THIRD EDITION

INTRODUCTION TO MATRICES AND LINEAR TRANSFORMATIONS

DANIEL T. FINKBEINER II

THIRD EDITION

INTRODUCTION TO MATRICES AND LINEAR TRANSFORMATIONS

DANIEL T. FINKBEINER II

Kenyon College

W. H. FREEMAN AND COMPANY

San Francisco

Library of Congress Cataloging in Publication Data

Finkbeiner, Daniel Talbot, 1919–
Introduction to matrices and linear transformations
(A Series of books in the mathematical sciences)
Includes index
1 Matrices 2 Algebras, Linear I Title
QA188 F56 1978 512 9'43 78-18257
ISBN 0-7167-0084-0

AMS 1970 subject classification 15-01

INTRODUCTION TO MATRICES AND LINEAR TRANSFORMATIONS
THIRD EDITION

Printed in the United States of America

9 8 7 6 5 4 3 2 1

PREFACE

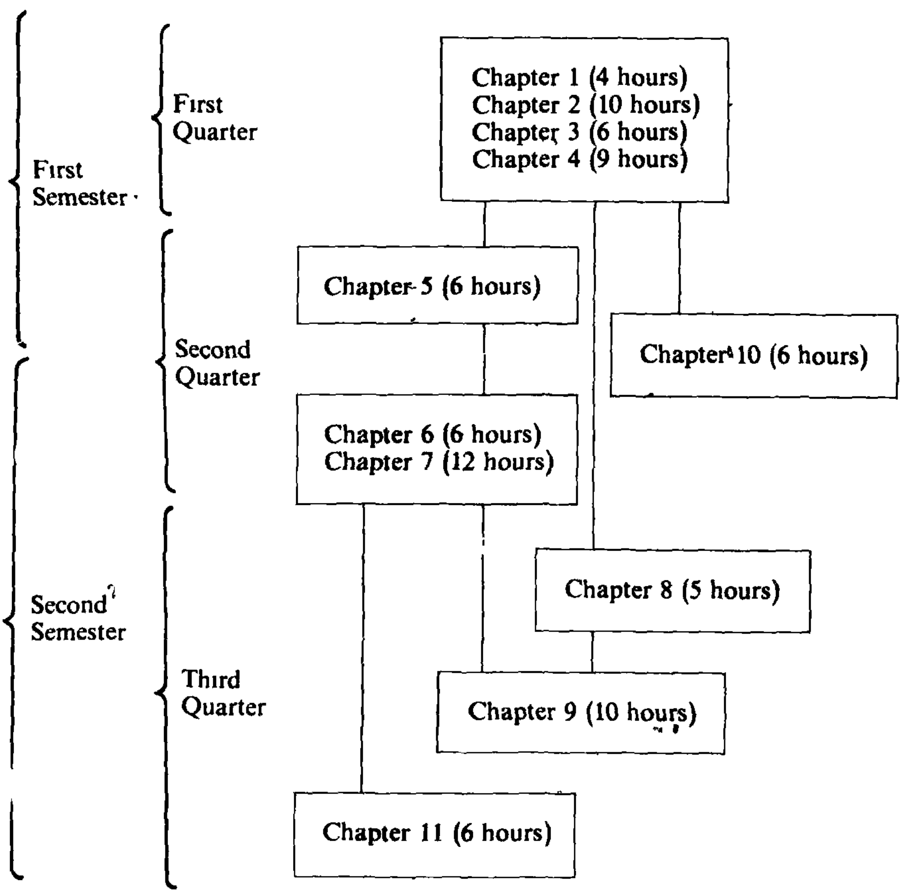
Persons familiar with earlier editions of this book will observe a number of changes and improvements in this new version. Although the previous approach and organization have been retained, virtually all the exposition has been rewritten, with more illustrative examples, new exercises, and an expanded solutions section. Students learn mathematics by practicing it, and practice can be stimulated by having detailed solutions available as guidance. Students should be encouraged to read *all* the exercises and to note especially those that extend the ideas of the text.

A change of notation has been adopted for the new edition in that linear mappings are treated as *left*-hand operators on vectors, in conformity with customary function notation.

More significantly, this revision begins concretely to allow students time to develop understanding before general concepts are introduced. The book begins with the familiar problem of solving a system of linear equations. This is used to introduce the concept of a vector and to motivate the ideas of vector and matrix algebra. Throughout the book Gaussian elimination is used as a unifying computational technique.

Metric notions of Euclidean space are introduced at an early stage to establish a familiar geometric setting in which the concepts of linear algebra can be interpreted. Euclidean spaces are reexamined more generally in Chapter 8.

