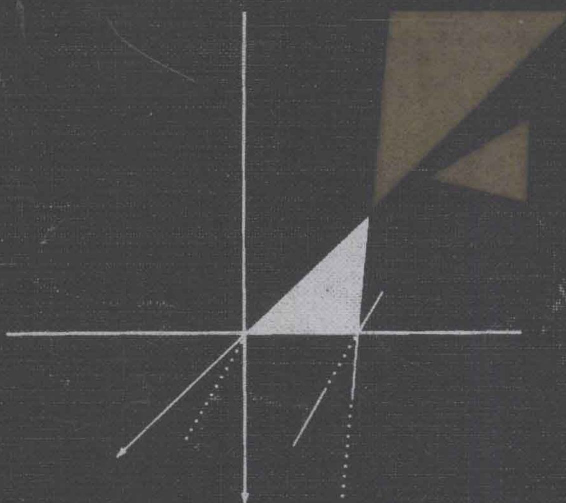
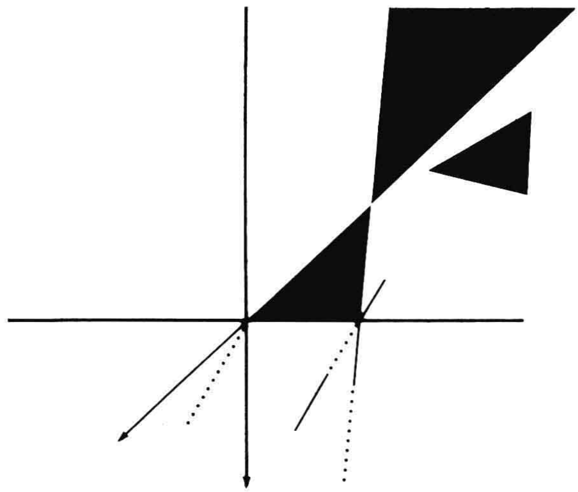


INTRODUCTION TO LINEAR ALGEBRA

PETER J. KAHN

HARPER'S SERIES IN MODERN MATHEMATICS



INTRODUCTION TO LINEAR ALGEBRA

PETER J. KAHN,
CORNELL UNIVERSITY

HARPER'S SERIES IN MODERN MATHEMATICS

NEW YORK / EVANSTON / LONDON

Introduction to Linear Algebra

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 67–12551

INTRODUCTION TO LINEAR ALGEBRA

I. N. HERSTEIN AND GIAN-CARLO ROTA, EDITORS

HARPER & ROW, PUBLISHERS

TO TANIA

TRUDE

OSCAR

JIM

SAM

ED

WADE

JOHN

BOB

"FUZZY"

PREFACE

This book is intended as an introduction to linear algebra for the good undergraduate novice mathematics student. It is almost entirely self-contained, including even a rather extensive treatment of elementary set theory. An unimportant assumption is made that the student has some familiarity with elementary differential and integral calculus—unimportant because the only references in the text to this subject appear in illustrative examples and not in the main body.

On the other hand the mathematical viewpoint of the text is basically rather sophisticated and abstract, the underlying philosophy being that this approach is ultimately more meaningful and useful. For example, linear equations are treated near the end of the book rather than near the beginning, the entire machinery of vector spaces and linear transformations having already been established in some detail. For another example, emphasis is placed on the geometric aspects of linear algebra. Thus determinants are introduced as scalar changes in volume effected by linear transformations.

To bridge the gap between the mathematical sophistication of the text and that of the student, a good deal of the text is devoted to motivational exposition. For example, the notion of vector-space isomorphism is heuristically examined in great detail. Or, for another example, before the standard abstract definition of determinant is presented (in terms of alternating n-linear forms), the entire concept is developed on a concrete level in terms of actual oriented area of oriented parallelograms in two-dimensional real Euclidean space. Exercises carry the student further, via oriented volume of oriented parallelepipeds, to three-dimensional real Euclidean space.

One result of this gap-bridging is a text that is, *in toto*, too long for a one-semester course. Nevertheless, by judicious omission of certain sections or portions of sections the numbers of which are indicated by daggers (†), most of the important topics in the text can be covered in one semester. Below we list two possible distributions of course time for text material. The column on the left covers, in one form or another, most of the basic concepts of the text.

The column on the right describes a course of more limited scope, more leisurely in nature, thus permitting the student more time to absorb and explore the elementary linear algebra ideas in greater depth. Needless to say, students with a reasonably good background in set theory may skip Chapter 2, after a cursory reading to acquaint themselves with the notation and terminology used in the text. For such students, a more ambitious program is, of course, possible.

