Computer Science and Applied Mathematics

THE THEORY OF MATRICES

SECOND EDITION

WITH APPLICATIONS

Peter Lancaster and Miron Tismenetsky

The Theory of Matrices
Second Edition
with Applications

This is a volume in
COMPUTER SCIENCE AND APPLIED MATHEMATICS
A Series of Monographs and Textbooks
Editor: WERNER RHEINBOLDT
A complete list of titles in this series appears at the end of this volume.

The Theory of Matrices Second Edition

with Applications

Peter Lancaster
Department of Mathematics
University of Calgary
Calgary, Alberta, Canada

Miron Tismenetsky
IBM Scientific Center
Technion City
Haifa, Israel

1985

ACADEMIC PRESS

(Harcourt Brace Jovanovich, Publishers)

Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Lancaster, Peter, Date
The theory of matrices.

(Computer science and applied mathematics)
Includes index.
I. Matrices. I. Tismenetsky, M. (Miron) II. Title.
III. Series.
QA188.L36 1984 512.9'434 83-15775
ISBN 0-12-435560-9 (alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9 8 7 6 5 4 3 2 1

To our wives,
Edna and Fania

This page intentionally left blank

Contents

6 The Jordan Canonical Form: A Geometric Approach

7 Matrix Polynomials and Normal Forms

8 The Variational Method

9 Functions of Matrices

This page intentionally left blank

Preface

In this book the authors try to bridge the gap between the treatments of matrix theory and linear algebra to be found in current textbooks and the mastery of these topics required to use and apply our subject matter in several important areas of application, as well as in mathematics itself. At the same time we present a treatment that is as self-contained as is reasonably possible, beginning with the most fundamental ideas and definitions. In order to accomplish this double purpose, the first few chapters include a complete treatment of material to be found in standard courses on matrices and linear algebra. This part includes development of a computational algebraic development (in the spirit of the first edition) and also development of the abstract methods of finite-dimensional linear spaces. Indeed, a balance is maintained through the book between the two powerful techniques of matrix algebra and the theory of linear spaces and transformations.

The later chapters of the book are devoted to the development of material that is widely useful in a great variety of applications. Much of this has become a part of the language and methodology commonly used in modern science and engineering. This material includes variational methods, perturbation theory, generalized inverses, stability theory, and so on, and has innumerable applications in engineering, physics, economics, and statistics, to mention a few.

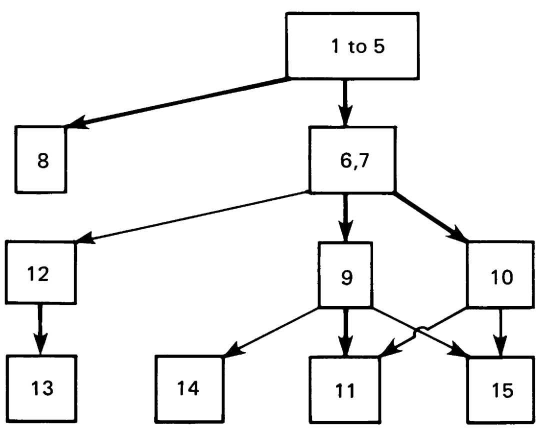
Beginning in Chapter 4 a few areas of application are developed in some detail. First and foremost we refer to the solution of constant-coefficient systems of differential and difference equations. There are also careful developments of the first steps in the theory of vibrating systems, Markov processes, and systems theory, for example.

The book will be useful for readers in two broad categories. One consists of those interested in a thorough reference work on matrices and linear algebra for use in their scientific work, whether in diverse applications or in mathematics itself. The other category consists of undergraduate or graduate students in a variety of possible programs where this subject matter is required. For example, foundations for courses offered in mathematics, computer science, or engineering programs may be found here. We address the latter audience in more detail.

The first seven chapters are essentially self-contained and require no formal prerequisites beyond college algebra. However, experience suggests that this material is most appropriately used as a second course in matrices or linear algebra at the sophomore or a more senior level.

There are possibilities for several different courses depending on specific needs and specializations. In general, it would not be necessary to work systematically through the first two chapters. They serve to establish notation, terminology, and elementary results, as well as some deeper results concerning determinants, which can be developed or quoted when required. Indeed, the first two chapters are written almost as compendia of primitive definitions, results, and exercises. Material for a traditional course in linear algebra, but with more emphasis on matrices, is then contained in Chapters 3–6, with the possibility of replacing Chapter 6 by Chapter 7 for a more algebraic development of the Jordan normal form including the theory of elementary divisors.