The book can be used flexibly for a variety of courses in linear algebra. In view of the differences in pace and emphasis of mathematics instruction at various institutions, the following time estimates should be regarded only as rough guidelines, reflecting experience at Kenyon College. The book contains enough material for a year of linear algebra, with Chapters 1–5 constituting a suitable first-semester course. (If time permits, Chapter 10 can be included as a significant application of special interest to students of business and economics. Alternatively, Chapter 8 can be covered to deepen the earlier exposure to inner product spaces.) Chapters 6–9 comprise a suitable second-semester course, with Chapters 10 and 11 as optional material if time permits..



Courses in linear algebra currently are elected by students ranging from freshmen to seniors and with widely different professional goals. Consequently, such courses must be paced carefully, with time allowed for review and assimilation of ideas. As an average rule an instructor should allow at least three class hours for every two sections of text material, allotting some class time for questions and discussions of exercises

Chapter interdependencies and an illustrative arrangement of courses are shown in the diagram, but each instructor should adjust this schedule to the needs of the class.

I am indebted to many persons—in particular, the Kenyon students who pointed out errors in an earlier draft of this material, to Wendell Lindstrom who offered thoughtful suggestions for improvement, and to Hope Weir who typed the manuscript with rare expertise. The errors and flaws that remain are my responsibility, and I shall welcome the help of readers in finding them and in sending me their comments and criticisms Special gratitude is expressed to Charles Rice, Richard Hoppe, Gerald Chaplin, George Fowler, Steve Alex, Danny Vaughn, James Carhart, and the late Thomas L. Bogardus, Jr., whose contributions to this work were indirect but most essential.

May 1978 DANIEL T. FINKBEINER II

CONTENTS

CHAPTER 1
LINEAR EQUATIONS

Linear algebra is concerned primarily with mathematical systems of a particular type (called *vector spaces*), functions of a particular type (called *linear mappings*), and the algebraic representation of such functions by matrices. If you have completed a course in calculus, you are already familiar with some examples of vector spaces, such as the real number system **R** and the Euclidean plane. You also have studied functions from **R** to **R**, so at least superficially the study of linear algebra appears to be a natural extension and generalization of your previous studies. But you should be forewarned that the degree of generalization is substantial and the methods of linear algebra are significantly different from those of calculus.

A glance at the Table of Contents will reveal many terms and topics that might be unfamiliar to you at this stage in your mathematical development. Therefore, as you study this material you will need to pay close attention to the definitions and theorems, assimilating each idea as it arises, gradually building your mathematical vocabulary and your ability to utilize new concepts and techniques. You are urged to make a practice of reading all the exercises and noting the results they contain, whether or not you solve them in detail.

The contents of this book are a blend of formal theory and computational techniques related to that theory. We begin with the problem, familiar from secondary school algebra, of solving a system of linear equations, thereby introducing the idea of a vector space informally. Vector spaces are not defined formally until Section 3 of Chapter 2. At that point, and from time to time thereafter, you are urged to study Appendix A.1, where algebraic systems are explained briefly but generally. You might not need that much generality to understand the concept of a vector space, but firm familiarity with the notion of an algebraic system will greatly accelerate your ability to feel comfortable with the ideas of linear algebra.

Individuals acquire mathematical sophistication and maturity at different rates, and you should not expect to achieve instant success in assimilating some of the more subtle concepts of this course. With patience, persistence, and plenty of practice with specific examples and exercises, you can anticipate steady progress in developing your capacity for abstract thought and careful reasoning. Moreover, you will greatly enhance your insight into the nature of mathematics and your appreciation of its power and beauty.

1.1 SYSTEMS OF LINEAR EQUATIONS

The central focus of this book is the concept of *linearity*. Persons who have studied mathematics through a first course in calculus already are familiar with examples of linearity in elementary algebra, coordinate geometry, and calculus, but they probably are not yet aware of the extent to which linear methods pervade mathematical theory and application. Such awareness will develop gradually throughout this book as we explore the properties and significance of linearity in various mathematical settings.

We begin with the familiar example of a line L in the real coordinate plane, which can be described algebraically by a *linear equation* in two variables:

$$L\colon ax + by = d.$$

A point (x_0, y_0) of the plane lies on the line L if and onlv if the real number $ax_0 + by_0$ has the value d. The formal expression

$$ax + by$$

is called a *linear combination* of x and y.