	Number of Weeks	
Chapter	Option 1 (with omissions)	Option 2 (without omissions)
1	$\frac{1}{3}$	—
2	2	3
3	3	4
4	4	4
5	2	4
6	2+	—
7	1+	—

The exercises are designed to be supplementary to text material. Occasionally, an exercise is an important extension of text material, perhaps cited subsequently in the text. In such a case, the exercise is indicated by an asterisk after its number.

At this point I would like to thank Mrs. Ollie Cullers, Mrs. Marjorie Proaper, and Mrs. Ellen Varney for their good typing, and my wife for her help at the final stages of manuscript preparation.

PETER J. KAHN

CONTENTS

CHAPTER 1

INTRODUCTION TO LINEARITY

It is widely recognized today that the concept of linearity is fundamental to most of modern mathematics, both abstract and applied. This chapter is intended to give the reader some idea of the scope of this concept. We shall avoid precise general definitions here. Instead, we proceed by presenting briefly a number of examples that involve the concept of linearity, pausing now and then to direct the attention of the reader to some of the relevant features of the examples.

There are two equally important ways of looking at linearity. We can think of it as a property of certain *sets*, or as a property of certain *functions*. In Chapter 2 we discuss the abstract notions of sets and functions in detail. For our purpose now, only a rough intuitive understanding of these ideas is necessary. The reader who is totally unfamiliar with them, however, should look at Definitions 2.1 and 2.15 before continuing.

First, we shall look at some examples of *linear sets* (as we shall call them temporarily).

(a) We begin with the set of all real numbers, which, henceforth, we call "**R**." Given any two real numbers x and y, their *sum*, $x + y$, and *product*, xy, are also real numbers. We express these facts by saying that **R** is *closed* with respect to addition and multiplication. Alternatively, we shall say that **R** is a *linear set* (with respect to these operations).

(b) Next, consider a straight line ℓ lying in the Cartesian plane, which we call "$\mathbf{R}^2$," and passing through the origin (see Figure 1). Suppose that ℓ is not vertical. Then, the equation of ℓ is of the form $y = ax$, for some fixed real number a.

The line ℓ consists of all ordered pairs of real numbers of the form (x, ax). Let us agree to define the "sum" of any two ordered pairs of real numbers (x_0, y_0) and (x_1, y_1) to be the ordered pair $(x_0 + x_1, y_0 + y_1)$. That is,

$$(x_0, y_0) + (x_1, y_1) = (x_0 + x_1, y_0 + y_1)$$

Moreover, given any real number r and any ordered pair of real numbers (x, y), let us agree to define the "product" of r and (x, y) to be (rx, ry). That is, $r(x, y = (rx, ry)$.

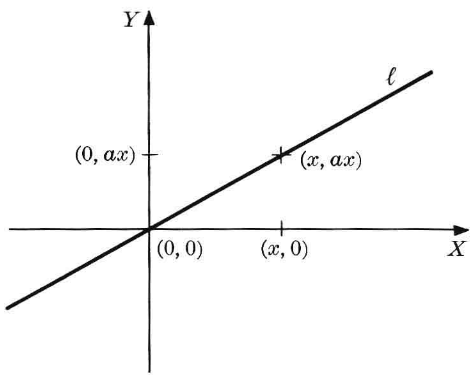


FIGURE 1

With these definitions in mind, it is easy to see that the sum of any two points on ℓ is again a point on ℓ. For

$$(x_0, ax_0) + (x_1, ax_1) = (x_0 + x_1, ax_0 + ax_1) = (x_0 + x_1, a(x_0 + x_1))$$

Moreover, every real multiple of a point on ℓ is again a point on ℓ. For

$$r(x, ax) = (rx, rax) = (rx, a(rx))$$

We express these facts by saying that ℓ is closed with respect to the operations of vector addition (i.e., "addition" of ordered pairs of real numbers) and real multiplication. Again, alternatively, we say that ℓ is a linear set.

(c) The Cartesian plane, $\mathbf{R}^2$, discussed above, is also a linear set. It consists of *all* ordered pairs of real numbers (x, y). In the above discussion, we defined the sum of any two such ordered pairs; it is again an ordered pair (of real numbers). Moreover, every real multiple of an ordered pair of real numbers is again an ordered pair of real numbers, by our definition of "real multiple."

Therefore, the plane is closed with respect to vector addition and real multiplication, and it is, therefore, a linear set.