More advanced courses can be based on selected material from subsequent chapters. The logical connections between these chapters are indicated below to assist in the process of course design. It is assumed that in order to absorb any of these chapters the reader has a reasonable grasp of the first seven, as well as some knowledge of calculus. In this sketch the stronger connections are denoted by heavier lines.



Prerequisite structure by chapters.

There are many exercises and examples throughout the book. These range from computational exercises to assist the reader in fixing ideas, to extensions of the theory not developed in the text. In some cases complete solutions are given, and in others hints for solution are provided. These are seen as an integral part of the book and the serious reader is expected to absorb the information in them as well as that in the text.

In comparison with the 1969 edition of "The Theory of Matrices" by the first author, this volume is more comprehensive. First, the treatment of material in the first seven chapters (four chapters in the 1969 edition) is completely rewritten and includes a more thorough development of the theory of linear spaces and transformations, as well as the theory of determinants.

Chapters 8–11 and 15 (on variational methods, functions of matrices, norms, perturbation theory, and nonnegative matrices) retain the character and form of chapters of the first edition, with improvements in exposition and some additional material. Chapters 12–14 are essentially extra material and include some quite recent ideas and developments in the theory of matrices. A treatment of linear equations in matrices and generalized inverses that is sufficiently detailed for most applications is the subject of Chapter 12. It includes a complete description of commuting matrices. Chapter 13 is a thorough treatment of stability questions for matrices and scalar polynomials. The classical polynomial criteria of the nineteenth century are developed in a systematic and self-contained way from the more recent inertia theory of matrices. Chapter 14 contains an introduction to the recently developed spectral theory of matrix polynomials in sufficient depth for many applications, as well as providing access to the more general theory of matrix polynomials.

The greater part of this book was written while the second author was a Research Fellow in the Department of Mathematics and Statistics at the University of Calgary. Both authors are pleased to acknowledge support during this period from the University of Calgary. Many useful comments on the first edition are embodied in the second, and we are grateful to many colleagues and readers for providing them. Much of our work has been influenced by the enthusiasms of co-workers I. Gohberg, L. Rodman, and L. Lerer, and it is a pleasure to acknowledge our continued indebtedness to them. We would like to thank H. K. Wimmer for several constructive suggestions on an early draft of the second edition, as well as other colleagues, too numerous to mention by name, who made helpful comments.

The secretarial staff of the Department of Mathematics and Statistics at the University of Calgary has been consistently helpful and skillful in preparing the typescript for this second edition. However, Pat Dalgetty bore the brunt of this work, and we are especially grateful to her. During the period of production we have also benefitted from the skills and patience demonstrated by the staff of Academic Press. It has been a pleasure to work with them in this enterprise.

P. Lancaster M. Tismenetsky
Calgary Haifa

This page intentionally left blank

CHAPTER 1

Matrix Algebra

An ordered array of mn elements a_{ij} $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ written in the form

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

is said to be a *rectangular* $m \times n$ *matrix*. These elements can be taken from an arbitrary field $\mathscr{F}$. However, for the purposes of this book, $\mathscr{F}$ will always be the set of all real or all complex numbers, denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively.

A matrix A may be written more briefly in terms of its elements as

$$A = [a_{ij}]_{i,j=1}^{m,n}, \qquad \text{or} \qquad A = [a_{ij}],$$

where a_{ij} $(1 \le i \le m,\ 1 \le j \le n)$ denotes the element of the matrix lying on the intersection of the ith row and the jth column of A.

Two matrices having the same number of rows (m) and columns (n) are matrices of the *same size*. Matrices of the same size

$$A = [a_{ij}]_{i,j=1}^{m,n} \qquad \text{and} \qquad B = [b_{ij}]_{i,j=1}^{m,n}$$

are equal if and only if all the corresponding elements are identical, that is, $a_{ij} = b_{ij}$ for $1 \le i \le m$ and $1 \le j \le n$.

The set of all $m \times n$ matrices with real elements will be denoted by $\mathbb{R}^{m \times n}$. Similarly, $\mathbb{C}^{m \times n}$ is the set of all $m \times n$ matrices with complex elements.