By analogy a linear combination of three variables has the form

$$ax + by + cz,$$

where a, b, and c are constants. Any equation of the form

$$ax + by + cz = d$$

is called a linear equation in three variables. If you have studied the geometry of three-dimensional space, you will recall that the graph of a linear equation in three variables is a *plane*, rather than a line. This is a significant observation: *the word linear refers to the algebraic form of an equation rather than to the geometric object that is its graph.* The two meanings coincide only for the case of two variables—that is, for the coordinate plane. In general, a linear equation in n variables has the form

$$c_1x_1 + c_2x_2 + c_3x_3 + \cdots + c_nx_n = d,$$

where at least one $c_i \neq 0$. For $n > 3$ the graph of this equation in n-dimensional space is called a *hyperplane.*

Applications of mathematics to science and social science frequently lead to the need to solve a *system* of several linear equations in several variables, the coefficients being real numbers:

$$(1.1)\qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= d_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= d_2, \\ &\ \ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= d_m. \end{aligned}$$

The number m of equations might be less than, equal to, or greater than the number n of variables. *A solution* of the System 1.1 is an ordered n-tuple $(c_1, \ldots, c_n)$ of real numbers having the property that the substitution

$$\begin{aligned} x_1 &= c_1, \\ x_2 &= c_2, \\ &\ \ \vdots \\ x_n &= c_n, \end{aligned}$$

simultaneously satisfies each of the m equations of the system. *The solution* of (1.1) is the set of *all* solutions, and *to solve* the system means to describe the set of all solutions. As we shall see, this set can be finite or infinite.

This problem is considered in algebra courses in secondary school for the case $m = 2 = n$, and sometimes for other small values of m and n. But a large scale linear model in contemporary economics might require the solution of a system of perhaps 83 equations in 115 unknowns. Hence we need to find very efficient procedures for solving (1.1), regardless of the values of m and n, in a finite number of computational steps. Any fixed set of instructions that is guaranteed to solve a particular type of problem in a finite number of steps is called an *algorithm*. Many algorithms exist for solving systems of linear equations, but one of the oldest methods, introduced by Gauss, is also one of the most efficient. Gaussian elimination, and various algorithms related to it, operate on the principle of exchanging the given system (1.1) for another system (1.1A) that has precisely the same set of solutions but one that is easier to solve. Then (1.1A) is exchanged for still another system (1.1B) that has the same solutions as (1.1) but is even easier than (1.1A) to solve. By increasing the ease of solution at each step, after m or fewer exchanges we obtain a system with the same solutions as (1.1) and in an algebraic form that easily produces the solution. For convenience, we say that two systems of linear equations are *equivalent* if and only if each solution of each system is also a solution of the other.

We first illustrate this idea with a specific example. Soon we shall be able to verify that the following two systems are equivalent, and for the moment we shall assume that they are.

$$\begin{aligned} 6x_1 + 2x_2 - x_3 + 5x_4 &= -8, \\ 3x_1 + 2x_2 + x_3 + 3x_4 &= -1, \\ 4x_1 + x_2 - x_3 + 3x_4 &= -6, \end{aligned} \quad \text{and} \quad \begin{aligned} x_1 \quad - x_3 + x_4 &= -3, \\ x_2 + 2x_3 \quad &= 4, \\ x_3 - x_4 &= 2. \end{aligned}$$

Obviously, we would prefer to solve the second system. To do so we let x_4 be any number, say c. Then

$$\begin{aligned} x_4 &= c, \\ x_3 &= 2 + x_4 = 2 + c, \\ x_2 &= 4 - 2x_3 = 4 - 2(2 + c) = -2c, \\ x_1 &= -3 + x_3 - x_4 = -3 + (2 + c) - c = -1, \end{aligned}$$

and we conclude that for any number c the ordered quadruple

$$\begin{pmatrix} -1 + 0c \\ 0 - 2c \\ 2 + c \\ 0 + c \end{pmatrix}$$

is a solution of the second system and hence of the first. Furthermore, it is easy to see that any solution of the second system must be of that form, and therefore we have produced the complete solution of the first system. There are infinitely many solutions because each value of c produces a different solution. When a system has infinitely many solutions, a complete description of all solutions involves one, two, or more arbitrary constants.

The second system is easy to solve because of its special algebraic form: one of the variables (x_1) appears with nonzero coefficient in the first equation but in no subsequent equation, another variable (x_2) appears with nonzero coefficient in the second equation but in no subsequent equation, and so on. A system of this nature is said to be in *echelon form*. To solve a system that already is in echelon form we first consider the last equation; we solve for the first variable of that equation in terms of the constant term and the subsequent variables. Each subsequent variable may be assigned an arbitrary value. In this case

$$x_4 = c,$$
$$x_3 = 2 + x_4 = 2 + c.$$

Then we consider the next to last equation; we solve for the first variable of that equation, assigning an arbitrary value to any subsequent variable whose value is not already assigned. For this example,

$$x_2 = 4 - 2x_3 = 4 - 2(2 + c) = -2c.$$

Continuing in the same way with each preceding equation, we eventually obtain the complete solution of the system.

