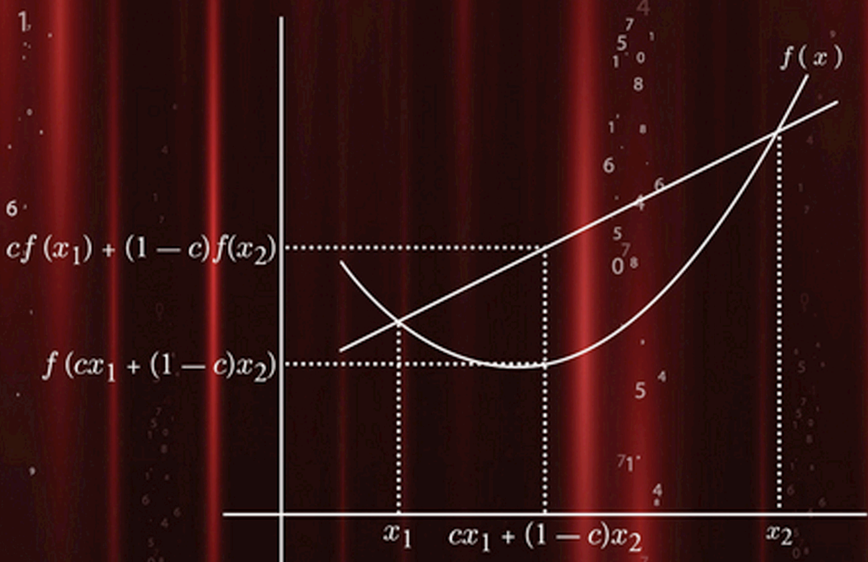
Wiley Series in Probability and Statistics

THIRD EDITION

MATRIX ANALYSIS FOR STATISTICS

James R. Schott



with website

WILEY

MATRIX ANALYSIS FOR STATISTICS

WILEY SERIES IN PROBABILITY AND STATISTICS

A complete list of the titles in this series appears at the end of this volume.

MATRIX ANALYSIS FOR STATISTICS

Third Edition

JAMES R. SCHOTT

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Names: Schott, James R., 1955- author.
Title: Matrix analysis for statistics / James R. Schott.
Description: Third edition. | Hoboken, New Jersey : John Wiley & Sons, 2016.
| Includes bibliographical references and index.
Identifiers: LCCN 2016000005| ISBN 9781119092483 (cloth) | ISBN 9781119092469
(epub)
Subjects: LCSH: Matrices. | Mathematical statistics.
Classification: LCC QA188 .S24 2016 | DDC 512.9/434–dc23 LC record available at
http://lccn.loc.gov/2016000005

Cover image courtesy of GettyImages/Alexmumu.

Typeset in 10/12pt TimesLTStd by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Susan, Adam, and Sarah

CONTENTS

PREFACE

As the field of statistics has developed over the years, the role of matrix methods has evolved from a tool through which statistical problems could be more conveniently expressed to an absolutely essential part in the development, understanding, and use of the more complicated statistical analyses that have appeared in recent years. As such, a background in matrix analysis has become a vital part of a graduate education in statistics. Too often, the statistics graduate student gets his or her matrix background in bits and pieces through various courses on topics such as regression analysis, multivariate analysis, linear models, stochastic processes, and so on. An alternative to this fragmented approach is an entire course devoted to matrix methods useful in statistics. This text has been written with such a course in mind. It also could be used as a text for an advanced undergraduate course with an unusually bright group of students and should prove to be useful as a reference for both applied and research statisticians.

Students beginning in a graduate program in statistics often have their previous degrees in other fields, such as mathematics, and so initially their statistical backgrounds may not be all that extensive. With this in mind, I have tried to make the statistical topics presented as examples in this text as self-contained as possible. This has been accomplished by including a section in the first chapter which covers some basic statistical concepts and by having most of the statistical examples deal with applications which are fairly simple to understand; for instance, many of these examples involve least squares regression or applications that utilize the simple concepts of mean vectors and covariance matrices. Thus, an introductory statistics course should provide the reader of this text with a sufficient background in statistics. An additional prerequisite is an undergraduate course in matrices or linear algebra, while a calculus background is necessary for some portions of the book, most notably, Chapter 8.

