

ELLIS HORWOOD SERIES IN MATHEMATICS AND ITS APPLICATIONS

Series Editor: Professor G. M. BELL, Chelsea College, University of London

The works in this series will survey recent research, and introduce new areas and up-to-date mathematical methods. Undergraduate texts on established topics will stimulate student interest by including present-day applications, and the series can also include selected volumes of lecture notes on important topics which need quick and early publication.

In all three ways it is hoped to render a valuable service to those who learn, teach, develop and use mathematics.

MATHEMATICAL THEORY OF WAVE MOTION

G. R. BALDOCK and T. BRIDGEMAN, University of Liverpool.

MATHEMATICAL MODELS IN SOCIAL MANAGEMENT AND LIFE SCIENCES

D. N. BURGHES and A. D. WOOD, Cranfield Institute of Technology.

MODERN INTRODUCTION TO CLASSICAL MECHANICS AND CONTROL

D. N. BURGHES, Cranfield Institute of Technology and A. DOWNS, Sheffield University.

CONTROL AND OPTIMAL CONTROL

D. N. BURGHES, Cranfield Institute of Technology and A. GRAHAM, The Open University, Milton Keynes.

TEXTBOOK OF DYNAMICS

F. CHORLTON, University of Aston, Birmingham.

VECTOR AND TENSOR METHODS

F. CHORLTON, University of Aston, Birmingham.

TECHNIQUES IN OPERATIONAL RESEARCH

VOLUME 1: QUEUEING SYSTEMS

VOLUME 2: MODELS, SEARCH, RANDOMIZATION

B. CONOLLY, Chelsea College, University of London

MATHEMATICS FOR THE BIOSCIENCES

G. EASON, C. W. COLES, G. GETTINBY, University of Strathclyde.

HANDBOOK OF HYPERGEOMETRIC INTEGRALS: Theory, Applications, Tables, Computer Programs

H. EXTON, The Polytechnic, Preston.

MULTIPLE HYPERGEOMETRIC FUNCTIONS

H. EXTON, The Polytechnic, Preston

COMPUTATIONAL GEOMETRY FOR DESIGN AND MANUFACTURE

I. D. FAUX and M. J. PRATT, Cranfield Institute of Technology.

APPLIED LINEAR ALGEBRA

R. J. GOULT, Cranfield Institute of Technology.

MATRIX THEORY AND APPLICATIONS FOR ENGINEERS AND MATHEMATICIANS

A. GRAHAM, The Open University, Milton Keynes.

APPLIED FUNCTIONAL ANALYSIS

D. H. GRIFFEL, University of Bristol.

GENERALISED FUNCTIONS: Theory, Applications

R. F. HOSKINS, Cranfield Institute of Technology.

MECHANICS OF CONTINUOUS MEDIA

S. C. HUNTER, University of Sheffield.

GAME THEORY: Mathematical Models of Conflict

A. J. JONES, Royal Holloway College, University of London.

USING COMPUTERS

B. L. MEEK and S. FAIRTHORNE, Queen Elizabeth College, University of London.

SPECTRAL THEORY OF ORDINARY DIFFERENTIAL OPERATORS

E. MULLER-PFEIFFER, Technical High School, Ergurt.

SIMULATION CONCEPTS IN MATHEMATICAL MODELLING

F. OLIVEIRA-PINTO, Chelsea College, University of London.

ENVIRONMENTAL AERODYNAMICS

R. S. SCORER, Imperial College of Science and Technology, University of London.

APPLIED STATISTICAL TECHNIQUES

K. D. C. STOODLEY, T. LEWIS and C. L. S. STANTON, University of Bradford.

LIQUIDS AND THEIR PROPERTIES: A Molecular and Macroscopic Treatise with Applications

H. N. V. TEMPERLEY, University College of Swansea, University of Wales and D. H. TREVENA, University of Wales, Aberystwyth.

GRAPH THEORY AND APPLICATIONS

H. N. V. TEMPERLEY, University College of Swansea.

Kronecker Products and Matrix Calculus: with Applications

ALEXANDER GRAHAM, M.A., M.Sc., Ph.D., C.Eng. M.I.E.E.
Senior Lecturer in Mathematics,
The Open University,
Milton Keynes



ELLIS HORWOOD LIMITED
Publishers · Chichester

Halsted Press: a division of
JOHN WILEY & SONS
New York · Brisbane · Chichester · Toronto

First published in 1981 by

ELLIS HORWOOD LIMITED

Market Cross House, Cooper Street, Chichester, West Sussex, PO19 1EB, England

The publisher's colophon is reproduced from James Gillison's drawing of the ancient Market Cross, Chichester.

Distributors:

Australia, New Zealand, South-east Asia:

Jacaranda-Wiley Ltd., Jacaranda Press,

JOHN WILEY & SONS INC.,

G.P.O. Box 859, Brisbane, Queensland 40001, Australia

Canada:

JOHN WILEY & SONS CANADA LIMITED

22 Worcester Road, Rexdale, Ontario, Canada.

Europe, Africa:

JOHN WILEY & SONS LIMITED

Baffins Lane, Chichester, West Sussex, England.

North and South America and the rest of the world:

Halsted Press: a division of

JOHN WILEY & SONS

605 Third Avenue, New York, N.Y. 10016, U.S.A.

© 1981 A. Graham/Ellis Horwood Ltd.

British Library Cataloguing in Publication Data

Graham, Alexander

Kronecker products and matrix calculus. —

(Ellis Horwood series in mathematics and its applications)

1. Matrices

I. Title

512.9'43 QA188

Library of Congress Card No. 81-7132 AACR2

ISBN 0-85312-391-8 (Ellis Horwood Limited, Library Edition)

ISBN 0-85312-427-2 (Ellis Horwood Limited, Student Edition)

ISBN 0-470-27300-3 (Halsted Press)

Typeset in Press Roman by Ellis Horwood Ltd.

Printed in Great Britain by R. J. Acford, Chichester

COPYRIGHT NOTICE —

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the permission of Ellis Horwood Limited, Market Cross House, Cooper Street, Chichester, West Sussex, England.

Table of Contents

Author's Preface	7
Symbols and Notation Used	9
Chapter 1 – Preliminaries	
1.1 Introduction.	11
1.2 Unit Vectors and Elementary Matrices.	11
1.3 Decompositions of a Matrix.	13
1.4 The Trace Function	16
1.5 The Vec Operator	18
Problems for Chapter 1.	20
Chapter 2 – The Kronecker Product	
2.1 Introduction.	21
2.2 Definition of the Kronecker Product.	21
2.3 Some Properties and Rules for Kronecker Products	23
2.4 Definition of the Kronecker Sum	30
2.5 The Permutation Matrix associating $\text{vec}X$ and $\text{vec}X'$	32
Problems for Chapter 2.	35
Chapter 3 – Some Applications for the Kronecker Product	
3.1 Introduction.	37
3.2 The Derivative of a Matrix.	37
3.3 Problem 1: solution of $AX + XB = C$	38
3.4 Problem 2: solution of $AX + XA = \mu X$	40
3.5 Problem 3: solution of $\dot{X} = AX + XB$	41
3.6 Problem 4: to find the transition matrix associated with the equation $\dot{X} = AX + XB$	42
3.7 Problem 5: solution of $AXB = C$	44
3.8 Problem 6: Pole assignment for a Multivariable System.	45

Table of Contents

Chapter 4 – Introduction to Matrix Calculus	
4.1 Introduction	51
4.2 The Derivatives of Vectors	52
4.3 The Chain rule for Vectors	54
4.4 The Derivative of Scalar Functions of a Matrix with respect to a Matrix	56
4.5 The Derivative of a Matrix with respect to one of its Elements and Conversely	60
4.6 The Derivatives of the Powers of a Matrix	67
Problems for Chapter 4	68
Chapter 5 – Further Development of Matrix Calculus including an Application of Kronecker Products	
5.1 Introduction	70
5.2 Derivatives of Matrices and Kronecker Products	70
5.3 The Determination of $(\partial \text{vec} X)/(\partial \text{vec} Y)$ for more complicated Equations	72
5.4 More on Derivatives of Scalar Functions with respect to a Matrix	75
5.5 The Matrix Differential	78
Problems for Chapter 5	80
Chapter 6 – The Derivative of a Matrix with respect to a Matrix	
6.1 Introduction	81
6.2 The Definition and some Results	81
6.3 Product Rules for Matrices	84
6.4 The Chain Rule for the Derivative of a Matrix with respect to Matrix	88
Problems for Chapter 6	92
Chapter 7 – Some Applications of Matrix Calculus	
7.1 Introduction	94
7.2 The Problems of Least Squares and Constrained Optimization in Scalar Variables	94
7.3 Problem 1: Matrix Calculus Approach to the Problems of Least Squares and Constrained Optimization	96
7.4 Problem 2: The General Least Squares Problem	100
7.5 Problem 3: Maximum Likelihood Estimate of the Multivariate Normal	102
7.6 Problem 4: Evaluation of the Jacobians of some Transformations	104
7.7 Problem 5: To Find the Derivative of an Exponential Matrix with respect to a Matrix	108
Solution to Problems	111
Tables of Formulae and Derivatives	121
Bibliography	126
Index	129

Author's Preface

My purpose in writing this book is to bring to the attention of the reader, some recent developments in the field of Matrix Calculus. Although some concepts, such as Kronecker matrix products, the vector derivative etc. are mentioned in a few specialised books, no book, to my knowledge, is totally devoted to this subject. The interested researcher must consult numerous published papers to appreciate the scope of the concepts involved.

Matrix calculus applicable to square matrices was developed by Turnbull [29,30] as far back as 1927. The theory presented in this book is based on the works of Dwyer and McPhail [15] published in 1948 and others mentioned in the Bibliography. It is more general than Turnbull's development and is applicable to non-square matrices. But even this more general theory has grave limitations, in particular it requires that in general the matrix elements are non constant and independent. A symmetric matrix, for example, is treated as a special case. Methods of overcoming some of these limitations have been suggested, but I am not aware of any published theory which is both quite general and simple enough to be useful.

The book is organised in the following way:

Chapter 1 concentrates on the preliminaries of matrix theory and notation which is found useful throughout the book. In particular, the simple and useful elementary matrix is defined. The vec operator is defined and many useful relations are developed. Chapter 2 introduces and establishes various important properties of the matrix Kronecker product.

Several applications of the Kronecker product are considered in Chapter 3. Chapter 4 introduces Matrix Calculus. Various derivatives of vectors are defined and the chain rule for vector differentiation is established. Rules for obtaining the derivative of a matrix with respect to one of its elements and conversely are discussed. Further developments in Matrix Calculus including derivatives of scalar functions of a matrix with respect to the matrix and matrix differentials are found in Chapter 5.

Chapter 6 deals with the derivative of a matrix with respect to a matrix.

