose careful reading and most perceptive comments the book might have been written in half the time but would have been half as good. Thanks are also due to Edward Millman for his editorship, to Holden-Day, Inc., for the production, and to Mrs. Mae Jean Ruehlman for the typing of the manuscript.

Burton W. Jones

CONTENTS

1

VECTOR SPACES

1.1 INTRODUCTION

What is linear algebra? Often in mathematics we begin with a simple idea, work it and knead it and stretch it until the final product has very little resemblance to what we started with. The only clue which remains is, perhaps, a word in the name. This process, of course, happens outside mathematics as well. Who would think that a beautiful glass *objet d'art* was made mostly of sand.

The word *linear* should have something to do with a line, and indeed it has in the beginning. We know that in analytic geometry the equation of a line has the form $ax + by + c = 0$. So we call $ax + by + c$ a *linear expression*. A property of this expression is that x and y each occur to the first degree. Taking our cue from this, we say that $ax + by$ is *linear in x and y* and that $ax + by + c = 0$ is a *linear equation*. On the other hand, we do not call the expression $2x + xy$ linear in x *and* y since the degree of xy is $1 + 1 = 2$ in x and y, though it is linear in x alone.

The next step in our generalization is to remove the restriction to two variables. We also call $ax + by + cz$ a linear expression, where a, b, c are thought of as numbers and x, y, z as variables. We are no longer concerned with a line, for the equation

$$ax + by + cz + d = 0$$

is satisfied by points (x,y,z) on a plane. And with an expression of this kind in four or more variables, we leave the realm of visual geometry completely but still, as in an afterimage, think of

$$ax + by + cz + dt + e = 0$$

for instance, as representing a "surface" in four dimensions which is in some vague sense "flat." Again $ax + by + cz + dt$ would be called a linear function or linear expression in the variables x, y, z, and t.

In mathematics, as well as outside of it, it is often easier and certainly more productive to define a concept not just by how it looks but by how it behaves. In the description above we had to use words like *degree* and *term* which are essentially dependent on what an expression looks like. If we consider instead a function, with which you are also familiar, we can concentrate on the behavior. To be somewhat specific, let us think in terms of a function of three variables: $f(x,y,z)$. Now

$$f(x,y,z) = ax + by + cz$$

looks linear. It has the property that

$$(ax + by + cz) + (ax' + by' + cz') = a(x + x') + b(y + y') + c(z + z')$$

In functional notation this shows up a little better. We have

Property 1 If $f(x,y,z)$ is a linear function, then

$$f(x,y,z) + f(x',y',z') = f(x + x', y + y', z + z')$$

Notice one consequence. If we let $x = x'$, $y = y'$, and $z = z'$, we have

$$f(x,y,z) + f(x,y,z) = f(2x,2y,2z)$$

that is,

$$2f(x,y,z) = f(2x,2y,2z)$$

This is a special instance of a second property defined as follows:

Property 2 If $f(x,y,z)$ is a linear function, then

$$f(cx,cy,cz) = cf(x,y,z)$$

for all numbers c. The operation involved is called *multiplication by a scalar*, the *scalar* being the number c.

We have shown above that if $c = 2$ this property follows from Property 1. Indeed it can be seen that Property 1 implies Property 2 for any positive integer c. But we could not use this method if c were $\sqrt{2}$ or -3, for instance. This shows, intuitively, that Property 1 is not sufficient for linearity. We can show, more rigorously, that Property 2 is not sufficient by the following example, where Property 2 holds but Property 1 does not. Let

$$f(x,y,z) = \sqrt{x^2 + y^2 + z^2}$$

This satisfies Property 2. But the following shows that it does not satisfy Property 1:

$$\sqrt{x^2 + y^2 + z^2} + \sqrt{x'^2 + y'^2 + z'^2} \neq \sqrt{(x + x')^2 + (y + y')^2 + (z + z')^2}$$

On the other hand, $f(x,y,z) = ax + by + dz$ does have Property 2 by the distributive property. Thus we formulate our definition of linearity of a function.

Definition A function $f(x,y,z)$ is said to be *linear* if it satisfies Properties 1 and 2.

This definition can be extended to any number of variables as follows:

Definition Let $f(x_1, x_2, \ldots, x_n)$ be a function of n variables. We call it *linear* if it satisfies the following two properties:

1. $f(x_1, x_2, \ldots, x_n) + f(y_1, y_2, \ldots, y_n) = f(x_1 + y_1, \ldots, x_n + y_n)$
2. $f(cx_1, cx_2, \ldots, cx_n) = cf(x_1, x_2, \ldots, x_n)$

It would be an oversimplification to say that linear algebra is the study of linear functions. But it would be not far from the truth to say that linear algebra is a study of mathematical systems which have the two properties above.