By selectively omitting some sections, all nine chapters of this book can be covered in a one-semester course. For instance, in a course targeted at students who end their educational careers with the masters degree, I typically omit Sections 2.10, 3.5, 3.7, 4.8, 5.4-5.7, and 8.6, along with a few other sections.

Anyone writing a book on a subject for which other texts have already been written stands to benefit from these earlier works, and that certainly has been the case here. The texts by Basilevsky (1983), Graybill (1983), Healy (1986), and Searle (1982), all books on matrices for statistics, have helped me, in varying degrees, to formulate my ideas on matrices. Graybill's book has been particularly influential, since this is the book that I referred to extensively, first as a graduate student, and then in the early stages of my research career. Other texts which have proven to be quite helpful are Horn and Johnson (1985, 1991), Magnus and Neudecker (1988), particularly in the writing of Chapter 8, and Magnus (1988).

I wish to thank several anonymous reviewers who offered many very helpful suggestions, and Mark Johnson for his support and encouragement throughout this project. I am also grateful to the numerous students who have alerted me to various mistakes and typos in earlier versions of this book. In spite of their help and my diligent efforts at proofreading, undoubtedly some mistakes remain, and I would appreciate being informed of any that are spotted.

JIM SCHOTT

Orlando, Florida

PREFACE TO THE SECOND EDITION

The most notable change in the second edition is the addition of a chapter on results regarding matrices partitioned into a 2×2 form. This new chapter, which is Chapter 7, has the material on the determinant and inverse that was previously given as a section in Chapter 7 of the first edition. Along with the results on the determinant and inverse of a partitioned matrix, I have added new material in this chapter on the rank, generalized inverses, and eigenvalues of partitioned matrices.

The coverage of eigenvalues in Chapter 3 has also been expanded. Some additional results such as Weyl's Theorem have been included, and in so doing, the last section of Chapter 3 of the first edition has now been replaced by two sections.

Other smaller additions, including both theorems and examples, have been made elsewhere throughout the book. Over 100 new exercises have been added to the problems sets.

The writing of a second edition of this book has also given me the opportunity to correct mistakes in the first edition. I would like to thank those readers who have pointed out some of these errors as well as those that have offered suggestions for improvement to the text.

Jim Schott

Orlando, Florida
September 2004

PREFACE TO THE THIRD EDITION

The third edition of this text maintains the same organization that was present in the previous editions. The major changes involve the addition of new material. This includes the following additions.

1. A new chapter, now Chapter 10, on inequalities has been added. Numerous inequalities such as Cauchy-Schwarz, Hadamard, and Jensen's, already appear in the earlier editions, but there are many important ones that are missing, and some of these are given in the new chapter. Highlighting this chapter is a fairly substantial section on majorization and some of the inequalities that can be developed from this concept.
2. A new section on oblique projections has been added to Chapter 2. The previous editions only covered orthogonal projections.
3. A new section on antieigenvalues and antieigenvectors has been added to Chapter 3.

Numerous other smaller additions have been made throughout the text. These include some additional theorems, the proofs of some results that previously had been given without proof, and some more examples involving statistical applications. Finally, more than 70 new problems have been added to the end-of-chapter problem sets.

Jim Schott

Orlando, Florida
December 2015

ABOUT THE COMPANION WEBSITE

This book is accompanied by a companion website:

www.wiley.com/go/Schott/MatrixAnalysis3e

The instructor's website includes:

- A solutions manual with solutions to selected problems

The student's website includes:

- A solutions manual with odd-numbered solutions to selected problems

1

A REVIEW OF ELEMENTARY MATRIX ALGEBRA

1.1 INTRODUCTION

In this chapter, we review some of the basic operations and fundamental properties involved in matrix algebra. In most cases, properties will be stated without proof, but in some cases, when instructive, proofs will be presented. We end the chapter with a brief discussion of random variables and random vectors, expected values of random variables, and some important distributions encountered elsewhere in the book.

1.2 DEFINITIONS AND NOTATION

Except when stated otherwise, a scalar such as α will represent a real number. A matrix A of size $m \times n$ is the $m \times n$ rectangular array of scalars given by

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

and sometimes it is simply identified as $A = (a_{ij})$. Sometimes it also will be convenient to refer to the (i, j)th element of A, as $(A)_{ij}$; that is, $a_{ij} = (A)_{ij}$. If $m = n$,

Matrix Analysis for Statistics, Third Edition. James R. Schott.