Author's Preface

This includes the derivation of expressions for the derivatives of both the matrix product and the Kronecker product of matrices with respect to a matrix. There is also the derivation of a chain rule of matrix differentiation. Various applications of at least some of the matrix calculus are discussed in Chapter 7.

By making use, whenever possible, of simple notation, including many worked examples to illustrate most of the important results and other examples at the end of each Chapter (except for Chapters 3 and 7) with solutions at the end of the book, I have attempted to bring a topic studied mainly at post-graduate and research level to an undergraduate level.

Symbols and Notation Used

$A, B, C \dots$	matrices
A'	the transpose of A
a_{ij}	the (i, j) th element of the matrix A
$[a_{ij}]$	the matrix A having a_{ij} as its (i, j) th element
I_m	the unit matrix of order $m \times m$
e_i	the unit vector
e	the one vector (having all elements equal to one)
E_{ij}	the elementary matrix
0_m	the zero matrix of order $m \times m$
δ_{ij}	the Kronecker delta
$A_{.i}$	the i th column of the matrix A
A_j	the j th row of A as a column vector
A_j'	the transpose of A_j . (a row vector)
$(A')_{.i}$	the i th column of the matrix A'
$(A')'_i$	the transpose of the i th column of A' (that is, a row vector)
$\text{tr } A$	the trace of A
$\text{vec } A$	an ordered stock of columns of A
$A \otimes B$	the Kronecker product of A and B
iff	if and only if
$\text{diag } \{A\}$	the square matrix having elements a_{11}, a_{22}, \dots along its diagonal and zeros elsewhere
$\frac{\partial Y}{\partial x_{rs}}$	a matrix of the same order as Y
$\frac{\partial y_{ij}}{\partial X}$	a matrix of the same order as X
E_{rs}	an elementary matrix of the same order as X
E_{ij}	an elementary matrix of the same order as Y

Preliminaries

1.1 INTRODUCTION

In this chapter we introduce some notation and discuss some results which will be found very useful for the development of the theory of both Kronecker products and matrix differentiation. Our aim will be to make the notation as simple as possible although inevitably it will be complicated. Some simplification may be obtained at the expense of generality. For example, we may show that a result holds for a square matrix of order $n \times n$ and state that it holds in the more general case when A is of order $m \times n$. We will leave it to the interested reader to modify the proof for the more general case.

Further, we will often write

$$\sum_{i,j} a_{ij} \quad \text{or} \quad \sum_i \sum_j a_{ij} \quad \text{or just} \quad \sum \sum a_{ij} \quad \text{instead of} \quad \sum_{i=1}^m \sum_{j=1}^n a_{ij},$$

when the summation limits are obvious from the context.

Many other simplifications will be used as the opportunities arise. Unless of particular importance, we shall not state the order of the matrices considered. It will be assumed that, for example, when taking the product AB or ABC the matrices are conformable.

1.2 UNIT VECTORS AND ELEMENTARY MATRICES

The unit vectors of order n are defined as

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad (1.1)$$

The one vector of order n is defined as

$$\mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (1.2)$$

From (1.1) and (1.2), obtain the relation

$$\mathbf{e} = \sum \mathbf{e}_i \quad (1.3)$$

The elementary matrix E_{ij} is defined as the matrix (of order $m \times n$) which has a unity in the (i, j) th position and all other elements are zero.

For example,

$$E_{23} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (1.4)$$

The relation between \mathbf{e}_i , \mathbf{e}_j and E_{ij} is as follows

$$E_{ij} = \mathbf{e}_i \mathbf{e}_j' \quad (1.5)$$

where \mathbf{e}_j' denotes the transposed vector (that is, the row vector) of \mathbf{e}_j .

Example 1.1

Using the unit vectors of order 3

(i) form E_{11} , E_{21} , and E_{23}

(ii) write the unit matrix of order 3×3 as a sum of the elementary matrices.

Solution

$$(i) \quad E_{11} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} [1 \ 0 \ 0] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$E_{21} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} [1 \ 0 \ 0] = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$E_{23} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} [0 \ 0 \ 1] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(ii) \quad I = E_{11} + E_{22} + E_{33} = \sum_{i=1}^3 \mathbf{e}_i \mathbf{e}_i'.$$

The Kronecker delta δ_{ij} is defined as

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

it can be expressed as

$$\delta_{ij} = \mathbf{e}_i' \mathbf{e}_j = \mathbf{e}_j' \mathbf{e}_i. \quad (1.6)$$

We can now determine some relations between unit vectors and elementary matrices.

$$\begin{aligned} E_{ij} \mathbf{e}_r &= \mathbf{e}_i \mathbf{e}_j' \mathbf{e}_r \quad (\text{by 1.5}) \\ &= \delta_{jr} \mathbf{e}_i \end{aligned} \quad (1.7)$$

and

$$\begin{aligned} \mathbf{e}_r' E_{ij} &= \mathbf{e}_r' \mathbf{e}_i \mathbf{e}_j' \\ &= \delta_{ri} \mathbf{e}_j'. \end{aligned} \quad (1.8)$$

Also

$$E_{ij} E_{rs} = \mathbf{e}_i \mathbf{e}_j' \mathbf{e}_r \mathbf{e}_s' = \delta_{jr} \mathbf{e}_i \mathbf{e}_s' = \delta_{jr} E_{is}. \quad (1.9)$$

In particular if $r = j$, we have

$$E_{ij} E_{js} = \delta_{jj} E_{is} = E_{is}$$

and more generally

$$E_{ij} E_{js} E_{sm} = E_{is} E_{sm} = E_{im}. \quad (1.10)$$

Notice from (1.9) that

$$E_{ij} E_{rs} = 0 \quad \text{if } j \neq r.$$

1.3 DECOMPOSITIONS OF A MATRIX

We consider a matrix A of order $m \times n$ having the following form

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = [a_{ij}] \quad (1.11)$$

We denote the n columns of A by A_1, A_2, \dots, A_n . So that

$$A_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad (j = 1, 2, \dots, n) \quad (1.12)$$

and the m rows of A by $A_{1.}, A_{2.}, \dots, A_{m.}$ so that

$$A_{i.} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \quad (i = 1, 2, \dots, m) \quad (1.13)$$

Both the $A_{.j}$ and the $A_{i.}$ are column vectors. In this notation we can write A as the (partitioned) matrix

$$A = [A_{.1} \ A_{.2} \ \dots \ A_{.n}] \quad (1.14)$$

or as

$$A = [A_{1.} \ A_{2.} \ \dots \ A_{m.}]' \quad (1.15)$$

(where the prime means 'the transpose of').

For example, let

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

so that

$$A_{1.} = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \quad \text{and} \quad A_{2.} = \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix}$$

then

$$[A_{1.} \ A_{2.}]' = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}' = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A.$$

The elements, the columns and the rows of A can be expressed in terms of the unit vectors as follows:

$$\text{The } j\text{th column } A_{.j} = A e_j \quad (1.16)$$

$$\text{The } i\text{th row } A_{i.}' = e_i' A. \quad (1.17)$$

So that

$$A_{i.} = (e_i' A)' = A' e_i. \quad (1.18)$$

The (i, j) th element of A can now be written as

$$a_{ij} = e_i' A e_j = e_i' A' e_i. \quad (1.19)$$

We can express A as the sum

$$A = \sum_i \sum_j a_{ij} E_{ij} \quad (1.20)$$

(where the E_{ij} are of course of the same order as A) so that

$$A = \sum_i \sum_j a_{ij} e_i e_j'. \quad (1.21)$$

From (1.16) and (1.21)

$$\begin{aligned} A_{.j} &= A e_j = \left(\sum_i \sum_l a_{il} e_l e_j' \right) e_j \\ &= \sum_i \sum_l a_{il} e_l (e_j' e_j) \\ &= \sum_i a_{ij} e_i. \end{aligned} \quad (1.22)$$

Similarly

$$A_{li} = \sum_j a_{lj} e_j \quad (1.23)$$

so that

$$A_{li}' = \sum_j a_{lj} e_j'. \quad (1.24)$$

It follows from (1.21), (1.22), and (1.24) that

$$A = \sum_j A_{.j} e_j' \quad (1.25)$$

and

$$A = \sum_l e_l A_{li}'. \quad (1.26)$$

Example 1.2

Write the matrix

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

as a sum of: (i) column vectors of A ; (ii) row vectors of A .

Solutions

(i) Using (1.25)

$$\begin{aligned} A &= A_{.1} e_1' + A_{.2} e_2' \\ &= \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \end{aligned}$$

Using (1.26)

$$\begin{aligned} A &= e_1 A_{1.}' + e_2 A_{2.}' \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} a_{21} & a_{22} \end{bmatrix}. \end{aligned}$$

There exist interesting relations involving the elementary matrices operating on the matrix A .

For example

$$\begin{aligned} E_{ij} A &= e_i e_j' A \quad (\text{by 1.5}) \\ &= e_i A_{j.}' \quad (\text{by 1.17}) \end{aligned} \quad (1.27)$$

16

Preliminaries

[Ch. I

Similarly

$$AE_{ij} = Ae_i e_j' = A_{.i} e_j' \quad (\text{by 1.16}) \quad (1.28)$$

so that

$$AE_{jj} = A_{.j} e_j' \quad (1.29)$$

$$AE_{ij} B = Ae_i e_j' B = A_{.i} B_{j.}' \quad (\text{by 1.28 and 1.27}) \quad (1.30)$$

$$\begin{aligned} E_{ij} AE_{rs} &= e_i e_j' A e_r e_s' & (\text{by 1.5}) \\ &= e_i a_{jr} e_s' & (\text{by 1.19}) \\ &= a_{jr} e_i e_s' = a_{jr} E_{is} \end{aligned} \quad (1.31)$$

In particular

$$E_{jj} AE_{rr} = a_{jr} E_{jr} \quad (1.32)$$

Example 1.3

Use elementary matrices and/or unit vectors to find an expression for

- (i) The product AB of the matrices $A = [a_{ij}]$ and $B = [b_{ij}]$.
- (ii) The k th column of the product AB
- (iii) The k th column of the product XYZ of the matrices $X = [x_{ij}]$, $Y = [y_{ij}]$ and $Z = [z_{ij}]$

Solutions

- (i) By (1.25) and (1.29)

$$A = \sum A_{.j} e_j' = \sum AE_{jj},$$

hence

$$\begin{aligned} AB &= \sum (AE_{jj})B = \sum (Ae_j)(e_j'B) \\ &= \sum A_{.j} B_{j.}' \quad (\text{by (1.16) and (1.17)}) \end{aligned}$$

- (ii) (a)

$$(AB)_{.k} = (AB)e_k = A(Be_k) = AB_{.k} \quad \text{by (1.16)}$$

- (b) From (i) above we can write

$$\begin{aligned} (AB)_{.k} &= \sum_j (Ae_j e_j' B) e_k = \sum_j (Ae_j)(e_j' B e_k) \\ &= \sum_j A_{.j} b_{jk} \quad \text{by (1.16) and (1.19)} . \end{aligned}$$

$$\begin{aligned} \text{(iii)} \quad (XYZ)_{.k} &= \sum_j z_{jk} (XY)_{.j} \quad \text{by (ii)(b) above} \\ &= \sum_j (z_{jk} X) Y_{.j} \quad \text{by (ii)(a) above.} \end{aligned}$$

1.4 THE TRACE FUNCTION

The trace (or the spur) of a square matrix A of order $(n \times n)$ is the sum of the diagonal terms

$$\sum_{i=1}^n a_{ii} .$$

We write

$$\text{tr } A = \sum a_{ii} . \quad (1.33)$$

From (1.19) we have

$$a_{ii} = \mathbf{e}_i' A \mathbf{e}_i ,$$

so that

$$\text{tr } A = \sum \mathbf{e}_i' A \mathbf{e}_i . \quad (1.34)$$

From (1.16) and (1.34) we find

$$\text{tr } A = \sum \mathbf{e}_i' A_{.i} \quad (1.35)$$

and from (1.17) and (1.34)

$$\text{tr } A = \sum A_i . \mathbf{e}_i' . \quad (1.36)$$

We can obtain similar expression for the trace of a product AB of matrices.

