Matrix Theory

MATRIX THEORY

Joel N. Franklin

PROFESSOR OF APPLIED MATHEMATICS
CALIFORNIA INSTITUTE OF TECHNOLOGY

Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Current Printing (last digit):

10 9 8 7 6

Library of Congress Catalog Card Number: 68-16345
Printed in the United States of America

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

MATRIX THEORY

Prentice-Hall
Series in Applied Mathematics

ARSAC, *Fourier Transforms and the Theory of Distributions*
BATES AND DOUGLAS, *Programming Language/One*
BAUMANN, FELICIANO, BAUER, AND SAMELSON, *Introduction to Algol*
BOWLES, *Computers in Humanistic Research*
CESCHINO AND KUNTZMAN, *Numerical Solution of Initial Value Problems*
DESMONDE, *Computers and Their Uses*
DESMONDE, *Real-Time Data Processing Systems: Introductory Concepts*
DEUTSCH, *Nonlinear Transformations of Random Processes*
DRESHER, *Games of Strategy: Theory and Applications*
EVANS, WALLACE, AND SUTHERLAND, *Simulation Using Digital Computers*
FLORES, *Computer Programming*
FLORES, *Computer Software: Programming Systems for Digital Computers*
FRANKLIN, *Matrix Theory*
FORSYTHE AND MOLER, *Computer Solution of Linear Algebraic Systems*
GOLDEN, FORTRAN IV: *Programming and Computing*
GOLDEN AND LEICHUS, *IBM 360: Programming and Computing*
GROVE, *Brief Numerical Methods*
HAHN, *Theory and Application of Liapunov's Direct Method*
HARTMANIS AND STEARNS, *Algebraic Structure Theory of Sequential Machines*
HULL, *Introduction to Computing*
LEHNIGK, *Stability Theorems for Linear Motions*
LOUDEN, *Programming the IBM 1130 and 1800*
MCCORMICK AND SALVADORI, *Numerical Analysis and FORTRAN Programming*
MARTIN, *Design of Real-Time Computer Systems*
MARTIN, *Programming Real-Time Computer Systems*
MINSKY, *Computation, Finite and Infinite Machines*
MOORE, *Interval Analysis*
MORETTI, *Functions of a Complex Variable*
NIXON, *Handbook of LaPlace Transformation*
PIPES, *Matrix Methods for Engineering*
SAXON AND STEYER, *Basic Principles of Data Processing*
SCHULTZ, *Digital Processing: A System Orientation*
SNYDER, *Chebyshev Methods in Numerical Approximation*
STANTON, *Numerical Methods for Science and Engineering*
STERLING AND POLLACK, *Introduction to Statistical Data Processing*
STROUD AND SECREST, *Gaussian Quadrature Formulas*
TRAUB, *Iterative Methods for the Solution of Equations*
VARGA, *Matrix Iterative Analysis*
WACHSPRESS, *Iterative Solution to Elliptic Systems and Applications to the Neutron Diffusion Equations of Reactor Physics*
WILKINSON, *Rounding Errors in Algebraic Processes*
ZIEGLER, *Time-Sharing Data Processing Systems*

PREFACE

The widespread applicability of high-speed digital computers has made it necessary for every modern engineer, mathematician, or scientist to have a knowledge of matrix theory. The connection between digital computation and matrices is almost obvious. Matrices represent *linear* transformations from a *finite* set of numbers to another finite set of numbers. Since many important problems are *linear*, and since digital computers with a finite memory manipulate only *finite* sets of numbers, the solution of linear problems by digital computation usually involves matrices.

This book developed from a course on matrix theory which I have given at Caltech since 1957. The course has been attended by graduate students, seniors, and juniors majoring in mathematics, economics, science, or engineering. The course was originally designed to be a preparation for courses in numerical analysis; but as the attendance increased through the years, I modified the syllabus to make it as useful as possible for the many different purposes of the students. In many fields—mathematical economics, quantum physics, geophysics, electrical network synthesis, crystallography, and structural engineering, to name a few—it has become increasingly popular to formulate and to solve problems in terms of matrices.

Ten years ago there were few texts on matrices; now there are many texts, with different points of view. This text is meant to meet many different needs. Because the book is mathematically rigorous, it can be used by students of pure and applied mathematics. Because it is oriented towards applications, it can be used by students of engineering, science, and the social sciences. Because it contains the basic preparation in matrix theory required for numerical analysis, it can be used by students whose main interest is their future use of computers.