Companion Website: www.wiley.com/go/Schott/MatrixAnalysis3e

then A is called a square matrix of order m, whereas A is referred to as a rectangular matrix when $m \neq n$. An $m \times 1$ matrix

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

is called a column vector or simply a vector. The element a_i is referred to as the ith component of $\boldsymbol{a}$. A $1 \times n$ matrix is called a row vector. The ith row and jth column of the matrix A will be denoted by $(A)_{i\cdot}$ and $(A)_{\cdot j}$, respectively. We will usually use capital letters to represent matrices and lowercase bold letters for vectors.

The diagonal elements of the $m \times m$ matrix A are $a_{11}, a_{22}, \ldots, a_{mm}$. If all other elements of A are equal to 0, A is called a diagonal matrix and can be identified as $A = \text{diag}(a_{11}, \ldots, a_{mm})$. If, in addition, $a_{ii} = 1$ for $i = 1, \ldots, m$ so that $A = \text{diag}(1, \ldots, 1)$, then the matrix A is called the identity matrix of order m and will be written as $A = I_m$ or simply $A = I$ if the order is obvious. If $A = \text{diag}(a_{11}, \ldots, a_{mm})$ and b is a scalar, then we will use A^b to denote the diagonal matrix $\text{diag}(a_{11}^b, \ldots, a_{mm}^b)$. For any $m \times m$ matrix A, D_A will denote the diagonal matrix with diagonal elements equal to those of A, and for any $m \times 1$ vector $\boldsymbol{a}$, $D_{\boldsymbol{a}}$ denotes the diagonal matrix with diagonal elements equal to the components of $\boldsymbol{a}$; that is, $D_A = \text{diag}(a_{11}, \ldots, a_{mm})$ and $D_{\boldsymbol{a}} = \text{diag}(a_1, \ldots, a_m)$.

A triangular matrix is a square matrix that is either an upper triangular matrix or a lower triangular matrix. An upper triangular matrix is one that has all of its elements below the diagonal equal to 0, whereas a lower triangular matrix has all of its elements above the diagonal equal to 0. A strictly upper triangular matrix is an upper triangular matrix that has each of its diagonal elements equal to 0. A strictly lower triangular matrix is defined similarly.

The ith column of the $m \times m$ identity matrix will be denoted by $\boldsymbol{e}_i$; that is, $\boldsymbol{e}_i$ is the $m \times 1$ vector that has its ith component equal to 1 and all of its other components equal to 0. When the value of m is not obvious, we will make it more explicit by writing $\boldsymbol{e}_i$ as $\boldsymbol{e}_{i,m}$. The $m \times m$ matrix whose only nonzero element is a 1 in the (i, j)th position will be identified as E_{ij}.

The scalar zero is written 0, whereas a vector of zeros, called a null vector, will be denoted by $\mathbf{0}$, and a matrix of zeros, called a null matrix, will be denoted by (0). The $m \times 1$ vector having each component equal to 1 will be denoted by $\mathbf{1}_m$ or simply $\mathbf{1}$ when the size of the vector is obvious.

1.3 MATRIX ADDITION AND MULTIPLICATION

The sum of two matrices A and B is defined if they have the same number of rows and the same number of columns; in this case,

$$A + B = (a_{ij} + b_{ij}).$$

The product of a scalar α and a matrix A is

$$\alpha A = A\alpha = (\alpha a_{ij}).$$

The premultiplication of the matrix B by the matrix A is defined only if the number of columns of A equals the number of rows of B. Thus, if A is $m \times p$ and B is $p \times n$, then $C = AB$ will be the $m \times n$ matrix which has its (i, j)th element, c_{ij}, given by

$$c_{ij} = (A)_{i\cdot}(B)_{\cdot j} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

A similar definition exists for BA, the postmultiplication of B by A, if the number of columns of B equals the number of rows of A. When both products are defined, we will not have, in general, $AB = BA$. If the matrix A is square, then the product AA, or simply A^2, is defined. In this case, if we have $A^2 = A$, then A is said to be an idempotent matrix.

The following basic properties of matrix addition and multiplication in Theorem 1.1 are easy to verify.