(d) Let A be any nonvacuous set whatsoever, and let $\mathbf{R}^A$ be the collection of all real-valued functions defined on A. That is, a typical member of $\mathbf{R}^A$ is a function f whose argument, x, ranges over the set A and whose values are real numbers (i.e., $f(x)$ is a real number, for every x in A). To describe in what way $\mathbf{R}^A$ can be considered to be a linear set, we shall describe how to add two members of $\mathbf{R}^A$ and how to multiply a member of $\mathbf{R}^A$ by any real number.

Let f and g be any members of $\mathbf{R}^A$. They are real-valued functions. Therefore, for any x in A, $f(x)$ and $g(x)$ are real numbers, which we know how to add. Put briefly, we shall add f and g by adding their function values $f(x)$ and $g(x)$. In other words, the sum, $f+g$, of the functions f and g is a certain function whose value at any given x in A is given by $f(x)+g(x)$; that is, $(f+g)(x) = f(x)+g(x)$. Similarly, we multiply f by a real number r by multiplying the function values of f by r. That is, rf is a function whose value at x is $rf(x)$ or $(rf)(x) = rf(x)$.

By definition, then, the sum of any two members of $\mathbf{R}^A$, or any real multiple of a member of $\mathbf{R}^A$, is again a member of $\mathbf{R}^A$. Therefore, $\mathbf{R}^A$ has the desired closure property, and, thus, it is a linear set.

(e) This example is similar to the previous one, but slightly more interesting and much more important. Let a and b be any two real numbers, $a < b$. Let $\mathscr{C}[a, b]$ be the set of all continuous, real-valued functions defined on the closed interval $[a, b]$. Define the sum of two such functions and real multiples of such functions as above. It is a well-known fact from elementary calculus that the sum of two continuous functions is continuous and that every real multiple of a continuous function is again continuous. These facts are alternatively expressed by saying that $\mathscr{C}[a, b]$ is closed with respect to addition and real multiplication of functions. Thus, $\mathscr{C}[a, b]$ is a linear set.

(f) Finally, let a, b be as above, and let $\mathscr{D}[a, b]$ be the set of all real-valued functions defined on the closed interval $[a, b]$ and possessing a continuous first-derivative on this interval. Define sums and real multiples of such functions as above. It is, again, a well-known result of elementary calculus that the sum of any two members of $\mathscr{D}[a, b]$ and any real multiple of a member of $\mathscr{D}[a, b]$ is again a member of $\mathscr{D}[a, b]$. Therefore, $\mathscr{D}[a, b]$ is a linear set.

We urge the reader to notice that in all of the above examples, our description did not stop with the specification of such and such a set. Essential to our description was the specification of certain operations, which we called "addition" and "real multiplication" because of their similarity to the usual notions of adding and multiplying real numbers. The set was then shown to be closed *with respect to these particular operations*. Henceforth, we shall never consider the description of a linear set to be complete unless the operations are specified. When referring to the sets $\mathbf{R}$ and $\mathbf{R}^2$, defined above, we shall always think of them as linear sets with respect to the operations defined above in Examples (a) and (c), respectively, unless explicitly stated otherwise.

EXERCISES / 1.1

1. Add the following pairs of real numbers, as prescribed in Example (b), above:

   a. (2, 4) and (4, 2)
   b. (2, 4) and (−2, −4)
   c. (0, 0) and (4, 2)
   d. (2, 4) and (3, 6)
   e. (2, 4) and (5, 7)

   Locate each of the pairs and their sum on the Cartesian plane.

2. Multiply each of the following pairs by 2/3, as prescribed in Example (b), above:

   a. (2, 4) b. (0, 0) c. (3, −7)

   Multiply them by 3/2; by −3/2. Locate them and all three of their multiples on the Cartesian plane.

3. Sketch, in the Cartesian plane, the graphs specified by the following equations:

   a. $y = 2x + 3$
   b. $x = 2 - 3y$
   c. $x^2 + y = 4$
   d. $y = \sqrt{2 - x^2}$
   e. $y = 0$

   Each of the above specified graphs is a certain set of ordered pairs of real numbers. Which of these sets is closed with respect to the operations of adding ordered pairs of real numbers and of multiplying them by real numbers, as defined in Example (b)? That is, which of the graphs are linear sets?

4. Consider the case of a vertical line lying in $\mathbf{R}^2$. What is its general equation (using the standard x, y notation)? Describe the ordered pairs of real numbers that lie on such a line. Which vertical lines, if any, are closed with respect to the operations defined in Example (b)?