For example

$$\text{tr } AB = \sum_i \mathbf{e}_i' A B \mathbf{e}_i . \quad (1.37)$$

$$= \sum_j \sum_i (\mathbf{e}_i' A \mathbf{e}_j) (\mathbf{e}_j' B \mathbf{e}_i) \quad (\text{See Ex. 1.3})$$

$$= \sum_j \sum_i a_{ij} b_{ji} . \quad (1.38)$$

Similarly

$$\text{tr } BA = \sum_j \mathbf{e}_j' B A \mathbf{e}_j$$

$$= \sum_i \sum_j (\mathbf{e}_j' B \mathbf{e}_i) (\mathbf{e}_i' A \mathbf{e}_j)$$

$$= \sum_i \sum_j b_{ji} a_{ij} . \quad (1.39)$$

From (1.38) and (1.39) we find that

$$\text{tr } AB = \text{tr } BA . \quad (1.40)$$

From (1.16), (1.17) and (1.37) we have

$$\text{tr } AB = \sum A_i . B_{.i} . \quad (1.41)$$

Also from (1.40) and (1.41)

$$\text{tr } AB = \sum B_i . A_{.i} . \quad (1.42)$$

Similarly

$$\text{tr } AB' = \sum A_i . B_i' . \quad (1.43)$$

and since $\text{tr } AB' = \text{tr } A'B$

$$\text{tr } AB' = \sum A_i . B_i' . \quad (1.44)$$

Two important properties of the trace are

$$\text{tr}(A+B) = \text{tr} A + \text{tr} B \quad (1.45)$$

and

$$\text{tr}(\alpha A) = \alpha \text{tr} A \quad (1.46)$$

where α is a scalar.

These properties show that trace is a linear function.

For real matrices A and B the various properties of $\text{tr}(AB')$ indicated above show that it is an inner product and is sometimes written as

$$\text{tr}(AB') = (A, B)$$

1.5 THE VEC OPERATOR

We shall make use of a vector valued function denoted by $\text{vec } A$ of a matrix A defined by Neudecker [22].

If A is of order $m \times n$

$$\text{vec } A = \begin{bmatrix} A_{.1} \\ A_{.2} \\ \vdots \\ A_{.n} \end{bmatrix} \quad (1.47)$$

From the definition it is clear that $\text{vec } A$ is a vector of order mn .

For example if

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

then

$$\text{vec } A = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix}.$$

Example 1.4

Show that we can write $\text{tr } AB$ as $(\text{vec } A')' \text{vec } B$

Solution

By (1.37)

$$\begin{aligned} \text{tr } AB &= \sum_i e_i' A B e_i \\ &= \sum_i A_i' B_{.i} \quad \text{by (1.16) and (1.17)} \\ &= \sum_i (A')_{.i}' B_{.i} \end{aligned}$$

(since the i th row of A is the i th column of A')

Hence (assuming A and B of order $n \times n$)

$$\begin{aligned} \text{tr } AB &= [(A')_{.1}(A')_{.2}' \dots (A')_{.n}'] \begin{bmatrix} B_{.1} \\ B_{.2} \\ \vdots \\ B_{.n} \end{bmatrix} \\ &= (\text{vec } A')' \text{vec } B \end{aligned}$$

Before discussing a useful application of the above we must first agree on notation for the transpose of an elementary matrix, we do this with the aid of an example.

$$\text{Let } X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix},$$

then an elementary matrix associated with will X will also be of order (2×3) .

For example, one such matrix is

$$E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The transpose of E_{12} is the matrix

$$E'_{12} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Although at first sight this notation for the transpose is sensible and is used frequently in this book, there are associated snags. The difficulty arises when the suffix notation is not only indicative of the matrix involved but also determines specific elements as in equations (1.31) and (1.32). On such occasions it will be necessary to use a more accurate notation indicating the matrix order and the element involved. Then instead of E_{12} we will write $E_{12}(2 \times 3)$ and instead of E'_{12} we write $E_{21}(3 \times 2)$.

More generally if X is a matrix of order $(m \times n)$ then the transpose of

$$E_{rs}(m \times n)$$

will be written as

$$E'_{rs}$$

unless an accurate description is necessary, in which case the transpose will be written as

$$E_{sr}(n \times m).$$

Now for the application of the result of Example 1.4 which will be used later on in the book.

From the above

$$\begin{aligned}\text{tr } E'_{rs}A &= (\text{vec } E_{rs})' (\text{vec } A) \\ &= a_{rs}\end{aligned}$$

where a_{rs} is the (r, s) th element of the matrix A .

We can of course prove this important result by a more direct method.

$$\begin{aligned}\text{tr } E'_{rs}A &= \sum_k \mathbf{e}'_k E'_{rs} A \mathbf{e}_k \\ &= \sum_{i, l, k} a_{ij} \mathbf{e}'_k \mathbf{e}_s \mathbf{e}'_r \mathbf{e}_i \mathbf{e}'_j \mathbf{e}_k \quad \left(\text{since } A = \sum_{i, j} a_{ij} E_{ij} \right) \\ &= \sum_{i, l, k} a_{ij} \delta_{ks} \delta_{ri} \delta_{jk} = a_{rs}\end{aligned}$$

Problems for Chapter 1

- (1) The matrix A is of order $(4 \times n)$ and the matrix B is of order $(n \times 3)$. Write the product AB in terms of the rows of A , that is, A_1, A_2, \dots and the columns of B , that is, B_1, B_2, \dots .
- (2) Describe in words the matrices
(a) AE_{lk} and (b) $E_{lk}A$.

Write these matrices in terms of an appropriate product of a row or a column of A and a unit vector.

- (3) Show that

$$(a) \text{tr } ABC = \sum_l A_l \cdot BC_l$$

$$(b) \text{tr } ABC = \text{tr } BCA = \text{tr } CAB$$

- (4) Show that $\text{tr } AE_{ij} = a_{ji}$
- (5) $B = [b_{ij}]$ is a matrix of order $(n \times n)$
 $\text{diag } \{B\} = \text{diag } \{b_{11}, b_{22}, \dots, b_{nn}\} = \sum b_{ii} E_{ii}$.
 Show that if

$$a_{ij} = \text{tr } BE_{ij} \delta_{ij}$$

then $A = [a_{ij}] = \text{diag } \{B\}$.

The Kronecker Product

2.1 INTRODUCTION

Kronecker product, also known as a **direct product** or a **tensor product** is a concept having its origin in group theory and has important applications in particle physics. But the technique has been successfully applied in various fields of matrix theory, for example in the solution of matrix equations which arise when using Lyapunov's approach to the stability theory. The development of the technique in this chapter will be as a topic within the scope of matrix algebra.

2.2 DEFINITION OF THE KRONECKER PRODUCT

Consider a matrix $A = [a_{ij}]$ of order $(m \times n)$ and a matrix $B = [b_{ij}]$ of order $(r \times s)$. The Kronecker product of the two matrices, denoted by $A \otimes B$ is defined as the partitioned matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix} \quad (2.1)$$

$A \otimes B$ is seen to be a matrix of order $(mr \times ns)$. It has mn blocks, the (i,j) th block is the matrix $a_{ij}B$ of order $(r \times s)$.

For example, let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},$$

then

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}.$$

The Kronecker Product

[Ch. 2

22

Notice that the Kronecker product is defined irrespective of the order of the matrices involved. From this point of view it is a more general concept than matrix multiplication. As we develop the theory we will note other results which are more general than the corresponding ones for matrix multiplication.

The Kronecker product arises naturally in the following way. Consider two linear transformations

$$x = Az \quad \text{and} \quad y = Bw$$

which, in the simplest case take the form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \quad (2.2)$$

We can consider the two transformations simultaneously by defining the following vectors

$$\mu = x \otimes y = \begin{bmatrix} x_1 y_1 \\ x_1 y_2 \\ x_2 y_1 \\ x_2 y_2 \end{bmatrix} \quad \text{and} \quad v = z \otimes w = \begin{bmatrix} z_1 w_1 \\ z_1 w_2 \\ z_2 w_1 \\ z_2 w_2 \end{bmatrix}. \quad (2.3)$$

To find the transformation between μ and v , we determine the relations between the components of the two vectors.

For example,

$$\begin{aligned} x_1 y_1 &= (a_{11} z_1 + a_{12} z_2) (b_{11} w_1 + b_{12} w_2) \\ &= a_{11} b_{11} (z_1 w_1) + a_{11} b_{12} (z_1 w_2) + a_{12} b_{11} (z_2 w_1) + a_{12} b_{12} (z_2 w_2). \end{aligned}$$

Similar expressions for the other components lead to the transformation

$$\mu = \begin{bmatrix} a_{11} b_{11} & a_{11} b_{12} & a_{12} b_{11} & a_{12} b_{12} \\ a_{11} b_{21} & a_{11} b_{22} & a_{12} b_{21} & a_{12} b_{22} \\ a_{21} b_{11} & a_{21} b_{12} & a_{22} b_{11} & a_{22} b_{12} \\ a_{21} b_{21} & a_{21} b_{22} & a_{22} b_{21} & a_{22} b_{22} \end{bmatrix} v$$

or

$$\mu = (A \otimes B)v,$$

that is

$$Az \otimes Bw = (A \otimes B)(z \otimes w). \quad (2.4)$$

Example 2.1

Let E_{ij} be an elementary matrix of order (2×2) defined in section 1.2 (see 1.4). Find the matrix

$$U = \sum_{i=1}^2 \sum_{j=1}^2 E_{i,j} \otimes E_{j,i}.$$

Solution

$$\begin{aligned}
 U &= E_{11} \otimes E_{11} + E_{1,2} \otimes E_{2,1} + E_{21} \otimes E_{12} + E_{2,2} \otimes E_{2,2} \\
 &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\
 &\quad + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}
 \end{aligned}$$

so that

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Note. U is seen to be a square matrix having columns which are unit vectors $e_i (i = 1, 2, \dots)$. It can be obtained from a unit matrix by a permutation of rows or columns. It is known as a permutation matrix (see also section 2.5).