Theorem 1.1 Let α and β be scalars and A, B, and C be matrices. Then, when the operations involved are defined, the following properties hold:

(a) $A + B = B + A$.
(b) $(A + B) + C = A + (B + C)$.
(c) $\alpha(A + B) = \alpha A + \alpha B$.
(d) $(\alpha + \beta)A = \alpha A + \beta A$.
(e) $A - A = A + (-A) = (0)$.
(f) $A(B + C) = AB + AC$.
(g) $(A + B)C = AC + BC$.
(h) $(AB)C = A(BC)$.

1.4 THE TRANSPOSE

The transpose of an $m \times n$ matrix A is the $n \times m$ matrix A' obtained by interchanging the rows and columns of A. Thus, the (i, j)th element of A' is a_{ji}. If A is $m \times p$ and B is $p \times n$, then the (i, j)th element of $(AB)'$ can be expressed as

$$((AB)')_{ij} = (AB)_{ji} = (A)_{j\cdot}(B)_{\cdot i} = \sum_{k=1}^{p} a_{jk} b_{ki}$$
$$= (B')_{i\cdot}(A')_{\cdot j} = (B'A')_{ij}.$$

Thus, evidently $(AB)' = B'A'$. This property along with some other results involving the transpose are summarized in Theorem 1.2.

Theorem 1.2 Let α and β be scalars and A and B be matrices. Then, when defined, the following properties hold:

(a) $(\alpha A)' = \alpha A'$.
(b) $(A')' = A$.
(c) $(\alpha A + \beta B)' = \alpha A' + \beta B'$.
(d) $(AB)' = B'A'$.

If A is $m \times m$, that is, A is a square matrix, then A' is also $m \times m$. In this case, if $A = A'$, then A is called a symmetric matrix, whereas A is called a skew-symmetric if $A = -A'$.

The transpose of a column vector is a row vector, and in some situations, we may write a matrix as a column vector times a row vector. For instance, the matrix E_{ij} defined in Section 1.2 can be expressed as $E_{ij} = \boldsymbol{e}_i \boldsymbol{e}_j'$. More generally, $\boldsymbol{e}_{i,m} \boldsymbol{e}_{j,n}'$ yields an $m \times n$ matrix having 1, as its only nonzero element, in the (i, j)th position, and if A is an $m \times n$ matrix, then

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \boldsymbol{e}_{i,m} \boldsymbol{e}_{j,n}'.$$

1.5 THE TRACE

The trace is a function that is defined only on square matrices. If A is an $m \times m$ matrix, then the trace of A, denoted by $\text{tr}(A)$, is defined to be the sum of the diagonal elements of A; that is,

$$\text{tr}(A) = \sum_{i=1}^{m} a_{ii}.$$

Now if A is $m \times n$ and B is $n \times m$, then AB is $m \times m$ and

$$\begin{aligned} \text{tr}(AB) &= \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} (A)_{i\cdot} (B)_{\cdot i} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij} = \sum_{j=1}^{n} (B)_{j\cdot} (A)_{\cdot j} \\ &= \sum_{j=1}^{n} (BA)_{jj} = \text{tr}(BA). \end{aligned}$$

This property of the trace, along with some others, is summarized in Theorem 1.3.

Theorem 1.3 Let α be a scalar and A and B be matrices. Then, when the appropriate operations are defined, we have the following properties:

(a) $\operatorname{tr}(A') = \operatorname{tr}(A)$.
(b) $\operatorname{tr}(\alpha A) = \alpha \operatorname{tr}(A)$.
(c) $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
(d) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.
(e) $\operatorname{tr}(A'A) = 0$ if and only if $A = (0)$.

1.6 THE DETERMINANT

The determinant is another function defined on square matrices. If A is an $m \times m$ matrix, then its determinant, denoted by $|A|$, is given by

$$\begin{aligned}|A| &= \sum (-1)^{f(i_1, \dots, i_m)} a_{1i_1} a_{2i_2} \cdots a_{mi_m} \\ &= \sum (-1)^{f(i_1, \dots, i_m)} a_{i_1 1} a_{i_2 2} \cdots a_{i_m m},\end{aligned}$$

where the summation is taken over all permutations $(i_1, \dots, i_m)$ of the set of integers $(1, \dots, m)$, and the function $f(i_1, \dots, i_m)$ equals the number of transpositions necessary to change $(i_1, \dots, i_m)$ to an increasing sequence of components, that is, to $(1, \dots, m)$. A transposition is the interchange of two of the integers. Although f is not unique, it is uniquely even or odd, so that $|A|$ is uniquely defined. Note that the determinant produces all products of m terms of the elements of the matrix A such that exactly one element is selected from each row and each column of A.