The book begins with a concise presentation of the theory of determinants. There follows a presentation of classical linear algebra, and then there is an optional chapter on the use of matrices to solve systems of linear differential equations. Next is a presentation of the most commonly used diagonalizations or triangularizations of Hermitian and non-Hermitian matrices. The following chapter presents a proof of the difficult and important matrix theorem of Jordan. Then there is a chapter on the variational principles and perturbation theory of matrices, which are used in applications and in numerical analysis. The book ends with a long chapter on matrix numerical analysis. This last chapter is an introduction to the subject of linear computations, which is discussed in depth in the advanced treatises of Householder, Varga, Wilkinson, and others.

The book presents certain topics which are relatively new in basic texts on matrix theory. There are sections on vector and matrix norms, on the condition-number of a matrix, on positive and irreducible matrices, on the numerical identification of stable matrices, and on the *QR* method for computing eigenvalues.

A course on matrix theory lasting between one and two academic quarters could be based on selections from the first six chapters. A full-year course could cover the entire book.

The book assumes very little mathematical preparation. Except for the single section on the continuous dependence of eigenvalues on matrices, the book assumes only a knowledge of elementary algebra and calculus. The book begins with the most elementary results about determinants, and it proceeds gradually to cover the basic preparation in matrix theory which is necessary for every modern mathematician, engineer, or scientist.

I wish to thank Dr. George Forsythe and Dr. Richard Dean for reading parts of the manuscript and for suggesting the inclusion of certain special topics.

JOEL N. FRANKLIN

Pasadena, California

NOTATION USED IN THIS BOOK

Some authors denote vectors by boldface (**x**, **y**, **z**), but we shall simply denote them by using lower-case English letters (x, y, z). If we wish to designate the components of a vector, x, we shall use subscripts; thus, $x_1, \ldots, x_n$ designates the components of x. Superscripts will be used to designate different vectors; thus, x^1, x^2, x^3 designates three vectors, and if there appears to be any chance of confusion with the powers of a scalar, we shall enclose the superscripts in parentheses—e.g., $x^{(1)}, x^{(2)}, x^{(3)}$. Thus, $x_1^{(2)}, \ldots, x_n^{(2)}$ designates the n components of the vector $x^{(2)}$, whereas $x_1^2, \ldots, x_n^2$ designates the squares of the n components of a vector x.

Matrices will be denoted by capital letters (A, B, C). Subscripts may be used to designate different matrices—e.g., $A_1, A_2, A_3, \ldots$—but A, A^2, A^3 will designate different powers of the same square matrix, A. Thus,

$$A^2 = A \cdot A, \qquad A^3 = A \cdot A \cdot A, \qquad A^4 = A \cdot A \cdot A \cdot A, \ldots$$

A lower case Greek letter—e.g., λ, α, γ—will always designate an ordinary real or complex number. A Greek letter will never be used to designate a vector or a matrix. We will also occasionally use subscripted English letters—e.g., $c_1, \ldots, c_n$—to designate real or complex numbers. Subscripted English letters will *never* be used to designate vectors; thus, z_3 cannot designate a vector, although it may be used to designate the third component of a vector z.

For the columns of a matrix A we will often use the notation $a^{(1)}, \ldots, a^{(n)}$ or simply $a^1, \ldots, a^n$. For the components of a matrix A we will write a_{ij}. Sometimes we shall write

$$A = (a_{ij}) \qquad \begin{array}{l} i = 1, \ldots, m \\ j = 1, \ldots, n \end{array}$$

to indicate that A is the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

If A has the same number, n, of rows and columns, we may write

$$A = (a_{ij}) \qquad i, j = 1, \ldots, n$$

The rows of a matrix are horizontal; the columns are vertical. Thus,

$$[a_{31}, \ldots, a_{3n}]$$

may be the third row of a matrix, whereas

$$\begin{bmatrix} a_{17} \\ \cdot \\ \cdot \\ \cdot \\ a_{m7} \end{bmatrix}$$

may be the seventh column. Since a column vector takes much space to print, we will sometimes refer to it by the prefix "col." Thus, the preceding vector may be written as col $(a_{17}, \ldots, a_{m7})$.