2.3 SOME PROPERTIES AND RULES FOR KRONECKER PRODUCTS

We expect the Kronecker product to have the usual properties of a product.

I If α is a scalar, then

$$A \otimes (\alpha B) = \alpha(A \otimes B) . \quad (2.5)$$

Proof

The (i, j) th block of $A \otimes (\alpha B)$ is

$$\begin{aligned}
 &[a_{ij}(\alpha B)] \\
 &= \alpha[a_{ij}B] \\
 &= \alpha[(i, j) \text{th block of } A \otimes B]
 \end{aligned}$$

The result follows.

II The product is distributive with respect to addition, that is

$$(a) \quad (A + B) \otimes C = A \otimes C + B \otimes C \quad (2.6)$$

$$(b) \quad A \otimes (B + C) = A \otimes B + A \otimes C \quad (2.7)$$

Proof

We will only consider (a). The (i, j) th block of $(A + B) \otimes C$ is

$$(a_{ij} + b_{ij})C .$$

The (i, j) th block of $A \otimes C + B \otimes C$ is

$$a_{ij}C + b_{ij}C = (a_{ij} + b_{ij})C .$$

24

The Kronecker Product

[Ch. 2

Since the two blocks are equal for every (i, j) , the result follows.

III The product is associative

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C. \quad (2.8)$$

IV There exists

$$\text{a zero element } 0_{mn} = 0_m \otimes 0_n \quad (2.9)$$

$$\text{a unit element } I_{mn} = I_m \otimes I_n.$$

The unit matrices are all square, for example I_m in the unit matrix of order $(m \times m)$.

Other important properties of the Kronecker product follow.

$$V \quad (A \otimes B)' = A' \otimes B' \quad (2.10)$$

Proof

The (i, j) th block of $(A \otimes B)'$ is

$$a_{ji} B'.$$

VI (The 'Mixed Product Rule').

$$(A \otimes B)(C \otimes D) = AC \otimes BD \quad (2.11)$$

provided the dimensions of the matrices are such that the various expressions exist.

Proof

The (i, j) th block of the left hand side is obtained by taking the product of the i th row block of $(A \otimes B)$ and the j th column block of $(C \otimes D)$, this is of the following form

$$\begin{aligned} & [a_{i1}B \ a_{i2}B \ \dots \ a_{in}B] \begin{bmatrix} c_{1j}D \\ c_{2j}D \\ \vdots \\ c_{nj}D \end{bmatrix} \\ &= \sum_r a_{ir} c_{rj} BD. \end{aligned}$$

The (i, j) th block of the right hand side is (by definition of the Kronecker product)

$$g_{ij} BD$$

where g_{ij} is the (i, j) th element of the matrix AC . But by the rule of matrix multiplications

$$g_{ij} = \sum_r a_{ir} c_{rj}.$$

Since the (i, j) th blocks are equal, the result follows.

VII Given $A(m \times m)$ and $B(n \times n)$ and subject to the existence of the various inverses,

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \quad (2.12)$$

Proof

Use (2.11)

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = AA^{-1} \otimes BB^{-1} = I_m \otimes I_n = I_{mn}$$

The result follows.

VIII (See (1.47))

$$\text{vec}(AYB) = (B' \otimes A) \text{vec } Y \quad (2.13)$$

Proof

We prove (2.13) for A , Y and B each of order $n \times n$. The result is true for $A(m \times n)$, $Y(n \times r)$, $B(r \times s)$. We use the solutions to Example 1.3(iii).

$$\begin{aligned} (AYB)_{\cdot k} &= \sum_j (b_{jk} A) Y_{\cdot j} \\ &= [b_{1k} A \quad b_{2k} A \quad \dots \quad b_{nk} A] \begin{bmatrix} Y_{\cdot 1} \\ Y_{\cdot 2} \\ \vdots \\ Y_{\cdot n} \end{bmatrix} \\ &= [B_{\cdot k}' \otimes A] \text{vec } Y \\ &= [(B')_k]' \otimes A \text{vec } Y \end{aligned}$$

since the transpose of the k th column of B is the k th row of B' ; the results follows.

Example 2.2

Write the equation

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

in a matrix-vector form.

Solution

The equation can be written as $AXI = C$. Use (2.12), to find

$$\text{vec}(AXI) = (I \otimes A) \text{vec } X = \text{vec } C,$$

so that

$$\begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & a_{11} & a_{12} \\ 0 & 0 & a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{12} \\ c_{22} \end{bmatrix}$$

Example 2.3

A and B are both of order $(n \times n)$, show that

- (i) $\text{vec } AB = (I \otimes A) \text{vec } B$
- (ii) $\text{vec } AB = (B' \otimes A) \text{vec } I$
- (iii) $\text{vec } AB = \Sigma(B')_{\cdot k} \otimes A_{\cdot k}$

Solution

- (i) (As in Example 2.2)
In (2.13) let $Y = B$ and $B = I$.
- (ii) In (2.13) let $Y = I$.
- (iii) In $\text{vec } AB = (B' \otimes A) \text{vec } I$

substitute (1.25), to obtain

$$\begin{aligned} \text{vec } AB &= \left[\sum_i (B')_{\cdot i} e'_i \otimes \sum_j A_{\cdot j} e_j \right] \text{vec } I \\ &= \left[\sum_i \sum_j ((B')_{\cdot i} \otimes A_{\cdot j}) (e'_i \otimes e_j) \right] \text{vec } I \end{aligned} \quad (\text{by 2.11})$$

The product $e'_i \otimes e_j$ is a one row matrix having a unit element in the $[(i-1)n + j]$ th column and zeros elsewhere. Hence the product

$$[(B')_{\cdot i} \otimes A_{\cdot j}] [e'_i \otimes e_j]$$

is a matrix having

$$(B')_{\cdot i} \otimes A_{\cdot j}$$

as its $[(i-1)n + j]$ th column and zeros elsewhere. Since $\text{vec } I$ is a one column matrix having a unity in the 1st, $(n+2)$ nd, $(2n+3)$ rd \dots n^2 rd position and zeros elsewhere, the product of

$$[(B')_{\cdot i} \otimes A_{\cdot j}] [e'_i \otimes e_j] \text{ and } \text{vec } I$$

is a one column matrix whose elements are all zeros unless i and j satisfy

$$(i-1)n + j = 1, \text{ or } n+2, \text{ or } 2n+3, \dots, \text{ or } n^2$$

that is

$$i = j = 1 \quad \text{or} \quad i = j = 2 \quad \text{or} \quad i = j = 3 \quad \text{or} \quad \dots, \quad i = j = n$$

in which case the one column matrix is

$$(B')_{\cdot i} \otimes A_{\cdot i} \quad (i = 1, 2, \dots, n).$$

The result now follows.

IX If $\{\lambda_i\}$ and $\{x_i\}$ are the eigenvalues and the corresponding eigenvectors for A and $\{\mu_j\}$ and $\{y_j\}$ are the eigenvalues and the corresponding eigenvectors for B , then

$$A \otimes B$$

has eigenvalues $\{\lambda_i \mu_j\}$ with corresponding eigenvectors $\{x_i \otimes y_j\}$.

Proof

By (2.11)

$$\begin{aligned} (A \otimes B)(x_i \otimes y_j) &= (Ax_i) \otimes (By_j) \\ &= (\lambda_i x_i) \otimes (\mu_j y_j) \\ &= \lambda_i \mu_j (x_i \otimes y_j) \end{aligned} \quad (\text{by 2.5})$$

The result follows.

X Given the two matrices A and B of order $n \times n$ and $m \times m$ respectively

$$|A \otimes B| = |A|^m |B|^n$$

where $|A|$ means the determinant of A .

Proof

Assume that $\lambda_1, \lambda_2, \dots, \lambda_n$ and $\mu_1, \mu_2, \dots, \mu_m$ are the eigenvalues of A and B respectively. The proof relies on the fact (see [18] p. 145) that the determinant of a matrix is equal to the product of its eigenvalues.

Hence (from Property IX above)

$$\begin{aligned} |A \otimes B| &= \prod_{i,j} \lambda_i \mu_j \\ &= \left(\lambda_1^m \prod_{j=1}^m \mu_j \right) \left(\lambda_2^m \prod_{j=1}^m \mu_j \right) \dots \left(\lambda_n^m \prod_{j=1}^m \mu_j \right) \\ &= (\lambda_1 \lambda_2 \dots \lambda_n)^m (\mu_1 \mu_2 \dots \mu_m)^n \\ &= |A|^m |B|^n. \end{aligned}$$

Another important property of Kronecker products follows.

$$A \otimes B = U_1(B \otimes A)U_2 \quad (2.14)$$

where U_1 and U_2 are permutation matrices (see Example 2.1).

Proof

Let $AYB' = X$, then by (2.13)

$$(B \otimes A) \text{vec } Y = \text{vec } X. \quad (1)$$

On taking transpose, we obtain

$$BY'A' = X'$$

So that by (2.13)

$$(A \otimes B) \text{vec } Y' = \text{vec } X'. \quad (2)$$

From example 1.5, we know that there exist permutation matrices U_1 and U_2 such that

$$\text{vec } X' = U_1 \text{vec } X \quad \text{and} \quad \text{vec } Y = U_2 \text{vec } Y'.$$

Substituting for $\text{vec } Y$ in (1) and multiplying both sides by U_1 , we obtain

$$U_1(B \otimes A)U_2 \text{vec } Y' = U_1 \text{vec } X. \quad (3)$$

Substituting for $\text{vec } X'$ in (2), we obtain

$$(A \otimes B) \text{vec } Y' = U_1 \text{vec } X. \quad (4)$$

The result follows from (3) and (4).

We will obtain an explicit formula for the permutation matrix U in section 2.5. Notice that U_1 and U_2 are independent of A and B except for the orders of the matrices.

XII If f is an analytic function, A is a matrix of order $(n \times n)$ and $f(A)$ exists, then

$$f(I_m \otimes A) = I_m \otimes f(A) \quad (2.15)$$

and

$$f(A \otimes I_m) = f(A) \otimes I_m \quad (2.16)$$

Proof

Since f is an analytic function it can be expressed as a power series such as

$$f(z) = a_0 + a_1 z + a_2 z^2 + \dots$$

so that

$$f(A) = a_0 I_n + a_1 A + a_2 A^2 + \dots = \sum_{k=0}^{\infty} a_k A^k,$$

where $A^0 = I_n$.