Using the formula for the determinant, we find that $|A| = a_{11}$ when $m = 1$. If A is 2×2, we have

$$|A| = a_{11}a_{22} - a_{12}a_{21},$$

and when A is 3×3, we get

$$\begin{aligned}|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.\end{aligned}$$

The following properties of the determinant in Theorem 1.4 are fairly straightforward to verify using the definition of a determinant.

Theorem 1.4 If α is a scalar and A is an $m \times m$ matrix, then the following properties hold:

(a) $|A'| = |A|$.
(b) $|\alpha A| = \alpha^m |A|$.

(c) If A is a diagonal matrix, then $|A| = a_{11} \cdots a_{mm} = \prod_{i=1}^{m} a_{ii}$.
(d) If all elements of a row (or column) of A are zero, $|A| = 0$.
(e) The interchange of two rows (or columns) of A changes the sign of $|A|$.
(f) If all elements of a row (or column) of A are multiplied by α, then the determinant is multiplied by α.
(g) The determinant of A is unchanged when a multiple of one row (or column) is added to another row (or column).
(h) If two rows (or columns) of A are proportional to one another, $|A| = 0$.

An alternative expression for $|A|$ can be given in terms of the cofactors of A. The minor of the element a_{ij}, denoted by m_{ij}, is the determinant of the $(m-1) \times (m-1)$ matrix obtained after removing the ith row and jth column from A. The corresponding cofactor of a_{ij}, denoted by A_{ij}, is then given as $A_{ij} = (-1)^{i+j} m_{ij}$.

Theorem 1.5 For any $i = 1, \ldots, m$, the determinant of the $m \times m$ matrix A can be obtained by expanding along the ith row,

$$|A| = \sum_{j=1}^{m} a_{ij} A_{ij}, \tag{1.1}$$

or expanding along the ith column,

$$|A| = \sum_{j=1}^{m} a_{ji} A_{ji}. \tag{1.2}$$

Proof. We will just prove (1.1), as (1.2) can easily be obtained by applying (1.1) to A'. We first consider the result when $i = 1$. Clearly

$$\begin{aligned} |A| &= \sum (-1)^{f(i_1, \ldots, i_m)} a_{1i_1} a_{2i_2} \cdots a_{mi_m} \\ &= a_{11} b_{11} + \cdots + a_{1m} b_{1m}, \end{aligned}$$

where

$$a_{1j} b_{1j} = \sum (-1)^{f(i_1, \ldots, i_m)} a_{1i_1} a_{2i_2} \cdots a_{mi_m},$$

and the summation is over all permutations for which $i_1 = j$. Since $(-1)^{f(j, i_2, \ldots, i_m)} = (-1)^{j-1} (-1)^{f(i_2, \ldots, i_m)}$, this implies that

$$b_{1j} = \sum (-1)^{j-1} (-1)^{f(i_2, \ldots, i_m)} a_{2i_2} \cdots a_{mi_m},$$

where the summation is over all permutations $(i_2, \ldots, i_m)$ of $(1, \ldots, j-1, j+1, \ldots, m)$. If C is the $(m-1) \times (m-1)$ matrix obtained from A by deleting its 1st row and jth column, then b_{1j} can be written

$$\begin{aligned} b_{1j} &= (-1)^{j-1} \sum (-1)^{f(i_1, \dots, i_{m-1})} c_{1i_1} \cdots c_{m-1 i_{m-1}} = (-1)^{j-1}|C| \\ &= (-1)^{j-1} m_{1j} = (-1)^{1+j} m_{1j} = A_{1j}, \end{aligned}$$

where the summation is over all permutations $(i_1, \dots, i_{m-1})$ of $(1, \dots, m-1)$ and m_{1j} is the minor of a_{1j}. Thus,