For those who prefer to think geometrically, we can also associate linearity with flatness as mentioned earlier. A plane can be characterized by the fact that if P and Q are any two points of the plane, then the line determined by P and Q lies entirely in the plane. It is not hard to make the connection, by means of analytic geometry, between this geometrical concept and linear functions, but we do this more efficiently later (see Exercise 8 of Sec. 1.7).

Why does one study linear algebra? One reason is that linear functions are simpler than functions which are quadratic or of higher degree. This means that linearity gives us a powerful tool for getting results, both within and outside of mathematics. As we develop the theory we shall point out some of its uses. (In fact, the uses which one makes of a theory often determine the direction in which it is to progress.) Also, the development has an intrinsic interest of its own.

1.2 VECTORS

Of all the examples of mathematical systems which have the property of linearity, vectors seem the most convenient to begin with since they are useful in considering other linear systems. Let us begin by discussing them very intuitively and recalling their properties. We think of a vector as a line segment with a point on one end or as an arrow without a tail. The length of a vector might be a measure of the magnitude of a force, say, and its direction shows the direction of the force. We could represent a vector in a plane by two points, say (1,3) and (6,9), together with the line segment between them and the designation of one as the initial point (the blunt end) and the other as the terminal point (the sharp end).

Now the vector from (1,3) to (6,9) has the same magnitude and direction as that from (0,0) to (6 − 1, 9 − 3), that is, from (0,0) to (5,6) (see Fig. 1.1). We call these two vectors *equivalent*. In general two vectors are defined to be equivalent if they have the same length and direction. The idea of direction is hard to pin down rigorously; and since we are operating on an intuitive level at this point it is probably best to leave it undefined.

There are a number of operations that can be performed on a vector. One is to change its length or magnitude; another is to change its direction; a third is to "move it around," that is, replace it by an equivalent vector as defined in the above paragraph. Consider now the operation of addition of two vectors. Here we can take our cue from the physical situation and then

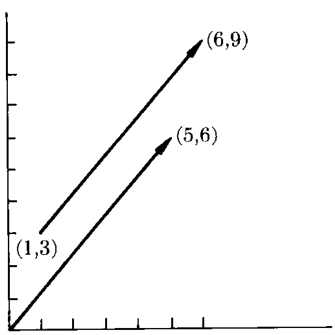

Figure 1.1

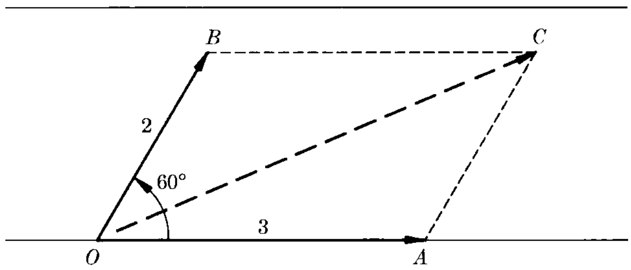

Figure 1.2

see what this amounts to algebraically. To be more specific, suppose a stream is flowing eastward at a rate of 3 miles per hour, and a person is rowing across it at an angle of 60° with the bank and at a rate of 2 miles per hour. We can ask the question: At what rate and in what direction is he approaching the opposite shore? The theory of vectors in physics informs us that we can find the answer by drawing a horizontal vector 3 units long to represent the direction and speed of the water and a vector of length 2 at an angle of 60° with it to represent the direction and speed (velocity) of the rower (see Fig. 1.2). Then if we "complete the parallelogram" as shown, the length of the vector OC gives the rower's rate across the stream and its direction determines his direction. The vector OC is called the *resultant* or *sum* of the vectors OA and OB. One can also find the resultant by completing a triangle as indicated in Fig. 1.3.

In this particular case, let us see what this vector addition amounts to algebraically. Consider the first figure placed on coordinate axes so that point A has the coordinates (3,0), and B is denoted by (1,2). Then it is not hard to see that C has the coordinates $(3 + 1, 0 + 2) = (4,2)$. We merely add the corresponding coordinates.