By Cayley Hamilton's theorem (see [18]) the right hand side of the equation for $f(A)$ is the sum of at most $(n + 1)$ matrices.

We now have

$$\begin{aligned}
 f(I_m \otimes A) &= \sum_{k=0}^{\infty} a_k (I_m \otimes A^k) \\
 &= \sum_{k=0}^{\infty} (I_m \otimes a_k A^k) && \text{by (2.11)} \\
 &= \sum_{k=0}^{\infty} (I_m \otimes a_k A^k) && \text{by (2.5)} \\
 &= I_m \otimes \sum_{k=0}^{\infty} a_k A^k && \text{by (2.7)} \\
 &= I_m \otimes f(A) .
 \end{aligned}$$

This proves (2.15); (2.16) is proved similarly.

We can write

$$\begin{aligned}
 f(A \otimes I_m) &= \sum_{k=0}^{\infty} a_k (A \otimes I_m)^k \\
 &= \sum_{k=0}^{\infty} a_k (A^k \otimes I_m) && \text{by (2.11)} \\
 &= \sum_{k=0}^{\infty} (a_k A^k \otimes I_m) \\
 &= \sum_{k=0}^{\infty} a_k A^k \otimes I_m && \text{by (2.6)} \\
 &= f(A) \otimes I_m
 \end{aligned}$$

This proves (2.16).

An important application of the above property is for

$$f(z) = e^z .$$

(2.15) leads to the result

$$e^{I_m \otimes A} = I_m \otimes e^A \tag{2.17}$$

and (2.16) leads to

$$e^{A \otimes I_m} = e^A \otimes I_m \tag{2.18}$$

Example 2.4

Use a direct method to verify (2.17) and (2.18).

Solution

$$e^{I_m \otimes A} = I_m \otimes I_n + I_m \otimes A + \frac{1}{2!} (I_m \otimes A)^2 + \dots$$

The right hand side is a block diagonal matrix, each of the m blocks is the sum

$$I_m + A + \frac{1}{2!} A^2 + \dots = e^A.$$

The result (2.17) follows.

$$\begin{aligned} e^A \otimes I_m &= (I_n \otimes I_m) + (A \otimes I_m) + \frac{1}{2!} (I_m \otimes A)^2 + \dots \\ &= (I_n \otimes I_m) + (A \otimes I_m) + \frac{1}{2!} (A^2 \otimes I_m) + \dots \\ &= (I_n + A + \frac{1}{2!} A^2 + \dots) \otimes I_m \\ &= e^A \otimes I_m \end{aligned}$$

$$\text{XIII } \text{tr}(A \otimes B) = \text{tr } A \text{ tr } B$$

Proof

Assume that A is of order $(n \times n)$

$$\begin{aligned} \text{tr}(A \otimes B) &= \text{tr}(a_{11}B) + \text{tr}(a_{22}B) + \dots + \text{tr}(a_{nn}B) \\ &= a_{11} \text{tr } B + a_{22} \text{tr } B + \dots + a_{nn} \text{tr } B \\ &= (a_{11} + a_{22} + \dots + a_{nn}) \text{tr } B \\ &= \text{tr } A \text{ tr } B. \end{aligned}$$

2.4 DEFINITION OF THE KRONECKER SUM

Given a matrix $A(n \times n)$ and a matrix $B(m \times m)$, their Kronecker Sum denoted by $A \oplus B$ is defined as the expression

$$A \oplus B = A \otimes I_m + I_n \otimes B \quad (2.19)$$

We have seen (Property IX) that if $\{\lambda_i\}$ and $\{\mu_j\}$ are the eigenvalues of A and B respectively, then $\{\lambda_i \mu_j\}$ are the eigenvalues of the product $A \otimes B$. We now show the equivalent and fundamental property for $A \oplus B$.

XIV If $\{\lambda_i\}$ and $\{\mu_j\}$ are the eigenvalues of A and B respectively, then $\{\lambda_i + \mu_j\}$ are the eigenvalues of $A \oplus B$.

Proof

Let x and y be the eigenvectors corresponding to the eigenvalues λ and μ of A and B respectively, then

$$\begin{aligned} (A \oplus B)(x \otimes y) &= (A \otimes I)(x \otimes y) + (I \otimes B)(x \otimes y) && \text{by (2.19)} \\ &= (Ax \otimes y) + (x \otimes By) && \text{by (2.11)} \\ &= \lambda(x \otimes y) + \mu(x \otimes y) \\ &= (\lambda + \mu)(x \otimes y) \end{aligned}$$

The result follows.

Example 2.5

Verify the Property XIV for

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

Solution

For the matrix A ;

$$\lambda_1 = 1 \quad \text{and} \quad x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\lambda_2 = 2 \quad \text{and} \quad x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

For the matrix B ;

$$\mu_1 = 1 \quad \text{and} \quad y_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mu_2 = 2 \quad \text{and} \quad y_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We find

$$C = A \oplus B = \begin{bmatrix} 2 & 0 & -1 & 0 \\ 2 & 0 & 0 & -1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

and $|pI - C| = p(p-1)(p-2)(p-3)$, so that the eigenvalues of $A \oplus B$ are

$$p = 0 = \lambda_1 + \mu_2 \quad \text{and} \quad x_1 \otimes y_2 = [0 \ 1 \ 0 \ 0]'$$

$$p = 1 = \lambda_2 + \mu_2 \quad \text{and} \quad x_2 \otimes y_2 = [0 \ 1 \ 0 \ -1]'$$

$$\text{and} \quad p = 2 = \lambda_1 + \mu_1 \quad \text{and} \quad x_1 \otimes y_1 = [1 \ 1 \ 0 \ 0]'$$

$$p = 3 = \lambda_2 + \mu_1 \quad \text{and} \quad x_2 \otimes y_1 = [1 \ 1 \ -1 \ -1]'$$

The Kronecker sum frequently turns up when we are considering equations of the form,

$$AX + XB = C \tag{2.20}$$

where $A(n \times n)$, $B(m \times m)$ and $X(n \times m)$.

Use (2.13) and solution to Example 2.3 to write the above in the form

$$(I_m \otimes A + B' \otimes I_n) \text{vec } X = \text{vec } C \tag{2.21}$$

or

$$(B' \oplus A) \text{vec } X = \text{vec } C.$$

It is interesting to note the generality of the Kronecker sum. For example,

$$\exp(A + B) = \exp A \exp B$$

if and only if A and B commute (see [18] p. 227)
 whereas $\exp(A \oplus B) = \exp(A \otimes I) \exp(I \otimes B)$
 even if A and B do not commute!

Example 2.6

Show that

$$\exp(A \oplus B) = \exp A \otimes \exp B$$

where $A(n \times n)$, $B(m \times m)$.

Solution

By (2.11)

$$(A \otimes I_m)(I_n \otimes B) = A \otimes B$$

and

$$(I_n \otimes B)(A \otimes I_m) = A \otimes B$$

hence $(A \otimes I_m)$ and $(I_n \otimes B)$ commute so that

$$\begin{aligned} \exp(A \oplus B) &= \exp(A \otimes I_m + I_n \otimes B) \\ &= \exp(A \otimes I_m) \exp(I_n \otimes B) \\ &= (\exp A \otimes I_m)(I_n \otimes \exp B) && \text{(by 2.15 and 2.16)} \\ &= \exp A \otimes \exp B && \text{(by 2.11)} \end{aligned}$$

2.5 THE PERMUTATION MATRIX ASSOCIATING $\text{vec } X$ AND $\text{vec } X'$

If $X = [x_{ij}]$ is a matrix of order $(m \times n)$ we can write (see (1.20))

$$X = \sum_i \sum_j x_{ij} E_{ij}$$

where E_{ij} is an elementary matrix of order $(m \times n)$. It follows that

$$X' = \sum \sum x_{ij} E'_{ij}.$$

so that

$$\text{vec } X' = \sum \sum x_{ij} \text{vec } E'_{ij}. \quad (2.22)$$

We can write (2.22) in a form of matrix multiplication as

$$\text{vec } X' = [\text{vec } E'_{11} : \text{vec } E'_{21} : \dots : \text{vec } E'_{m1} : \text{vec } E'_{12} : \dots : \text{vec } E'_{mn}]$$

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{m1} \\ x_{12} \\ \vdots \\ x_{mn} \end{bmatrix}$$

that is

$$\text{vec } X' = [\text{vec } E'_{11}; \text{vec } E'_{21}; \dots \text{vec } E'_{m1}; \text{vec } E'_{12}; \dots \text{vec } E'_{mn}] \text{vec } X.$$

So the permutation matrix associating $\text{vec } X$ and $\text{vec } X'$ is

$$U = [\text{vec } E'_{11}; \text{vec } E'_{21}; \dots \text{vec } E'_{mn}] . \quad (2.23)$$

Example 2.7

Given

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}, \quad \text{determine the matrix } U$$

such that

$$\text{vec } X' = U \text{vec } X.$$

Solution

$$E'_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad E'_{21} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad E'_{12} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E'_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$E'_{13} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad E'_{23} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence by (2.23)

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We now obtain the permutation matrix U in a useful form as a Kronecker product of elementary matrices.

As it is necessary to be precise about the suffixes of the elementary matrices, we will use the notation explained at the end of Chapter 1.

As above, we write

$$X' = \sum_{r=1}^m \sum_{s=1}^n x_{rs} E_{sr} \quad (n \times m).$$

By (1.31) we can write

$$X' = \sum_{r,s} E_{sr} (n \times m) X E_{sr} (n \times m).$$

Hence,

$$\begin{aligned} \text{vec } X' &= \text{vec} \sum_{r,s} E_{sr} (n \times m) X E_{sr} (n \times m) \\ &= \sum_{r,s} [E_{rs} (m \times n) \otimes E_{sr} (n \times m)] \text{vec } X. \end{aligned} \quad \text{by (2.13)}$$

It follows that

$$U = \sum_{r,s} E_{rs} (m \times n) \otimes E_{sr} (n \times m) \quad (2.24)$$

or in our less rigorous notation

$$U = \sum_{r,s} E_{rs} \otimes E'_{rs}. \quad (2.25)$$

Notice that U is a matrix of order $(mn \times mn)$.

At first sight it may appear that the evaluation of the permutation matrices U_1 and U_2 in (2.14) using (2.24) is a major task. In fact this is one of the examples where the practice is much easier than the theory.

We can readily determine the form of a permutation matrix — as in Example 2.7. So the only real problem is to determine the orders of the two matrices.

Since the matrices forming the product (2.14) must be conformable, the orders of the matrices U_1 and U_2 are determined respectively by the number of rows and the number of columns of $(A \otimes B)$.