5.* Show that, if (x_0, y_0) and (x_1, y_1) are any two ordered pairs of real numbers in $\mathbf{R}^2$ not equal to (0, 0) and thought of as opposite vertices of a parallelogram of which another vertex is (0, 0), then the sum $(x_0, y_0) + (x_1, y_1)$ [as defined in Example (b)] is the fourth vertex of the parallelogram. This result provides a geometric interpretation of addition of ordered pairs.

6.* Consider the set of all ordered triples (x, y, z) of real numbers. Define the sum of two triples, (x_0, y_0, z_0) and (x_1, y_1, z_1) by

$$(x_0, y_0, z_0) + (x_1, y_1, z_1) = (x_0 + x_1, y_0 + y_1, z_0 + z_1),$$

and, for any real number r, define the product of r and (x, y, z) by

$$r(x, y, z) = (rx, ry, rz)$$

* An asterisk following an exercise number, here and throughout the book, indicates an exercise that is an important extension of text material (see Preface).

Now, let a, b, c be *any fixed* real numbers and consider the set S of all triples (x, y, z) satisfying the equation

$$ax + by + cz = 0$$

Is S closed with respect to the operations defined above? Justify your answer. Show that the set of all *triples of real numbers is a linear set with respect to the above operations.* Call this set of triples $\mathbf{R}^3$. Henceforth, when we refer to $\mathbf{R}^3$, we shall always think of it as a linear set with respect to the operations defined above.

Now, we look at some examples of *linear functions* (as we shall call them temporarily).

(g) Consider the function given by the equation

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

where the variable x is any real number (that is, any member of $\mathbf{R}$). Let x_0 and x_1 be any given real numbers, and notice that

$$f(x_0 + x_1) = 2(x_0 + x_1) = 2x_0 + 2x_1 = f(x_0 + fx_1)$$

Moreover, for any real numbers r and x,

$$f(rx) = 2rx = r(2x) = rf(x)$$

We express these two properties of f by saying that f respects the operations of addition and real multiplication. This characteristic of f is closely related to the closure property of linear sets. For the graph of f is, of course, just a line through the origin having slope 2. According to Example (b), above, such a line is closed with respect to the operations of adding ordered pairs of real numbers and real multiplication of such ordered pairs. A typical ordered pair on the line is of the form $(x, 2x) = (x, f(x))$. The sum of two ordered pairs of this form is given by

$$\begin{aligned}(x_0, f(x_0)) + (x_1, f(x_1)) &= (x_0, 2x_0) + (x_1, 2x_1)\\ &= (x_0 + x_1, 2x_0 + 2x_1) = (x_0 + x_1,\ 2(x_0 + x_1))\\ &= (x_0 + x_1, f(x_0 + x_1)),\end{aligned}$$

which is again of the same form. Similarly,

$$r(x, f(x)) = (rx, rf(x)) = (rx, f(rx)),$$

which is, again, on the line, for any r. Thus, the fact that f respects the operations of $\mathbf{R}$ is really the same as the fact that the graph of f is a linear set. We therefore say that f is a *linear function.*

(h) Next, we shall define a real-valued function E whose argument y ranges over the linear set $\mathbf{R}^A$ of Example (d)! That is, for every y in $\mathbf{R}^A$ (i.e., y is a real-valued function of a variable x ranging over the set A), $E(y)$ is going to be a certain real number.

To define E, we choose an arbitrary member a of the set A and hold it fixed throughout the entire discussion. The real number $E(y)$ is then defined to be $y(a)$, the value of y for the argument $x = a$. That is, $E(y) = y(a)$. We chose the letter "E" for this function to emphasize the fact that E is an *evaluation* function: given any y, $E(y)$ is the evaluation of y at a.

Now, let y_0 and y_1 be any two members of $\mathbf{R}^A$. Remember that by the definition in Example (d), the sum $y_0 + y_1$ satisfies

$$(y_0 + y_1)(a) = y_0(a) + y_1(a)$$

In terms of E, this means that $E(y_0 + y_1) = E(y_0) + E(y_1)$. Similarly, for any real number r,

$$(ry)(a) = r(y(a)),$$

so that $E(ry) = rE(y)$.