1.1 Special Types of Matrices

If the number of rows of a matrix is equal to the number of columns, that is, $m = n$, then the matrix is *square* or *of order n*:

$$A = [a_{ij}]_{i,j=1}^{n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

The elements $a_{11}, a_{22}, \ldots, a_{nn}$ of a square matrix form its *main diagonal*, whereas the elements $a_{1n}, a_{2,n-1}, \ldots, a_{n1}$ generate the *secondary diagonal* of the matrix A.

Square matrices whose elements above (respectively, below) the main diagonal are zeros,

$$A_1 = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & a_{n,n-1} & a_{nn} \end{bmatrix}, \qquad A_2 = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix},$$

are called *lower-* (respectively, *upper-*) *triangular matrices.*

Diagonal matrices are a particular case of triangular matrices, for which all the elements lying outside the main diagonal are equal to zero:

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} = \operatorname{diag}[a_{11}, a_{22}, \ldots, a_{nn}].$$

If $a_{11} = a_{22} = \cdots = a_{nn} = a$, then the diagonal matrix A is called a *scalar matrix*;

$$A = \begin{bmatrix} a & 0 & \cdots & 0 \\ 0 & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a \end{bmatrix} = \operatorname{diag}[a, a, \ldots, a].$$

In particular, if $a = 1$, the matrix A becomes the *unit*, or *identity matrix*

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix},$$

and in the case $a = 0$, a square *zero-matrix*

$$0 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & & \ddots & \vdots \\ \vdots & \ddots & & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}$$

is obtained. A rectangular matrix with all its elements zero is also referred to as a zero-matrix.

A square matrix A is said to be a *Hermitian* (or *self-adjoint*) matrix if the elements on the main diagonal are real and whenever two elements are positioned symmetrically with respect to the main diagonal, they are mutually complex conjugate. In other words, Hermitian matrices are of the form

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \bar{a}_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & a_{nn} \end{bmatrix},$$

so that $a_{ji} = \bar{a}_{ij}$ for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, and $\bar{a}_{ij}$ denotes the complex conjugate of the number a_{ij}.

If all the elements located symmetrically with respect to the main diagonal are *equal*, then a square matrix is said to be *symmetric*:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}.$$

It is clear that, in the case of a *real matrix* (i.e., consisting of real numbers), the notions of Hermitian and symmetric matrices coincide.

Returning to rectangular matrices, note particularly those matrices having only one column (*column-matrix*) or one row (*row-matrix*) of length, or size, n:

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \qquad \boldsymbol{c}^{\mathrm{T}} = [c_1 \quad c_2 \quad \cdots \quad c_n].$$

The reason for the T symbol, denoting a row-matrix, will be made clear in Section 1.5.

Such $n \times 1$ and $1 \times n$ matrices are also referred to as *vectors* or *ordered n-tuples*, and in the cases $n = 1, 2, 3$ they have an obvious geometrical

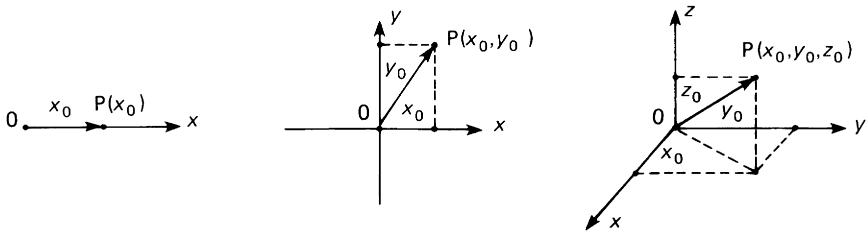


Fig. 1.1 Coordinates and position vectors.

meaning as the coordinates of a point P (or as components of the vector $\overrightarrow{OP}$) in one-, two-, or three-dimensional space with respect to the coordinate axes (Fig. 1.1).

For example, a point P in the three-dimensional Euclidean space, having Cartesian coordinates (x_0, y_0, z_0), and the vector $\overrightarrow{OP}$, are associated with the 1×3 row-matrix $[x_0 \quad y_0 \quad z_0]$. The location of the point P, as well as of the vector $\overrightarrow{OP}$, is described completely by this (*position*) vector.