The superscripted letter e has a particular meaning throughout the book. The vector e^j is a column-vector with its jth component equal to 1 and with all other components equal to 0. Thus, if there are five components,

$$e^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad e^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad e^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad e^4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad e^5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

CONTENTS

1 DETERMINANTS

1.1 INTRODUCTION

Suppose that we wish to solve the equations

$$\begin{aligned} 2x_1 + 7x_2 &= 4 \\ 3x_1 + 8x_2 &= 5 \end{aligned} \tag{1}$$

To solve for x_1 we multiply the first equation by 8, multiply the second equation by 7, and subtract the resulting equations. This procedure gives

$$[(8\cdot 2) - (7\cdot 3)]x_1 = (8\cdot 4) - (7\cdot 5)$$

Thus, $x_1 = -3/-5 = \frac{3}{5}$. To find x_2, multiply the second equation by 2, the first equation by 3, and subtract. The result is

$$[(8\cdot 2) - (7\cdot 3)]x_2 = (2\cdot 5) - (3\cdot 4)$$

Thus $x_2 = -2/-5 = \frac{2}{5}$. To generalize these equations, write

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$$

In the example (1)

$$\begin{aligned} a_{11} &= 2, \qquad & a_{12} &= 7, \qquad & b_1 &= 4 \\ a_{21} &= 3, \qquad & a_{22} &= 8, \qquad & b_2 &= 5 \end{aligned}$$

Solving for x_1 and x_2, we find

$$\begin{aligned}(a_{11}a_{22} - a_{12}a_{21})x_1 &= b_1a_{22} - b_2a_{12} \\ (a_{11}a_{22} - a_{12}a_{21})x_2 &= a_{11}b_2 - a_{21}b_1\end{aligned} \tag{2}$$

If we introduce the notation

$$ab - cd = \begin{vmatrix} a & d \\ c & b \end{vmatrix} \tag{3}$$

formula (2) yields

$$\begin{aligned}x_1 &= \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{21}a_{12}} \\ x_2 &= \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{21}a_{12}}\end{aligned} \tag{4}$$

if the denominator in each fraction is not zero.

A rectangular collection of numbers is a matrix. Thus,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{bmatrix} \tag{5}$$

are matrices. So is the number 17—a single number is a 1×1 matrix. Systems of linear equations are completely specified by the matrices containing their coefficients and by their right-hand sides. *A vector is a matrix with only one column or only one row.* Thus

$$[a_{21}a_{22}] \quad \text{and} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{6}$$

are vectors. The first is a *row vector*; the second, a *column vector*.

A determinant is a single number computed from a square matrix. We speak of "the determinant of a matrix," and in the next section, we shall define the determinant of an $n \times n$ matrix. For $n = 2$ we define

$$\det \begin{bmatrix} a & d \\ c & b \end{bmatrix} = ab - cd = \begin{vmatrix} a & d \\ c & b \end{vmatrix} \tag{7}$$

For example, the collection of four coefficients

$$\begin{bmatrix} 2 & 7 \\ 3 & 8 \end{bmatrix}$$

from equations (1) is a *matrix* with a *determinant* that is the single number

$$\begin{vmatrix} 2 & 7 \\ 3 & 8 \end{vmatrix} = 16 - 21 = -5$$

As we saw in (4), determinants appear in the solution of linear equations.

1.2 THE DEFINITION OF A DETERMINANT

We wish to generalize our solution of two equations in two unknowns to n equations in n unknowns. For $n = 3$ write

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{1}$$

If x_2 and x_3 are eliminated, we can obtain by a long computation the formula

$$\Delta x_1 = \Delta_1 \tag{2}$$

where

$$\begin{aligned} \Delta = a_{11}a_{22}a_{33} &+ a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} \\ -a_{11}a_{23}a_{32} &- a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{aligned} \tag{3}$$

and where Δ_1 is found by substituting b_1, b_2, b_3, respectively, for a_{11}, a_{21}, a_{31} in the expression for Δ. A second enormous computation would yield

$$\Delta x_2 = \Delta_2 \tag{4}$$

where Δ_2 is found by substituting b_1, b_2, b_3, respectively, for a_{12}, a_{22}, a_{32} in the expression for Δ. A third computation would give

$$\Delta x_3 = \Delta_3 \tag{5}$$

where Δ_3 is found by substituting b_1, b_2, b_3, respectively, for a_{13}, a_{23}, a_{33}. In a later section these results will be proved and generalized. Here we only wish to motivate the definition of the determinant.

Every term in the expression for Δ has the form $a_{1j}a_{2k}a_{3l}$ where j, k, l is a permutation of 1, 2, 3. With each term there is a sign $\pm 1 = s(j, k, l)$, which we call the *sign of the permutation* j, k, l. Thus

$$\begin{array}{lll} s(1, 2, 3) = 1, & s(3, 1, 2) = 1, & s(2, 3, 1) = 1 \\ s(1, 3, 2) = -1, & s(2, 1, 3) = -1, & s(3, 2, 1) = -1 \end{array} \tag{6}$$

Evidently,

$$\begin{array}{ll} s(j, k, l) = 1 & \text{if } (k - j)(l - j)(l - k) > 0 \\ s(j, k, l) = -1 & \text{if } (k - j)(l - j)(l - k) < 0 \end{array} \tag{7}$$

Now (3) takes the form

$$\Delta = \sum_{(j,k,l)} s(j, k, l) a_{1j}a_{2k}a_{3l} \tag{8}$$

where the summation extends over all six permutations j, k, l.