Example 2.8

Let $A = [a_{ij}]$ be a matrix of order (2×3) , and $B = [b_{ij}]$ be a matrix of order (2×2) .

Determine the permutation matrices U_1 and U_2 such that

$$A \otimes B = U_1 (B \otimes A) U_2$$

Solution

$(A \otimes B)$ is of the order (4×6)

From the above discussion we conclude that U_1 is of order (4×4) and U_2 is of order (6×6) .

$$U_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Another related matrix which will be used (in Chapter 6) is

$$\bar{U} = \sum_{r,s} E_{rs} \otimes E_{rs}. \quad (2.26)$$

When the matrix X is of order $(m \times n)$, \bar{U} is of order $(m^2 \times n^2)$.

Problems of Chapter 2

(1) Given

$$U = \sum_{r,s} E_{rs} (m \times n) \otimes E_{sr} (n \times m).$$

Show that

$$U^{-1} = U' = \sum_{r,s} E_{sr} (n \times m) \otimes E_{rs} (m \times n)$$

(2) $A = [a_{ij}]$, $B = [b_{ij}]$ and $Y = [y_{ij}]$ are matrices all of order (2×2) , use a direct method to evaluate

(a) (i) AYB

(ii) $B' \otimes A$.

(b) Verify (2.13) that

$$\text{vec } AYB = (B' \otimes A) \text{vec } Y.$$

(3) Given

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

(a) Calculate

$$A \otimes B \quad \text{and} \quad B \otimes A.$$

(b) Find matrices U_1 and U_2 such that

$$A \otimes B = U_1 (B \otimes A) U_2.$$

(4) Given

$$A = \begin{bmatrix} 3 & 4 \\ -2 & -3 \end{bmatrix}, \quad \text{calculate}$$

(a) $\exp(A)$

(b) $\exp(A \otimes I)$.

Verify (2.16), that is

$$\exp(A) \otimes I = \exp(A \otimes I).$$

The Kronecker Product

[Ch. 2]

36

(5) Given

$$A = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \text{ calculate}$$

(a) $A^{-1} \otimes B^{-1}$ and

(b) $(A \otimes B)^{-1}$.

Hence verify (2.12), that is

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

(6) Given

$$A = \begin{bmatrix} 3 & -1 \\ 4 & -2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}, \text{ find}$$

(a) The eigenvalues and eigenvectors of A and B .(b) The eigenvalues and eigenvectors of $A \otimes B$.

(c) Verify Property IX of Kronecker Products.

(7) A, B, C and D are matrices such that A is similar to C , and B is similar to D .Show that $A \otimes B$ is similar to $C \otimes D$.

CHAPTER 3

Some Applications of the Kronecker Product

3.1 INTRODUCTION

There are numerous applications of the Kronecker product in various fields including statistics, economics, optimisation and control. It is not our intention to discuss applications in all these fields, just a selected number to give an idea of the problems tackled in some of the literature mentioned in the Bibliography. There is no doubt that the interested reader will find there various other applications hopefully in his own field of interest.

A number of the applications involve the derivative of a matrix — it is a well known concept (for example see [18] p. 229) which we now briefly review.

3.2 THE DERIVATIVE OF A MATRIX

Given the matrix

$$A(t) = [a_{ij}(t)]$$

the derivative of the matrix, with respect to a scalar variable t , denoted by $(d/dt)A(t)$ or just dA/dt or $\dot{A}(t)$ is defined as the matrix

$$\frac{d}{dt}A(t) = \left[\frac{d}{dt} a_{ij}(t) \right]. \quad (3.1)$$

Similarly, the integral of the matrix is defined as

$$\int A(t) dt = \left[\int a_{ij}(t) dt \right]. \quad (3.2)$$

For example, given

$$A = \begin{bmatrix} 2t^2 & 4 \\ \sin t & 2 + t^2 \end{bmatrix}$$

then

$$\frac{d}{dt} A = \begin{bmatrix} 4t & 0 \\ \cos t & 2t \end{bmatrix} \quad \text{and} \quad \int A dt = \begin{bmatrix} \frac{2}{3}t^3 & 4t \\ -\cos t & 2t + t^3/3 \end{bmatrix} + C$$

where C is a constant matrix.

One important property follows immediately. Given conformable matrices $A(t)$ and $B(t)$, then

$$\frac{d}{dt} [AB] = \frac{dA}{dt} B + A \frac{dB}{dt} . \quad (3.3)$$

Example 3.1

Given

$$C = A \otimes B$$

(each matrix is assumed to be a function of t) show that

$$\frac{dC}{dt} = \frac{dA}{dt} \otimes B + A \otimes \frac{dB}{dt} \quad (3.4)$$

Solution

On differentiating the (i,j) th block of $A \otimes B$, we obtain

$$\frac{d}{dt} (a_{ij} B) = \frac{da_{ij}}{dt} B + a_{ij} \frac{dB}{dt}$$

which is the (i,j) th partition of

$$\frac{dA}{dt} \otimes B + A \otimes \frac{dB}{dt} ,$$

the result follows.

3.3 PROBLEM 1

Determine the condition for the equation

$$AX + XB = C$$

to have a unique solution.

Solution

We have already considered this equation and wrote it (2.21) as

$$(B' \oplus A) \text{vec } X = \text{vec } C$$

or

$$GX = c \quad (3.5)$$

where $G = B' \oplus A$ and $c = \text{vec } C$.

Equation (3.5) has a unique solution iff G is nonsingular, that is iff the eigenvalues of G are all nonzero. Since, by Property XIV (see section 2.4), the eigenvalues of G are $\{\lambda_i + \mu_j\}$ (note that the eigenvalues of the matrix B' are the same as the eigenvalues of B). Equation (3.5) has a unique solution iff

$$\lambda_i + \mu_j \neq 0 \quad (\text{all } i \text{ and } j).$$

We have thus proved that $AX + BX = C$ has a unique solution iff A and $(-B)$ have no eigenvalue in common.

If on the other hand, A and $(-B)$ have common eigenvalues then the existence of solutions depends on the rank of the augmented matrix

$$[G:c].$$

If the rank of $[G:c]$ is equal to the rank of G , then solutions do exist, otherwise the set of equations

$$AX + XB = C$$

is not consistent.

Example 3.2

Obtain the solution to

$$AX + XB = C$$

where

$$(i) \quad A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 3 \\ -2 & 2 \end{bmatrix}.$$

$$(ii) \quad A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -3 & 4 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 0 & 5 \\ 2 & -9 \end{bmatrix}.$$

Solution

Writing the equation in the form of (3.5) we obtain,

$$(i) \quad \begin{bmatrix} -2 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 4 & 0 & 1 & -1 \\ 0 & 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 2 \end{bmatrix}$$

where for convenience we have denoted

$$X = \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix}.$$

On solving we obtain the unique solution

$$X = \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}.$$

(ii) In case (ii) A and $(-B)$ have one eigenvalue ($\lambda = 1$) in common. Equation (3.5) becomes

$$\begin{bmatrix} -2 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 4 & 0 & 0 & -1 \\ 0 & 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 5 \\ -9 \end{bmatrix}$$

and $\text{rank } G = \text{rank } [G; c]$.

G is seen to be singular, but

$$\text{rank } G = \text{rank } [G; c] = 3$$

hence at least one solution exists. In fact two linearly independent solutions are

$$X_1 = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$$

any other solution is a linear combination of X_1 and X_2 .

3.4 PROBLEM 2

Determine the condition for the equation

$$AX - XA = \mu X \tag{3.6}$$

to have a nontrivial solution.

Solution

We can write (3.6) as

$$Hx = \mu x \tag{3.7}$$

where $H = I \otimes A - A' \otimes I$ and

$$x = \text{vec } X.$$

(3.7) has a nontrivial solution for x iff

$$|\mu I - H| = 0$$

that is iff μ is an eigenvalue of H . But by a simple generalisation of Property XIV,

section 2.4, the eigenvalues of H are $\{(\lambda_i - \lambda_j)\}$ where $\{\lambda_i\}$ are the eigenvalues of A . Hence (3.6) has a nontrivial solution iff

$$\mu = \lambda_i - \lambda_j .$$

Example 3.3

Determine the solutions to (3.6) when

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \mu = -2 .$$

Solution

$\mu = -2$ is an eigenvalue of H , hence we expect a nontrivial solution. Equation (3.7) becomes

$$\begin{bmatrix} 0 & 0 & -2 & 0 \\ 2 & 2 & 0 & -2 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = -2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

On solving, we obtain

$$X = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

3.5 PROBLEM 3

Use the fact (see [18] p. 230) that the solution to

$$\dot{x} = Ax, \quad x(0) = c \tag{3.8}$$

$$x = \exp(At) c \tag{3.9}$$

to solve the equation

$$\dot{X} = AX + XB, \quad X(0) = C \tag{3.10}$$

where $A(n \times n)$, $B(m \times m)$ and $X(n \times m)$.

Solution

Using the vec operator on (3.10) we obtain

$$\dot{x} = Gx, \quad x(0) = c \tag{3.11}$$

where

$$x = \text{vec } X, \quad c = \text{vec } C$$

and

$$G = I_m \otimes A + B' \otimes I_n .$$

By (3.9) the solution to (3.11) is

$$\begin{aligned}\text{vec } X &= \exp \{ (I_m \otimes A)t + (B' \otimes I_n)t \} \text{vec } C \\ &= [\exp (I_m \otimes A)t] [\exp (B' \otimes I_n)t] \text{vec } C \quad (\text{see Example 2.6}) \\ &= [I_m \otimes \exp (At)] [\exp (B't) \otimes I_n] \text{vec } C \quad \text{by (2.17) and (2.18)}.\end{aligned}$$

We now make use of the result

$$\text{vec } AB = (B' \otimes I) \text{vec } A$$

(in (2.13) put $A = I$ and $Y = A$) in conjunction with the fact that

$$[\exp (B't)] = \exp (Bt),$$

to obtain

$$(\exp (B't) \otimes I_n) \text{vec } C = \text{vec } [C \exp (Bt)].$$

Using the result of Example 2.3(1), we finally obtain

$$\text{vec } X = \text{vec } [\exp (At) C \exp (Bt)] \quad (3.12)$$

So that $X = \exp (At) C \exp (Bt)$.

Example 3.4

Obtain the solution to (3.10) when

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

Solution

(See [18] p. 227)

$$\exp (At) = \begin{bmatrix} e^t & e^t - e^{2t} \\ 0 & e^{2t} \end{bmatrix}, \quad \exp (Bt) = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$$

hence

$$X = \begin{bmatrix} -e^{2t} - e^{3t} & 1 - e^t \\ e^{3t} & e^t \end{bmatrix}.$$

3.6 PROBLEM 4

We consider a problem similar to the previous one but in a different context.

An important concept in Control Theory is the **transition matrix**.