Borrowing some geometrical language, the *length* of a vector (or position vector) is defined by the natural generalization of Euclidean geometry: for a vector $\boldsymbol{b}$ with elements $b_1, b_2, \ldots, b_n$ the length is

$$|\boldsymbol{b}| \triangleq (|b_1|^2 + |b_2|^2 + \cdots + |b_n|^2)^{1/2}.$$

Note that, throughout this book, the symbol $\triangleq$ is employed when a relation is used as a definition.

1.2 The Operations of Addition and Scalar Multiplication

Since vectors are special cases of matrices, the operations on matrices will be defined in such a way that, in the particular cases of column matrices and of row matrices, they correspond to the familiar operations on position vectors. Recall that, in three-dimensional Euclidean space, the sum of two position vectors is introduced as

$$[x_1 \quad y_1 \quad z_1] + [x_2 \quad y_2 \quad z_2] \triangleq [x_1 + x_2 \quad y_1 + y_2 \quad z_1 + z_2].$$

This definition yields the parallelogram law of vector addition, illustrated in Fig. 1.2.

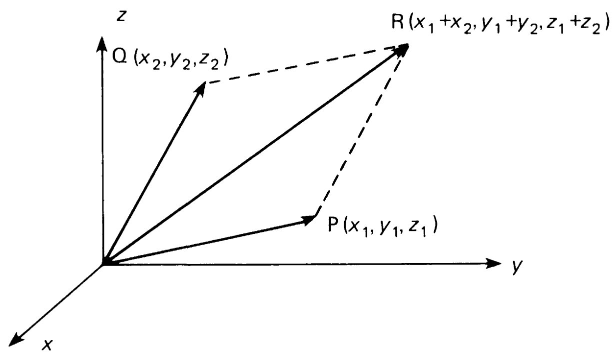


Fig. 1.2 The parallelogram law.

For ordered n-tuples written in the form of row- or column-matrices, this operation is naturally extended to

$$[x_1 \quad x_2 \quad \cdots \quad x_n] + [y_1 \quad y_2 \quad \cdots \quad y_n] \triangleq [x_1 + y_1 \quad x_2 + y_2 \quad \cdots \quad x_n + y_n]$$

or

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \triangleq \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

That is, the *elements* (or *components*, or *coordinates*) of the resulting vector are merely the sums of the corresponding elements of the vectors. Note that only vectors of the same size may be added.

Now the following definition of the sum of two matrices $A = [a_{ij}]_{i,j=1}^{m,n}$ and $B = [b_{ij}]_{i,j=1}^{m,n}$ of the same order is natural:

$$A + B \triangleq [a_{ij} + b_{ij}]_{i,j=1}^{m,n}.$$

The properties of the real and complex numbers (which we refer to as *scalars*) lead obviously to the *commutative* and *associative laws* of matrix addition.

Exercise 1. Show that, for matrices of the same size,

$$A + B = B + A,$$
$$(A + B) + C = A + (B + C). \quad \square$$

These rules allow easy definition and computation of the sum of several matrices of the same size. In particular, it is clear that the sum of any number of $n \times n$ upper- (respectively, lower-) triangular matrices is an upper- (respectively, lower-) triangular matrix. Note also that the sum of several diagonal matrices of the same order is a diagonal matrix.

The operation of subtraction on matrices is defined as for numbers. Namely, the *difference* of two matrices A and B of the same size, written $A - B$, is a matrix X that satisfies

$$X + B = A.$$

Obviously,

$$A - B = [a_{ij} - b_{ij}]_{i,j=1}^{m,n},$$

where

$$A = [a_{ij}]_{i,j=1}^{m,n}, \qquad B = [b_{ij}]_{i,j=1}^{m,n}.$$

It is clear that the zero-matrix plays the role of the zero in numbers: a matrix does not change if the zero-matrix is added to it or subtracted from it.

Before introducing the operation of multiplication of a matrix by a scalar, recall the corresponding definition for (position) vectors in three-dimensional Euclidean space: If $\boldsymbol{a}^{\mathrm{T}} = [a_1 \quad a_2 \quad a_3]$ and α denotes a real number, then the vector $\alpha\boldsymbol{a}^{\mathrm{T}}$ is defined by

$$\alpha\boldsymbol{a}^{\mathrm{T}} \triangleq [\alpha a_1 \quad \alpha a_2 \quad \alpha a_3].$$

Thus, in the product of a vector with a scalar, each element of the vector is multiplied by this scalar.