For an $n \times n$ matrix $(a_{ij})(i, j = 1, \ldots, n)$ we define the determinant Δ as the sum of $n!$ terms

$$\Delta = \sum_{(j_1, \ldots, j_n)} s(j_1, \ldots, j_n) a_{1j_1} a_{2j_2} \ldots a_{nj_n} \tag{9}$$

The summation extends over all $n!$ permutations $j_1, \ldots, j_n$ of $1, \ldots, n$. The sign $s(j_1, \ldots, j_n)$ is defined as

$$s(j_1, \ldots, j_n) = \text{sign} \prod_{1 < p < q < n} (j_q - j_p) \tag{10}$$

In other words, $s = 1$ if the product of all $j_q - j_p$ for $q > p$ is positive; $s = -1$ if the product is negative.

For example, the determinant of a 5×5 matrix has 120 terms. One of the terms is $-a_{13}a_{25}a_{31}a_{42}a_{54}$. The minus sign appears because

$$\begin{aligned} s(3, 5, 1, 2, 4) &= \text{sign } (5 - 3)\cdot(1 - 3)(1 - 5)\cdot(2 - 3)(2 - 5)(2 - 1)\cdot \\ &\qquad (4 - 3)(4 - 5)(4 - 1)(4 - 2) \\ &= (1)\cdot(-1)(-1)\cdot(-1)(-1)(1)\cdot(1)(-1)(1)(1) = -1 \end{aligned}$$

It should be emphasized that so far we have proved nothing. We merely have a definition (9) which we hope will be useful. For $n = 2$, the definition gives

$$\Delta = a_{11}a_{22} - a_{12}a_{21}$$

which is consistent with the definition given in the preceding section. For $n = 1$, we define $\Delta = a_{11}$, $s(1) = 1$.

PROBLEMS

1. Verify formula (7) for the six permutations in formula (6).
2. In the expansion (9) of the determinant of a 7×7 matrix there will be a term $\pm a_{17}a_{26}a_{35}a_{44}a_{53}a_{62}a_{71}$. Is the sign plus or is it minus?
3. Consider the equations

$$\begin{aligned} x_1 - x_2 + x_3 &= 2 \\ 2x_1 + x_2 - x_3 &= 0 \\ 3x_1 + 2x_2 - x_3 &= 4 \end{aligned}$$

Evaluate the determinants Δ, Δ_1, Δ_2, Δ_3 and solve for x_1, x_2, x_3 by formulas (2), (4), and (5).

1.3 PROPERTIES OF DETERMINANTS

The definitions (9) and (10) in the last section show that, to study determinants, we must study permutations.

Theorem 1. *If two numbers in the permutation* $j_1, \ldots, j_n$ *are interchanged, the sign of the permutation is reversed.* For example,

$$s(5, \mathbf{1}, 3, 2, \mathbf{4}) = -s(5, \mathbf{4}, 3, 2, \mathbf{1})$$

Proof. Suppose that the two numbers are adjacent, say k and l, in the permutation $j_1, \ldots, k, l, \ldots, j_n$. When k and l are interchanged, the product $\prod(j_s - j_r)$ for $s > r$ is unchanged except that the single term $l - k$ becomes $k - l$. Therefore, the sign is reversed.

If k and l are not adjacent, let them be separated by m numbers. Move k to the right by m successive interchanges of adjacent numbers so that k is just to the left of l. Now move l to the left by $m + 1$ successive interchanges of adjacent numbers. The effect of these interchanges is simply to interchange k and l in the original permutation. Since the sign is reversed an odd number of times, namely $m + m + 1$ times, the sign is reversed when k and l are interchanged.

Theorem 2. *Let the permutation* $j_1, \ldots, j_n$ *be formed from* $1, 2, \ldots, n$ *by* t *successive interchanges of pairs of numbers. Then* $s(j_1, \ldots, j_n) = (-1)^t$.

Proof. According to the last theorem, the sign of the permutation $1, 2, \ldots, n$ is reversed t times as the permutation $j_1, \ldots, j_n$ is formed. The result follows because $s(1, 2, \ldots, n) = 1$. For example,

$$s(4, 3, 2, 1) = -s(1, 3, 2, 4) = s(1, 2, 3, 4) = +1$$