What we need, therefore, is a process that leads from a given system of linear equations to an equivalent system that is in echelon form. And that is precisely the process that Gaussian elimination provides, as we now shall see. Beginning with a system in the form (1.1), we can assume that x_1 has a nonzero coefficient in at least one of the m equations. Furthermore, because the solution of a system does not depend on the order in which the equations are written, we can assume further that $a_{11} \neq 0$. Thus we can solve the first equation for x_1 in terms of the other variables:

$$x_1 = a_{11}^{-1}(d_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n).$$

We then replace x_1 by this expression in each of the *other* equations. The resulting equations then contain variables x_2 through x_m, and after collecting the coefficients of each of these variables we obtain the equivalent system

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= d_1, \\
b_{22}x_2 + b_{23}x_3 + \cdots + b_{2n}x_n &= e_2, \\
&\;\;\vdots \\
b_{m2}x_2 + b_{m3}x_3 + \cdots + b_{mn}x_n &= e_m.
\end{aligned} \tag{1.1A}
$$

At this stage we need not be concerned with explicit formulas for the new coefficient b_{ij} and the new constants e_i, where $i \geq 2$ and $j \geq 2$. Such formulas result immediately from a bit of routine algebra, and we record the results here for future reference.

$$b_{ij} = a_{ij} - a_{i1}a_{11}^{-1}a_{1j},$$

$$e_i = d_i - a_{i1}a_{11}^{-1}d_1.$$

The system (1.1A) is said to be obtained from (1.1) by means of a *pivot operation* on the nonzero entry a_{11}.

The second stage of Gaussian elimination leaves the first equation of (1.1A) untouched but repeats the pivot process on the reduced system of $m - 1$ equations in $n - 1$ variables:

$$
\begin{aligned}
b_{22}x_2 + b_{23}x_3 + \cdots + b_{2n}x_n &= e_2, \\
b_{32}x_2 + b_{33}x_3 + \cdots + b_{3n}x_n &= e_3, \\
&\;\;\vdots \\
b_{m2}x_2 + b_{m3}x_3 + \cdots + b_{mn}x_n &= e_m.
\end{aligned}
$$

Conceivably each coefficient b_{i2} is zero; if so, we look at the coefficients b_{i3}, in order, and continue in this way until we find the first *nonzero* coefficient, say b_{rs}. Again because we can write these equations in any order without changing the solutions, we can assume that $r = 2$. Then we pivot on b_{2s}; that is, we solve for x_s as

$$x_s = b_{2s}^{-1}(e_2 - b_{2,s+1}x_{s+1} - \cdots - b_{2n}x_n),$$

and substitute this expression for x_s into each of the last $m - 2$ equations.

Together with the original first equation the new system, equivalent to (1.1) and to (1.1A), is of this form:

$$
\begin{aligned}
a_{11}x_1 + \cdots + a_{1s}x_s + a_{1,s+1}x_{s+1} + \cdots + a_{1n}x_n &= d_1, \\
b_{2s}x_s + b_{2,s+1}x_{s+1} + \cdots + b_{2n}x_n &= e_2, \\
c_{3,s+1}x_{s+1} + \cdots + c_{3n}x_n &= f_3, \\
&\vdots \\
c_{m,s+1}x_{s+1} + \cdots + c_{mn}x_n &= f_m.
\end{aligned} \tag{1.1B}
$$

Then the pivot process is repeated again on the last $m-2$ equations of (1.1B), leaving the first two equations untouched. Continuing in this manner, we eventually obtain a system that is equivalent to (1.1) and is in echelon form.

To illustrate the method of Gaussian elimination we return to our previous example of three equations in four unknowns. The first equation is

$$6x_1 + 2x_2 - x_3 + 5x_4 = -8.$$

We pivot on the coefficient 6 by solving for x_1,

$$x_1 = \tfrac{1}{6}(-8 - 2x_2 + x_3 - 5x_4),$$

substituting this expression in the last two equations, and collecting like terms. The result, which you should verify on scratch paper, is the equivalent system,

$$
\begin{aligned}
6x_1 + 2x_2 - x_3 + 5x_4 &= -8, \\
x_2 + \tfrac{3}{2}x_3 + \tfrac{1}{2}x_4 &= 3, \\
-\tfrac{1}{3}x_2 - \tfrac{1}{3}x_3 - \tfrac{1}{3}x_4 &= -\tfrac{2}{3}.
\end{aligned}
$$

Now we pivot on the coefficient 1 by solving the second equation for x_2,

$$x_2 = 3 - \tfrac{3}{2}x_3 - \tfrac{1}{2}x_4,$$

substituting this expression for x_2 in the third equation, and collecting like terms. Again you should verify that the result is

$$
\begin{aligned}
6x_1 + 2x_2 - x_3 + 5x_4 &= -8, \\
x_2 + \tfrac{3}{2}x_3 + \tfrac{1}{2}x_4 &= 3, \\
\tfrac{1}{6}x_3 - \tfrac{1}{6}x_4 &= \tfrac{1}{3}.
\end{aligned}
$$