This operation has a simple geometrical meaning for real vectors and scalars (see Fig. 1.3). That is, the length of the vector $\alpha\boldsymbol{a}^{\mathrm{T}}$ is $|\alpha|$ times the length of the vector $\boldsymbol{a}^{\mathrm{T}}$, and its orientation does not change if $\alpha > 0$ and it reverses if $\alpha < 0$.

Passing from (position) vectors to the general case, the product of the matrix $A = [a_{ij}]$ with a scalar α is the matrix C with elements $c_{ij} = \alpha a_{ij}$, that is, $C \triangleq [\alpha a_{ij}]$. We also write $C = \alpha A$. The following properties of scalar

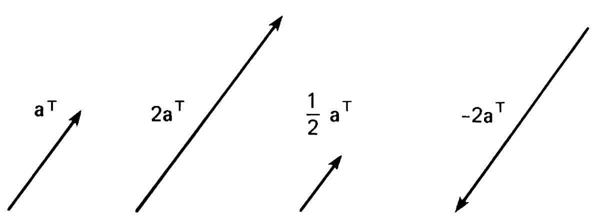


Fig. 1.3 Scalar multiplication.

multiplication and addition of matrices follow immediately from the definitions and from the corresponding properties of numbers.

Exercise 2. Check that for any two matrices of the same size and any scalars α, β

$$\begin{aligned} 0A &= 0, \\ (\alpha + \beta)A &= \alpha A + \beta A, \\ \alpha(A + B) &= \alpha A + \alpha B, \\ \alpha(\beta A) &= (\alpha\beta)A. \quad \square \end{aligned}$$

Note that by writing $(-1)B \triangleq -B$ we may, alternatively, define the difference $A - B$ as the sum of the matrices A and $-B$.

1.3 Matrix Multiplication

Like other operations on matrices, the notion of the product of two matrices can be motivated by some concepts of vector algebra. Recall that in three-dimensional Euclidean space, the *scalar* (or *dot*, or *inner*) *product* of the (position) vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is an important and useful concept. It is defined by

$$\boldsymbol{a} \cdot \boldsymbol{b} \triangleq |\boldsymbol{a}||\boldsymbol{b}| \cos \alpha, \tag{1}$$

where α denotes the angle between the given vectors. Clearly, for nonzero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the scalar product is equal to zero if and only if $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal.

If the coordinates a_1, a_2, a_3 and b_1, b_2, b_3 of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively, are known, then the computation of the scalar product is simple;

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 + a_3 b_3. \tag{2}$$

The crux of the definition of matrix multiplication lies in the convention of *defining* the product of *row vector* $\boldsymbol{a}^{\mathrm{T}}$ with *column vector* $\boldsymbol{b}$ (in this order) to be the sum on the right of Eq. (2). Thus,

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = [a_1 \quad a_2 \quad a_3]\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \triangleq a_1 b_1 + a_2 b_2 + a_3 b_3, \tag{3}$$

and this definition will now supersede the "dot" notation of Eqs. (1) or (2). Not unnaturally, it is described as "row-into-column" multiplication.

In contrast to Eq. (1), the definition in (3) lends itself immediately to broad generalizations that will be very useful indeed. First, the elements of $\boldsymbol{a}$ and $\boldsymbol{b}$ may be from any field (i.e., they are not necessarily real numbers), and second, the vectors may equally well have n elements each, for any positive integer n.

Thus, if $\boldsymbol{a}$, $\boldsymbol{b}$ are each vectors with n elements (the elements can be either real or complex numbers), define

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = [a_1 \quad a_2 \quad \cdots \quad a_n]\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \triangleq \sum_{r=1}^{n} a_r b_r. \tag{4}$$

This idea will now be exploited further in the full definition of matrix multiplication.

Let A be any matrix consisting of m rows $\boldsymbol{a}_i^{\mathrm{T}}$ $(1 \leq i \leq m)$ of length l (so A is $m \times l$) and let B denote a matrix containing n columns $\boldsymbol{b}_j$ $(1 \leq j \leq n)$ of the same length l (so that B is $l \times n$). Then the *product* of A and B is defined to be the $m \times n$ matrix $C = [c_{ij}]$, where

$$c_{ij} = \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{b}_j \qquad (1 \leq i \leq m, \quad 1 \leq j \leq n).$$