Very briefly, associated with the equations

$$\dot{X} = A(t)X \quad \text{or} \quad \dot{x} = A(t)x$$

is the transition matrix $\Phi_1(t, \tau)$ having the following two properties

$$\dot{\Phi}_1(t, \tau) = A(t)\Phi_1(t, \tau) \quad (3.13)$$

and

$$\Phi_1(t, t) = I$$

[For simplicity of notation we shall write Φ for $\Phi(t, \tau)$.] If A is a constant matrix, it is easily shown that

$$\Phi = \exp(At) .$$

Similarly, with the equation

$$\dot{X} = XB \quad \text{so that} \quad \dot{X}' = B'X'$$

we associate the transition matrix Φ_2 such that

$$\Phi_2 = B'\Phi_2 . \quad (3.14)$$

The problem is to find the transition matrix associated with the equation

$$\dot{X} = AX + XB \quad (3.15)$$

given the transition matrices Φ_1 and Φ_2 defined above.

Solution

We can write (3.15) as

$$\dot{x} = Gx$$

where x and G were defined in the previous problem.

We define a matrix ψ as

$$\psi(t, \tau) = \Phi_2(t, \tau) \otimes \Phi_1(t, \tau) . \quad (3.16)$$

We obtain by (3.4)

$$\begin{aligned} \dot{\psi} &= \dot{\Phi}_2 \otimes \Phi_1 + \Phi_2 \otimes \dot{\Phi}_1 \\ &= (B'\Phi_2) \otimes \Phi_1 + \Phi_2 \otimes (A\Phi_1) && \text{by (3.13) and (3.14)} \\ &= (B'\Phi_2) \otimes (I\Phi_1) + (I\Phi_2) \otimes (A\Phi_1) \\ &= [B' \otimes I + I \otimes A] [\Phi_2 \otimes \Phi_1] . && \text{by (2.11)} \end{aligned}$$

Hence

$$\dot{\psi} = G\psi .$$

Also

$$\begin{aligned} \psi(t, t) &= \Phi_2(t, t) \otimes \Phi_1(t, t) \\ &= I \otimes I \\ &= I . \end{aligned} \quad (3.18)$$

The two equations (3.17) and (3.18) prove that ψ is the transition matrix for (3.15)

Example 3.5

Find the transition matrix for the equation

$$\dot{X} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} X + X \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} .$$

Solution

In this case both A and B are constant matrices. From Example 3.4.

$$\Phi_1 = \exp(At) = \begin{bmatrix} e^t & e^t - e^{2t} \\ 0 & e^{2t} \end{bmatrix}$$

$$\Phi_2 = \exp(Bt) = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$$

so that

$$\psi = \Phi_2 \otimes \Phi_1 = \begin{bmatrix} e^{2t} & e^{2t} - e^{3t} & 0 & 0 \\ 0 & e^{3t} & 0 & 0 \\ 0 & 0 & 1 & 1 - e^t \\ 0 & 0 & 0 & e^t \end{bmatrix}.$$

For this equation

$$G = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and it is easily verified that

$$\dot{\psi} = G\psi$$

and

$$\psi(0) = I.$$

3.7 PROBLEM 5

Solve the equation

$$AXB = C$$

where all matrices are of order $n \times n$.

Solution

Using (2.13) we can write (3.19) in the form

$$Hx = c \tag{3.20}$$

where $H = B' \otimes A$, $x = \text{vec } X$ and $c = \text{vec } C$.

The criteria for the existence and the uniqueness of a solution to (3.20) are well known (see for example [18]).

The above method of solving the problem is easily generalised to the linear equation of the form

$$A_1XB_1 + A_2XB_2 + \dots + A_rXB_r = C \tag{3.21}$$

Equation (3.21) can be written as for example (3.20) where this time

$$H = B'_1 \otimes A_1 + B'_2 \otimes A_2 + \dots + B'_r \otimes A_r.$$

Example 3.6

Find the matrix X , given

$$A_1 X B_1 + A_2 X B_2 = C$$

where

$$A_1 = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 4 & -6 \\ 0 & 8 \end{bmatrix}.$$

Solution

For this example it is found that

$$H = B'_1 \otimes A_1 + B'_2 \otimes A_2 = \begin{bmatrix} 2 & 2 & -2 & -3 \\ 1 & -1 & 1 & 2 \\ 0 & 2 & 2 & 5 \\ -4 & -2 & -5 & -4 \end{bmatrix}$$

and $c' = [4 \ 0 \ -6 \ 8]$.

It follows that

$$x = H^{-1}c = \begin{bmatrix} 1 \\ -1 \\ -2 \\ 0 \end{bmatrix}$$

so that

$$X = \begin{bmatrix} 1 & -2 \\ -1 & 0 \end{bmatrix}$$

3.8 PROBLEM 6

This problem is to determine a constant output feedback matrix K so that the closed loop matrix of a system has preassigned eigenvalues.

A multivariable system is defined by the equations

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{3.22}$$

where $A(n \times n)$, $B(n \times m)$ and $C(r \times n)$ are constant matrices. u , x and y are column vectors of order m , n and r respectively.

We are concerned with a system having an output feedback law of the form

$$u = Ky \quad (3.23)$$

where $K(m \times r)$ is the constant control matrix to be determined.

On substituting (3.23) into (3.22), we obtain the equations of the closed loop system

$$\begin{aligned} \dot{x} &= (A + BKC)x \\ y &= Cx \end{aligned} \quad (3.24)$$

The problem can now be restated as follows:

Given the matrices A , B , and C , determine a matrix K such that

$$\begin{aligned} |\lambda I - A - BKC| &= a_0 + a_1\lambda + \dots + a_{n-1}\lambda^{n-1} + \lambda^n \quad (\text{say}) \\ &= 0 \text{ for preassigned values } \lambda = \lambda_1, \lambda_2, \dots, \lambda_n. \end{aligned} \quad (3.25)$$

Solution

Various solutions exist to this problem. We are interested in the application of the Kronecker product and will follow a method suggested in [24].

We consider a matrix $H(n \times n)$ whose eigenvalues are the desired values $\lambda_1, \lambda_2, \dots, \lambda_n$, that is

$$|\lambda I - H| = 0 \quad \text{for } \lambda = \lambda_1, \lambda_2, \dots, \lambda_n \quad (3.26)$$

and

$$|\lambda I - H| = a_0 + a_1\lambda + \dots + a_{n-1}\lambda^{n-1} + \lambda^n \quad (3.27)$$

Let

$$A + BKC = H$$

so that

$$BKC = H - A = Q \quad (\text{say}) \quad (3.28)$$

Using (2.13) we can write (3.28) as

$$(C' \otimes B) \text{vec } K = \text{vec } Q \quad (3.29)$$

or more simply as

$$Pk = q \quad (3.30)$$

where $P = C' \otimes B$, $k = \text{vec } K$ and $q = \text{vec } Q$.

Notice that P is of order $(n^2 \times mr)$ and k and q are column vectors of order mr and n^2 respectively.

The system of equations (3.30) is overdetermined unless of course $m = n = r$, in which case can be solved in the usual manner — assuming a solution does exist!

In general, to solve the system for k we must consider the subsystem of linearly independent equations, the remaining equations being linearly dependent

on this subsystem. In other words we determine a nonsingular matrix $T(n^2 \times n^2)$ such that

$$TP = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \quad (3.31)$$

where P_1 is the matrix of the coefficients of the linearly independent equations of the system (3.30) and P_2 is a null matrix.

Premultiplying both sides of (3.30) by T and making use of (3.31), we obtain

$$TPk = Tq$$

or

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} k = \begin{bmatrix} u \\ v \end{bmatrix} \quad (3.32)$$

If the rank of P is mr , then P_1 is of order $(mr \times mr)$, P_2 is of order $([n^2 - mr] \times mr)$ and u and v are of order mr and $(n^2 - mr)$ respectively.

A sufficient condition for the existence of a solution to (3.32) or equivalently to (3.30) is that

$$v = 0 \quad (3.33)$$

in (3.32).

If the condition (3.33) holds and $\text{rank } P_1 = mr$, then

$$k = P_1^{-1}u \quad (3.34)$$

The condition (3.33) depends on an appropriate choice of H . The underlying assumption being made is that a matrix H satisfying this condition does exist. This in turn depends on the system under consideration, for example whether it is controllable.

Some obvious choices for the form of matrix H are: (a) diagonal, (b) upper or lower triangular, (c) companion form or (d) certain combinations of the above forms.

Although forms (a) and (b) are well known, the companion form is less well documented.

Very briefly, the matrix

$$H = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & -a_2 & & -a_{n-1} \end{bmatrix}$$

is said to be in 'companion' form, it has the associated characteristic equation

$$|\lambda I - H| = a_0 + a_1\lambda + \dots + a_{n-1}\lambda^{n-1} + \lambda^n = 0 \quad (3.35)$$

Example 3.7

Determine the feedback matrix K so that the two input – two output system

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ 3 & 3 & 1 \\ 2 & -3 & 2 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} u$$

has closed loop eigenvalues $(-1, -2, -3)$.

Solution

We must first decide on the form of the matrix H .

Since (see (3.28))

$$H - A = BKC$$

and the first row of B is zero, it follows that the first row of

$$H - A$$

must be zero.

We must therefore choose H in the companion form.

Since the characteristic equation of H is

$$(\lambda + 1)(\lambda + 2)(\lambda + 3) = \lambda^3 + 6\lambda^2 + 11\lambda + 6 = 0.$$

$$H = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix} \quad (\text{see (3.35)})$$

and hence (see (3.28))

$$Q = \begin{bmatrix} 0 & 0 & 0 \\ -3 & -3 & 0 \\ -8 & -8 & -8 \end{bmatrix}.$$

$$P = C' \otimes B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

An appropriate matrix T is the following

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It follows that

$$TP = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

and

$$Tq = \begin{bmatrix} 0 \\ -8 \\ -3 \\ -8 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}$$

50

Some Applications of the Kronecker Product

[Ch. 3]

Since

$$P_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad P_1^{-1} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

so that (see (3.34))

$$k = P_1^{-1}u = \begin{bmatrix} -3 \\ 0 \\ 0 \\ -8 \end{bmatrix}.$$

Hence

$$K = \begin{bmatrix} -3 & 0 \\ 0 & -8 \end{bmatrix}$$

Introduction to Matrix Calculus

4.1 INTRODUCTION

It is becoming ever increasingly clear that there is a real need for matrix calculus in fields such as multivariate analysis. There is a strong analogy here with matrix algebra which is such a powerful and elegant tool in the study of linear systems and elsewhere.

Expressions in multivariate analysis can be written in terms of scalar calculus, but the compactness of the equivalent relations in terms of matrices not only leads to a better understanding of the problems involved, but also encourages the consideration of problems which may be too complex to tackle by scalar calculus.