$$|A| = \sum_{j=1}^{m} a_{1j} b_{1j} = \sum_{j=1}^{m} a_{1j} A_{1j},$$

as is required. To prove (1.1) when $i > 1$, let D be the $m \times m$ matrix for which $(D)_{1\cdot} = (A)_{i\cdot}$, $(D)_{j\cdot} = (A)_{j-1\cdot}$, for $j = 2, \dots, i$, and $(D)_{j\cdot} = (A)_{j\cdot}$ for $j = i + 1, \dots, m$. Then $A_{ij} = (-1)^{i-1} D_{1j}$, $a_{ij} = d_{1j}$ and $|A| = (-1)^{i-1}|D|$. Thus, since we have already established (1.1) when $i = 1$, we have

$$|A| = (-1)^{i-1}|D| = (-1)^{i-1} \sum_{j=1}^{m} d_{1j} D_{1j} = \sum_{j=1}^{m} a_{ij} A_{ij},$$

and so the proof is complete. □

Our next result indicates that if the cofactors of a row or column are matched with the elements from a different row or column, the expansion reduces to 0.

Theorem 1.6 If A is an $m \times m$ matrix and $k \neq i$, then

$$\sum_{j=1}^{m} a_{ij} A_{kj} = \sum_{j=1}^{m} a_{ji} A_{jk} = 0. \tag{1.3}$$

Example 1.1 We will find the determinant of the 5×5 matrix given by

$$A = \begin{bmatrix} 2 & 1 & 2 & 1 & 1 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 1 \end{bmatrix}.$$

Using the cofactor expansion formula on the first column of A, we obtain

$$|A| = 2 \begin{vmatrix} 0 & 3 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \end{vmatrix},$$

and then using the same expansion formula on the first column of this 4×4 matrix, we get

$$|A| = 2(-1)\begin{vmatrix} 3 & 0 & 0 \\ 2 & 2 & 0 \\ 1 & 1 & 1 \end{vmatrix}.$$

Because the determinant of the 3×3 matrix above is 6, we have

$$|A| = 2(-1)(6) = -12.$$

Consider the $m \times m$ matrix C whose columns are given by the vectors $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$; that is, we can write $C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m)$. Suppose that, for some $m \times 1$ vector $\boldsymbol{b} = (b_1, \ldots, b_m)'$ and $m \times m$ matrix $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$, we have

$$\boldsymbol{c}_1 = A\boldsymbol{b} = \sum_{i=1}^{m} b_i \boldsymbol{a}_i.$$

Then, if we find the determinant of C by expanding along the first column of C, we get

$$\begin{aligned} |C| &= \sum_{j=1}^{m} c_{j1} C_{j1} = \sum_{j=1}^{m} \left(\sum_{i=1}^{m} b_i a_{ji} \right) C_{j1} \\ &= \sum_{i=1}^{m} b_i \left(\sum_{j=1}^{m} a_{ji} C_{j1} \right) = \sum_{i=1}^{m} b_i |(\boldsymbol{a}_i, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_m)|, \end{aligned}$$

so that the determinant of C is a linear combination of m determinants. If B is an $m \times m$ matrix and we now define $C = AB$, then by applying the previous derivation on each column of C, we find that

$$\begin{aligned} |C| &= \left| \left(\sum_{i_1=1}^{m} b_{i_1 1} \boldsymbol{a}_{i_1}, \ldots, \sum_{i_m=1}^{m} b_{i_m m} \boldsymbol{a}_{i_m} \right) \right| \\ &= \sum_{i_1=1}^{m} \cdots \sum_{i_m=1}^{m} b_{i_1 1} \cdots b_{i_m m} |(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})| \\ &= \sum b_{i_1 1} \cdots b_{i_m m} |(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})|, \end{aligned}$$

where this final sum is only over all permutations of $(1, \ldots, m)$, because Theorem 1.4(h) implies that

$$|(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})| = 0$$

if $i_j = i_k$ for any $j \neq k$. Finally, reordering the columns in $|(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})|$ and using Theorem 1.4(e), we have

$$|C| = \sum b_{i_1 1} \cdots b_{i_m m} (-1)^{f(i_1, \ldots, i_m)} |(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)| = |B||A|.$$

This very useful result is summarized in Theorem 1.7.