Thus

$$AB = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_m^{\mathrm{T}} \end{bmatrix} [\boldsymbol{b}_1 \quad \boldsymbol{b}_2 \quad \cdots \quad \boldsymbol{b}_n]$$

$$\triangleq \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_1 & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_n \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b}_1 & \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b}_n \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{b}_1 & \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{b}_n \end{bmatrix} \equiv [\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{b}_j]_{i,j=1}^{m,n}.$$

Using (4) we get the multiplication formula

$$AB = \left[\sum_{k=1}^{l} a_{ik} b_{kj}\right]_{i,j=1}^{m,n}. \tag{5}$$

Exercise 1. Show that

$$\begin{bmatrix} 2 & -1 & 0 \\ 0 & 3 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 7 & 3 \end{bmatrix}. \quad \square$$

It is important to note that the product of two matrices is defined if and only if the number of columns of the first factor is equal to the number of rows of the second. When this is the case, A and B are said to be *conformable* (with respect to matrix multiplication).

Exercise 2. If $\boldsymbol{a}$ is a column matrix of size m and $\boldsymbol{b}^{\mathrm{T}}$ is a row matrix of size n show that $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$ is an $m \times n$ matrix. (This matrix is sometimes referred to as an *outer product* of $\boldsymbol{a}$ and $\boldsymbol{b}$.) ☐

The above definition of matrix multiplication is standard; it focusses on the *rows* of the first factor and the *columns* of the second. There is another representation of the product AB that is frequently useful and, in contrast, is written in terms of the columns of the first factor and the rows of the second.

First note carefully the conclusion of Exercise 2. Then let A be $m \times l$ and B be $l \times n$ (so that A and B are conformable for matrix multiplication). Suppose that the columns of A are $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_l$ and the rows of B are $\boldsymbol{b}_1^{\mathrm{T}}, \boldsymbol{b}_2^{\mathrm{T}}, \ldots, \boldsymbol{b}_l^{\mathrm{T}}$. Then it is claimed that

$$AB = \sum_{k=1}^{l} \boldsymbol{a}_k \boldsymbol{b}_k^{\mathrm{T}}, \tag{6}$$

and of course each product $\boldsymbol{a}_k \boldsymbol{b}_k^{\mathrm{T}}$ is an $m \times n$ matrix.

Exercise 3. Establish the formula in (6). ☐

Matrix multiplication preserves *some* properties of multiplication of scalars.

Exercise 4. Check that for any matrices of appropriate orders

$$AI = IA = A,$$

$$A(B + C) = AB + AC, \qquad (B + C)D = BD + CD \qquad \text{(distributive laws)},$$

$$A(BC) = (AB)C \qquad \text{(associative law)}. \quad \square$$

However, the matrix product does *not* retain some important properties enjoyed by multiplication of scalars, as the next exercises demonstrate.

Exercise 5. Show, setting

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 0 \end{bmatrix}$$

or

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix},$$

that $AB \neq BA$. ☐

Thus, matrix multiplication is, in general, *not commutative*.

Exercise 6. Show that $AB = 0$ does not imply that either A or B is a zero-matrix. (In other words, construct matrices $A \neq 0$, $B \neq 0$ for which $AB = 0$).

Exercise 7. Check that if $AB = AC$, $A \neq 0$, it does not always follow that $B = C$. □

A particular case of matrix multiplication is that of multiplication of a matrix by itself. Obviously, such multiplication is possible if and only if the matrix is square. So let A denote a square matrix and let p be any positive integer. The matrix

$$A^p \triangleq \underbrace{AA \cdots A}_{p \text{ times}}$$

is said to be the pth *power* of the matrix A, and A^0 is defined to be the identity matrix with the size of A. Note also that $A^1 \triangleq A$.

Exercise 8. Verify that

$$\begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix}^3 = \begin{bmatrix} 8 & -12 \\ 0 & 8 \end{bmatrix}.$$

Exercise 9. Prove the *exponent laws*;

$$A^p A^q = A^{p+q},$$
$$(A^p)^q = A^{pq},$$

where p, q are any nonnegative integers. □

1.4 Special Kinds of Matrices Related to Multiplication

Some types of matrices having special properties with regard to matrix multiplication are presented in this section. As indicated in Exercise 1.3.5, two matrices do not generally commute: $AB \neq BA$. But when equality does hold, we say that the matrix A *commutes* with B; a situation to be studied in some depth in section 12.4.