1.3 VECTORS AS n-TUPLES

We can preserve most of the intuitive properties of vectors and become much more precise if we largely confine ourselves to vectors whose initial points are at the origin of coordinates; any vector is equivalent to one of these. One advantage of doing this is that we can then identify any such vector by the

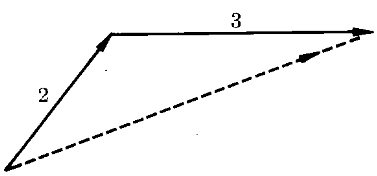

Figure 1.3

coordinates of its terminal point. In this sense we can then think of the ordered pair (a,b) as being the vector whose initial point is the origin and whose terminal point is that defined by the pair (a,b). In three dimensions, we similarly call the ordered triple (a,b,c) a vector, thinking of the physical vector that starts at the origin and has the point (a,b,c) as its terminal point. What we lose by this point of view is the opportunity to consider vectors whose initial points are not at the origin. We shall see that the gain much more than counterbalances the loss. So we have the following definition:

Definition A *vector* is a sequence of n numbers $(a_1, a_2, \ldots, a_n)$, for n some positive integer, that is, an ordered n-tuple of numbers. The a's are called the *components* of the vector. (Note that, for instance, (1,2,3) and (3,2,1) are different vectors.)

We may think of these numbers as being rational, real, complex, or, in fact, numbers of any field (see Appendix A). In this book we generally use lowercase Greek letters to represent vectors, and lowercase Roman letters to represent numbers. Thus we might write $\alpha = (a_1, a_2, \ldots, a_n)$.

1.4 OPERATIONS ON VECTORS

When adding vectors, we use the algebraic equivalent of the physical vector sum illustrated in the previous section. Thus the *resultant* or *sum* of the two vectors from the origin to the points (a_1,a_2) and (b_1,b_2) is the vector from the origin to $(a_1 + b_1, a_2 + b_2)$, as may be seen geometrically. So we define the sum of two vectors in the following way:

Vector Addition If $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$, then

$$\alpha + \beta = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

Note that we can add two vectors only if the number of components of one is the same as the number of components of the other; to add two vectors we add corresponding components.

We can stretch or contract a vector, that is, change its length and/or change its direction. Algebraically this amounts to multiplying each component of the n-tuple by a number. So our next operation is

Multiplication by a Scalar If $\alpha = (a_1, a_2, \ldots, a_n)$ and c is a number, then

$$c\alpha = \alpha c = (ca_1, ca_2, \ldots, ca_n)$$

The number c is, in such a connection, often called a *scalar*. We usually write it to the left of the vector, but sometimes (see Sec. 1.7) it is more convenient to put it on the right. The two vectors are the same since the number c is commutative with the a's.

Two vectors are equal if and only if they are the same. For instance, $(2,3) \neq (3,2)$, but $(\frac{4}{2},\frac{6}{3}) = (2,2)$. The vectors all of whose components are zero, that is, $(0, 0, \ldots, 0)$, are called *zero* or *null* vectors. We use the symbol θ to denote the null vector regardless of the number of components, with the understanding that if we write, for example, $\alpha + \theta$, then α and θ have the same number of components. From the definition of vector addition, we have

$$\alpha + \theta = \theta + \alpha = \alpha$$

The null vector is the additive identity.

The additive inverse of α is

$$-\alpha = (-1)(a_1, a_2, \ldots, a_n) = (-a_1, -a_2, \ldots, -a_n)$$

since $\alpha + (-\alpha) = (-\alpha) + \alpha = \theta$.

It should be noted that there is again no conflict between multiplication by a scalar and vector addition. For example, as we saw above, $\alpha + \alpha = 2\alpha$ whether by vector addition or multiplication by a scalar. It is important to notice that our definitions of addition of vectors and multiplication by a scalar have been precisely in accord with the requirements of linearity imposed in the first section of this chapter. That is, if in the two properties of linearity which we stated in Sec. 1.1 we omit the symbol f, we have vector addition and multiplication by a scalar.

We have thus set up a kind of model for vectors. Just as the vectors themselves are models for physical concepts like force and velocity, so our n-tuples are models for vectors. We have preserved all the properties of "physical" vectors in our mathematical model except equivalence. (This could be recovered if needed.) What have we gained, besides preciseness and a simple method for adding vectors? The answer to this question should become clear in later sections.

EXERCISES

1. In each case below when the indicated sum exists, express it as a single vector; when it does not exist, explain why.

(*a*) $(1,2,3) + (3,4,5)$ (*b*) $3(1,0,5) + 2(1,0)$
(*c*) $3(1,0,5) + 2(\frac{1}{2},\sqrt{3},4)$ (*d*) $5(1,0,7) - \sqrt{3}(4,\frac{1}{2},7)$
(*e*) $(1,0,-6,7) + 0(1,2,3)$ (*f*) $2(1,0,5,7) + (2,3,-1,7) + 0(5,6,7)$

2. Find the vector α satisfying each of the following equations:

(*a*) $\alpha + (3,4,5) = (7,1,-7)$ (*b*) $3\alpha = (1,2,3)$
(*c*) $\alpha + 2(3,1,0) = (2,0,3)$ (*d*) $5\alpha + (7,-1,3) = 7\alpha$