Theorem 1.7 If both A and B are square matrices of the same order, then

$$|AB| = |A||B|.$$

1.7 THE INVERSE

An $m \times m$ matrix A is said to be a nonsingular matrix if $|A| \neq 0$ and a singular matrix if $|A| = 0$. If A is nonsingular, a nonsingular matrix denoted by A^{-1} and called the inverse of A exists, such that

$$AA^{-1} = A^{-1}A = I_m. \tag{1.4}$$

This inverse is unique because, if B is another $m \times m$ matrix satisfying the inverse formula (1.4) for A, then $BA = I_m$, and so

$$B = BI_m = BAA^{-1} = I_m A^{-1} = A^{-1}.$$

The following basic properties of the matrix inverse in Theorem 1.8 can be easily verified by using (1.4).

Theorem 1.8 If α is a nonzero scalar, and A and B are nonsingular $m \times m$ matrices, then the following properties hold:

(a) $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$.
(b) $(A')^{-1} = (A^{-1})'$.
(c) $(A^{-1})^{-1} = A$.
(d) $|A^{-1}| = |A|^{-1}$.
(e) If $A = \text{diag}(a_{11}, \ldots, a_{mm})$, then $A^{-1} = \text{diag}(a_{11}^{-1}, \ldots, a_{mm}^{-1})$.
(f) If $A = A'$, then $A^{-1} = (A^{-1})'$.
(g) $(AB)^{-1} = B^{-1} A^{-1}$.

As with the determinant of A, the inverse of A can be expressed in terms of the cofactors of A. Let $A_{\#}$, called the adjoint of A, be the transpose of the matrix of cofactors of A; that is, the (i, j)th element of $A_{\#}$ is A_{ji}, the cofactor of a_{ji}. Then

$$AA_{\#} = A_{\#}A = \text{diag}(|A|, \ldots, |A|) = |A| I_m,$$

because $(A)_{i\cdot}(A_{\#})_{\cdot i} = (A_{\#})_{i\cdot}(A)_{\cdot i} = |A|$ follows directly from (1.1) and (1.2), and $(A)_{i\cdot}(A_{\#})_{\cdot j} = (A_{\#})_{i\cdot}(A)_{\cdot j} = 0$, for $i \neq j$ follows from (1.3). The equation above then yields the relationship

$$A^{-1} = |A|^{-1} A_{\#}$$

when $|A| \neq 0$. Thus, for instance, if A is a 2×2 nonsingular matrix, then

$$A^{-1} = |A|^{-1} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Similarly when $m = 3$, we get $A^{-1} = |A|^{-1} A_{\#}$, where

$$A_{\#} = \begin{bmatrix} a_{22}a_{33} - a_{23}a_{32} & -(a_{12}a_{33} - a_{13}a_{32}) & a_{12}a_{23} - a_{13}a_{22} \\ -(a_{21}a_{33} - a_{23}a_{31}) & a_{11}a_{33} - a_{13}a_{31} & -(a_{11}a_{23} - a_{13}a_{21}) \\ a_{21}a_{32} - a_{22}a_{31} & -(a_{11}a_{32} - a_{12}a_{31}) & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix}.$$

The relationship between the inverse of a matrix product and the product of the inverses, given in Theorem 1.8(g), is a very useful property. Unfortunately, such a nice relationship does not exist between the inverse of a sum and the sum of the inverses. We do, however, have Theorem 1.9 which is sometimes useful.

Theorem 1.9 Suppose A and B are nonsingular matrices, with A being $m \times m$ and B being $n \times n$. For any $m \times n$ matrix C and any $n \times m$ matrix D, it follows that if $A + CBD$ is nonsingular, then

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}.$$

Proof. The proof simply involves verifying that $(A + CBD)(A + CBD)^{-1} = I_m$ for $(A + CBD)^{-1}$ given above. We have

$$\begin{aligned} &(A + CBD)\{A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}\} \\ &\quad = I_m - C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} + CBDA^{-1} \\ &\qquad - CBDA^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \\ &\quad = I_m - C\{(B^{-1} + DA^{-1}C)^{-1} - B \\ &\qquad + BDA^{-1}C(B^{-1} + DA^{-1}C)^{-1}\}DA^{-1} \\ &\quad = I_m - C\{B(B^{-1} + DA^{-1}C)(B^{-1} + DA^{-1}C)^{-1} - B\}DA^{-1} \\ &\quad = I_m - C\{B - B\}DA^{-1} = I_m, \end{aligned}$$

and so the result follows. □

The expression given for $(A + CBD)^{-1}$ in Theorem 1.9 involves the inverse of the matrix $B^{-1} + DA^{-1}C$. It can be shown (see Problem 7.12) that the conditions of the theorem guarantee that this inverse exists. If $m = n$ and C and D are identity matrices, then we obtain Corollary 1.9.1 of Theorem 1.9.