Exercise 1. Find all matrices that commute with the matrices

$$\text{(a)} \quad \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \qquad \text{(b)} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Answers.

(a) $\begin{bmatrix} d-c & 0 \\ c & d \end{bmatrix}$, (b) $\begin{bmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{bmatrix}$,

where a, b, c, d denote arbitrary scalars. □

It follows from Exercise 1.3.4 that the identity matrix commutes with any square matrix. It turns out that there exist other matrices having this property. A general description of such matrices is given in the next exercise. Following that, other special cases of commuting matrices are illustrated.

Exercise 2. Show that a matrix commutes with any square conformable matrix if and only if it is a scalar matrix.

Exercise 3. Prove that diagonal matrices of the same size commute.

Exercise 4. Show that (nonnegative integer) powers of the same square matrix commute.

Exercise 5. Prove that if a matrix A commutes with a diagonal matrix $\operatorname{diag}[a_1, a_2, \ldots, a_n]$, where $a_i \neq a_j$, $i \neq j$, then A is diagonal. □

Another important class of matrices is formed by those that satisfy the equation

$$A^2 = A. \tag{1}$$

Such a matrix A is said to be *idempotent*.

Exercise 6. Check that the matrices

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

are idempotent.

Exercise 7. Show that if A is idempotent, and if p is a positive integer, then $A^p = A$.

Exercise 8. Describe the class of 2×2 idempotent matrices. □

A square matrix A is called *nilpotent* if there is a positive integer p such that

$$A^p = 0. \tag{2}$$

Obviously, it follows from (2) that any integer kth power of A with $k \geq p$ is also a zero matrix. Hence the notion of the *least* integer p_0 satisfying (2) is reasonable and is referred to as the *degree of nilpotency* of the matrix A.

Exercise 9. Check that the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

is nilpotent with degree of nilpotency equal to 3.

Exercise 10. Prove that every upper-triangular matrix having zeros on the main diagonal is nilpotent.

Exercise 11. Find all nilpotent 2×2 matrices having degree of nilpotency equal to two.

Answer.

$$\begin{bmatrix} a & b \\ c & -a \end{bmatrix},$$

where $a^2 + bc = 0$ and not all of a, b, c are zero. □

The last type of matrix to be considered in this section is the class of *involutory matrices.* They are those square matrices satisfying the condition

$$A^2 = I. \tag{3}$$

Exercise 12. Check that for any angle θ,

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

is an involutory matrix.

Exercise 13. Show that any matrix having 1 or -1 in all positions on the secondary diagonal and zeros elsewhere is involutory.

Exercise 14. Show that a matrix A is involutory if and only if

$$(I - A)(I + A) = 0.$$

Exercise 15. Find all 2×2 involutory matrices.

Answer.

$$\begin{bmatrix} a & b \\ c & -a \end{bmatrix},$$

with $a^2 + bc = 1$, or

$$\begin{bmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{bmatrix}$$

with $\varepsilon = \pm 1$.

Exercise 16. Show that if A is an involutory matrix, then the matrix $B = \frac{1}{2}(I + A)$ is idempotent. $\square$

1.5 Transpose and Conjugate Transpose

In this section some possibilities for forming new matrices from a given one are discussed.

Given an $m \times n$ matrix $A = [a_{ij}]$, the $n \times m$ matrix obtained by interchanging the rows and columns of A is called the *transpose* of the matrix A and is denoted by A^{T}. In more detail, if

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \text{then} \quad A^{\mathrm{T}} \triangleq \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Note that the rows of A become the columns of A^{T} and the columns of A are the rows of its transpose. Furthermore, in view of this definition our notation $\boldsymbol{a}^{\mathrm{T}}$ for a row matrix in the previous sections just means that it is the transpose of the corresponding column matrix $\boldsymbol{a}$.

Exercise 1. The matrices

$$A = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$$

are mutually transposed matrices: $A = B^{\mathrm{T}}$ and $B = A^{\mathrm{T}}$. $\square$

The following properties of the transpose are simple consequences of the definition.

Exercise 2. Let A and B be any matrices of an appropriate order and let $\alpha \in \mathscr{F}$. Prove that

$$(A^{\mathrm{T}})^{\mathrm{T}} = A,$$

$$(\alpha A)^{\mathrm{T}} = \alpha A^{\mathrm{T}},$$

$$(A + B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}},$$

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}. \quad \square$$