In this example, the argument of the function E and the values of E lie in different sets, $\mathbf{R}^A$ and $\mathbf{R}$, respectively. In the equation $E(y_0 + y_1) = E(y_0) + E(y_1)$, the "+" on the left denotes the "addition" in $\mathbf{R}^A$, whereas the "+" on the right denotes addition in $\mathbf{R}$. We can summarize the equation verbally by saying that E respects the addition operations of $\mathbf{R}^A$ and $\mathbf{R}$. Similarly, E respects the real multiplication operations of $\mathbf{R}^A$ and $\mathbf{R}$.

Thus, E is a function that takes values in a linear set and whose argument ranges over a linear set. Moreover, E respects the operations of these linear sets. Therefore, we say that E is a linear function.

(i) Consider the linear set $\mathscr{C}[a, b]$ of Example (e). We define a real-valued function $\mathscr{I}$ whose value $\mathscr{I}(f)$, for any given f in $\mathscr{C}[a, b]$, is

$$\int_a^b f(x)\,dx$$

That is,

$$\mathscr{I}(f) = \int_a^b f(x)\,dx,$$

for every continuous, real-valued function f defined on the closed interval $[a, b]$.

Now, let f_0 and f_1 be any two such functions. Then, by definition of $\mathscr{I}$,

$$\begin{aligned}\mathscr{I}(f_0 + f_1) &= \int_a^b (f_0 + f_1)(x)\,dx \\ &= \int_a^b f_0(x)\,dx + \int_a^b f_1(x)\,dx \\ &= \mathscr{I}(f_0) + \mathscr{I}(f_1)\end{aligned}$$

In a similar way, we can show that $\mathscr{I}(rf) = r\mathscr{I}(f)$, for any real number r and any function f in $\mathscr{C}[a, b]$.

Therefore, $\mathscr{I}$ respects the operations of $\mathscr{C}[a, b]$ and $\mathbf{R}$ so that it is a linear function. This example can be summarized briefly by saying that *the definite integral is a real-valued linear function on* $\mathscr{C}[a, b]$.

(j) We define a function D whose argument ranges over $\mathscr{D}[a, b]$, the linear set of Example (f), and whose values range over the linear set $\mathscr{C}[a, b]$. We shall show that D respects the operations of $\mathscr{D}[a, b]$ and $\mathscr{C}[a, b]$.

Put very simply, for every f in $\mathscr{D}[a, b]$, we let $D(f)$ be the first derivative of f. That is, for every x in $[a, b]$, the value of $D(f)$ at x is $f'(x)$, or $(D(f))(x) = f'(x)$.

We leave it to the reader to show that for any f_0, f_1 and f in $\mathscr{D}[a, b]$, and for any r in $\mathbf{R}$,

$$D(f_0 + f_1) = D(f_0) + D(f_1)$$

$$D(rf) \qquad = rD(f)$$

This example can be summarized by saying that *differentiation is a linear function on* $\mathscr{D}[a, b]$.

(k) Let f be any real-valued function of a real variable that has a continuous first derivative. Choose any real number x_0, and consider the tangent line to the graph of f at $(x_0, f(x_0))$ (see Figure 2).

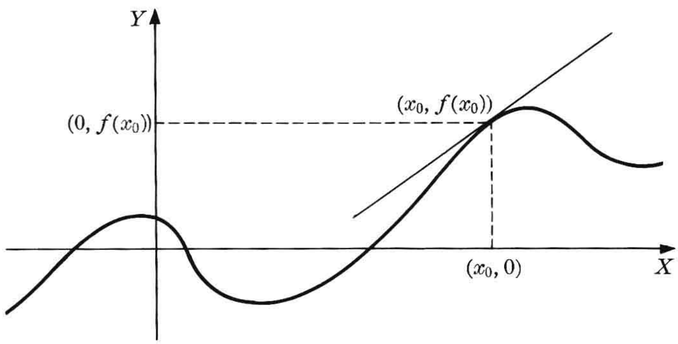


FIGURE 2

The equation of this line is

$$y - f(x_0) = f'(x_0)(x - x_0)$$

It is easy to see (cf. Exercises 1.1. **3a**, **b**) that, unless $f(x_0) = x_0 f'(x_0)$, the tangent line is *not* linear.

However, if we translate the X and Y axes so that (0, 0) goes to $(x_0, f(x_0))$, then the tangent line will go through the origin in the new coordinate system. We shall call the new Y-axis the "dY-axis" and the new X-axis, the "dX-axis." The corresponding variables are called "dy" and "dx," respectively. Actually, we should indicate dependence on x_0 in some way, but to simplify notation we avoid explicit reference to x_0. The situation is illustrated in Figure 3.