Corollary 1.9.1 Suppose that A, B and $A+B$ are all $m \times m$ nonsingular matrices. Then

$$(A+B)^{-1} = A^{-1} - A^{-1}(B^{-1} + A^{-1})^{-1}A^{-1}.$$

We obtain Corollary 1.9.2 of Theorem 1.9 when $n = 1$.

Corollary 1.9.2 Let A be an $m \times m$ nonsingular matrix. If $\boldsymbol{c}$ and $\boldsymbol{d}$ are both $m \times 1$ vectors and $A + \boldsymbol{c}\boldsymbol{d}'$ is nonsingular, then

$$(A + \boldsymbol{c}\boldsymbol{d}')^{-1} = A^{-1} - A^{-1}\boldsymbol{c}\boldsymbol{d}'A^{-1}/(1 + \boldsymbol{d}'A^{-1}\boldsymbol{c}).$$

Example 1.2 Theorem 1.9 can be particularly useful when m is larger than n and the inverse of A is fairly easy to compute. For instance, suppose we have $A = I_5$,

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 1 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad D' = \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 0 & 1 \\ 1 & 0 \\ -1 & 1 \end{bmatrix},$$

from which we obtain

$$G = A + CBD = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ -1 & 6 & 4 & 3 & 1 \\ -1 & 2 & 2 & 0 & 1 \\ -2 & 6 & 4 & 3 & 2 \\ -1 & 4 & 3 & 2 & 2 \end{bmatrix}.$$

It is somewhat tedious to compute the inverse of this 5×5 matrix directly. However, the calculations in Theorem 1.9 are fairly straightforward. Clearly, $A^{-1} = I_5$ and

$$B^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix},$$

so that

$$(B^{-1} + DA^{-1}C) = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} -2 & 0 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 2 & 5 \end{bmatrix}$$

and

$$(B^{-1} + DA^{-1}C)^{-1} = \begin{bmatrix} 2.5 & 0.5 \\ -1 & 0 \end{bmatrix}.$$

Thus, we find that

$$G^{-1} = I_5 - C(B^{-1} + DA^{-1}C)^{-1}D$$

$$= \begin{bmatrix} -1 & 1.5 & -0.5 & -2.5 & 2 \\ -3 & 3 & -1 & -4 & 3 \\ 3 & -2.5 & 1.5 & 3.5 & -3 \\ 2 & -2 & 0 & 3 & -2 \\ -1 & 0.5 & -0.5 & -1.5 & 2 \end{bmatrix}.$$

1.8 PARTITIONED MATRICES

Occasionally we will find it useful to partition a given matrix into submatrices. For instance, suppose A is $m \times n$ and the positive integers m_1, m_2, n_1, n_2 are such that $m = m_1 + m_2$ and $n = n_1 + n_2$. Then one way of writing A as a partitioned matrix is

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is $m_1 \times n_1$, A_{12} is $m_1 \times n_2$, A_{21} is $m_2 \times n_1$, and A_{22} is $m_2 \times n_2$. That is, A_{11} is the matrix consisting of the first m_1 rows and n_1 columns of A, A_{12} is the matrix consisting of the first m_1 rows and last n_2 columns of A, and so on. Matrix operations can be expressed in terms of the submatrices of the partitioned matrix. For example, suppose B is an $n \times p$ matrix partitioned as

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where B_{11} is $n_1 \times p_1$, B_{12} is $n_1 \times p_2$, B_{21} is $n_2 \times p_1$, B_{22} is $n_2 \times p_2$, and $p = p_1 + p_2$. Then the premultiplication of B by A can be expressed in partitioned form as

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

Matrices can be partitioned into submatrices in other ways besides this 2×2 partitioned form. For instance, we could partition only the columns of A, yielding the expression

$$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix},$$

where A_1 is $m \times n_1$ and A_2 is $m \times n_2$. A more general situation is one in which the rows of A are partitioned into r groups and the columns of A are partitioned into c groups so that A can be written as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1c} \\ A_{21} & A_{22} & \cdots & A_{2c} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rc} \end{bmatrix},$$