We have already defined the derivative of a matrix with respect to a scalar (see (3.1)), we now generalise this concept. The process is frequently referred to as *formal* or *symbolic* matrix differentiation. The basic definitions involve the partial differentiation of scalar matrix functions with respect to all the elements of a matrix. These derivatives are the elements of a matrix, of the same order as the original matrix, which is defined as the *derived matrix*. The words 'formal' and 'symbolic' refer to the fact that the matrix derivatives are defined without the rigorous mathematical justification which we expect for the corresponding scalar derivatives. This is not to say that such justification cannot be made, rather the fact is that this topic is still in its infancy and that appropriate mathematical basis is being laid as the subject develops. With this in mind we make the following observations about the notation used. In general the elements of the matrices A, B, C, \dots will be constant scalars. On the other hand the elements of the matrices X, Y, Z, \dots are scalar variables and we exclude the possibility that any element can be a constant or zero. In general we will also demand that these elements are independent. When this is not the case, for example when the matrix X is symmetric, is considered as a special case. The reader will appreciate the necessity for these restrictions when he considers the partial derivatives of (say) a matrix X with respect to one of its elements x_{rs} . Obviously the derivative is undefined if x_{rs} is a constant. The derivative is E_{rs} if x_{rs} is independent of all the other elements of X , but is $E_{rs} + E_{sr}$ if X is symmetric.

There have been attempts to define the derivative when x_{rs} is a constant (or zero) but, as far as this author knows, no rigorous mathematical theory for the general case has been proposed and successfully applied.

4.2 THE DERIVATIVES OF VECTORS

Let x and y be vectors of orders n and m respectively. We can define various derivatives in the following way [15]:

- (1) The derivative of the vector y with respect to vector x is the matrix

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \quad (4.1)$$

of order $(n \times m)$ where y_1, y_2, \dots, y_m and x_1, x_2, \dots, x_n are the components of y and x respectively.

- (2) The derivatives of a scalar with respect to a vector. If y is a scalar

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} \quad (4.2)$$

- (3) The derivative of a vector y with respect to a scalar x

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix} \quad (4.3)$$

Example 4.1

Given

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

and

$$y_1 = x_1^2 - x_2$$

$$y_2 = x_3^2 + 3x_2$$

Obtain $\partial y / \partial x$.

Solution

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 & -1 \\ -1 & 3 \\ 0 & 2x_3 \end{bmatrix}.$$

In multivariate analysis, if x and y are of the same order, the absolute value of the determinant of $\partial x / \partial y$, that is of

$$\left| \frac{\partial x}{\partial y} \right|$$

is called the Jacobian of the transformation determined by

$$y = y(x).$$

Example 4.2

The transformation from spherical to cartesian co-ordinates is defined by $x = r \sin \theta \cos \psi$, $y = r \sin \theta \sin \psi$, and $z = r \cos \theta$ where $r > 0$, $0 < \theta < \pi$ and $0 \leq \psi < 2\pi$.

Obtain the Jacobian of the transformation.

Solution

Let

$$x = y_1, \quad y = y_2, \quad z = y_3$$

and

$$r = y_1, \quad \theta = y_2, \quad \psi = y_3,$$

$$J = \left| \frac{\partial x}{\partial y} \right| = \begin{vmatrix} \sin y_2 \cos y_3 & \sin y_2 \sin y_3 & \cos y_2 \\ y_1 \cos y_2 \cos y_3 & y_1 \cos y_2 \sin y_3 & -y_1 \sin y_2 \\ -y_1 \sin y_2 \sin y_3 & y_1 \sin y_2 \cos y_3 & 0 \end{vmatrix}$$

$$= y_1^2 \sin y_2.$$

Definitions (4.1), (4.2) and (4.3) can be used to obtain derivatives to many frequently used expressions, including quadratic and bilinear forms.

For example consider

$$y = x'Ax$$

Using (4.2) it is not difficult to show that

$$\begin{aligned}\frac{\partial y}{\partial x} &= Ax + A'x \\ &= 2Ax \text{ if } A \text{ is symmetric.}\end{aligned}$$

We can of course differentiate the vector $2Ax$ with respect to x , by definition (4.1).

$$\begin{aligned}\frac{\partial}{\partial x} \left(\frac{\partial y}{\partial x} \right) &= \frac{\partial}{\partial x} (2Ax) \\ &= 2A' = 2A \text{ (if } A \text{ is symmetric).}\end{aligned}$$

The following table summarises a number of vector derivative formulae.

y scalar or a vector	$\frac{\partial y}{\partial x}$
Ax	A'
$x'A$	A
$x'x$	$2x$
$x'Ax$	$Ax + A'x$

(4.4)

4.3 THE CHAIN RULE FOR VECTORS

Let

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} \quad \text{and} \quad z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}.$$

Using the definition (4.1), we can write

$$\left(\frac{\partial z}{\partial x} \right)' = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_n} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial z_m}{\partial x_1} & \frac{\partial z_m}{\partial x_2} & \cdots & \frac{\partial z_m}{\partial x_n} \end{bmatrix}. \quad (4.5)$$

Assume that

$$z = y(x)$$

so that

$$\frac{\partial z_l}{\partial x_j} = \sum_{q=1}^r \frac{\partial z_l}{\partial y_q} \frac{\partial y_q}{\partial x_j} \quad \begin{array}{l} l = 1, 2, \dots, m \\ j = 1, 2, \dots, n. \end{array}$$

Then (4.5) becomes

$$\begin{aligned} \left(\frac{\partial z}{\partial x} \right)' &= \begin{bmatrix} \sum \frac{\partial z_1}{\partial y_q} \frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_1}{\partial y_q} \frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_1}{\partial y_q} \frac{\partial y_q}{\partial x_n} \\ \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_n} \\ \vdots & & & \\ \sum \frac{\partial z_m}{\partial y_q} \frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_m}{\partial y_q} \frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_m}{\partial y_q} \frac{\partial y_q}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial z_1}{\partial y_1} \frac{\partial z_1}{\partial y_2} \cdots \frac{\partial z_1}{\partial y_r} \\ \frac{\partial z_2}{\partial y_1} \frac{\partial z_2}{\partial y_2} \cdots \frac{\partial z_2}{\partial y_r} \\ \vdots \\ \frac{\partial z_m}{\partial y_1} \frac{\partial z_m}{\partial y_2} \cdots \frac{\partial z_m}{\partial y_r} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} \frac{\partial y_1}{\partial x_2} \cdots \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} \frac{\partial y_2}{\partial x_2} \cdots \frac{\partial y_2}{\partial x_n} \\ \vdots \\ \frac{\partial y_r}{\partial x_1} \frac{\partial y_r}{\partial x_2} \cdots \frac{\partial y_r}{\partial x_n} \end{bmatrix} \\ &= \left(\frac{\partial z}{\partial y} \right)' \left(\frac{\partial y}{\partial x} \right)' \quad (\text{by (4.1)}) \\ &= \left(\frac{\partial y}{\partial x} \frac{\partial z}{\partial y} \right)' \end{aligned}$$

on transporting both sides, we finally obtain

$$\frac{\partial z}{\partial y} = \frac{\partial y}{\partial x} \frac{\partial z}{\partial y} \quad (4.6)$$

4.4 THE DERIVATIVE OF SCALAR FUNCTIONS OF A MATRIX WITH RESPECT TO THE MATRIX

Let $X = [x_{ij}]$ be a matrix of order $(m \times n)$ and let

$$y = f(X)$$

be a scalar function of X .

The derivative of y with respect to X , denoted by

$$\frac{\partial y}{\partial X}$$

is defined as the following matrix of order $(m \times n)$

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y}{\partial x_{ij}} \end{bmatrix} = \sum_{i,j} E_{ij} \frac{\partial y}{\partial x_{ij}} \quad (4.7)$$

where E_{ij} is an elementary matrix of order $(m \times n)$.

Definition

When $X = [x_{ij}]$ is a matrix of order $(m \times n)$ and $y = f(X)$ is a scalar function of X , then $\partial f(X)/\partial X$ is known as a gradient matrix.

Example 4.3

Given the matrix $X = [x_{ij}]$ of order $(n \times n)$ obtain $\partial y/\partial X$ when $y = \text{tr } X$.

Solution

$y = \text{tr } X = x_{11} + x_{22} + \cdots + x_{nn} = \text{tr } X'$ (see 1.33) hence by (4.7)

$$\frac{\partial y}{\partial X} = I_n.$$

An important family of derivatives with respect to a matrix involves functions of the determinant of a matrix, for example

$$y = |X| \quad \text{or} \quad y = |AX|.$$

We will consider a general case, say we have a matrix $Y = [y_{ij}]$ whose components are functions of a matrix $X = [x_{ij}]$, that is

$$y_{ij} = f_{ij}(x)$$

where $x = [x_{11} \ x_{12} \ \cdots \ x_{mn}]'$.

We will determine

$$\frac{\partial |Y|}{\partial x_{rs}}$$

which will allow us to build up the matrix

$$\frac{\partial |Y|}{\partial X}.$$

Using the chain rule we can write

$$\frac{\partial |Y|}{\partial x_{rs}} = \sum_i \sum_j \frac{\partial |Y|}{\partial y_{ij}} \cdot \frac{\partial y_{ij}}{\partial x_{rs}}.$$

$$\text{But } |Y| = \sum_j y_{ij} Y_{ij}$$

where Y_{ij} is the cofactor of the element y_{ij} in $|Y|$. Since the cofactors Y_{i1}, Y_{i2}, \dots are independent of the element y_{ij} , we have

$$\frac{\partial |Y|}{\partial y_{ij}} = Y_{ij}.$$

It follows that

$$\frac{\partial |Y|}{\partial x_{rs}} = \sum_i \sum_j Y_{ij} \frac{\partial y_{ij}}{\partial x_{rs}}. \quad (4.8)$$

Although we have achieved our objective in determining the above formula, it can be written in an alternate and useful form.

With

$$a_{ij} = Y_{ij} \quad \text{and} \quad b_{ij} = \frac{\partial y_{ij}}{\partial x_{rs}}$$

we can write (4.8) as

$$\begin{aligned} \frac{\partial |Y|}{\partial x_{rs}} &= \sum_i \sum_j a_{ij} b_{ij} = \sum_i \sum_j a_{ij} b_{ij} e'_j e_j \\ &= \sum_i \sum_j a_{ij} e'_j b_{ij} e_j \\ &= \sum_i A_i' B_i. \quad (\text{by (1.23) and (1.24)}) \\ &= \text{tr}(AB') = \text{tr}(B'A) \quad (\text{by (1.43)}) \end{aligned}$$

where $A = [a_{ij}]$ and $B = [b_{ij}]$.

Assuming that Y is of order $(k \times k)$ let

$$\begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1k} \\ Y_{21} & Y_{22} & \dots & Y_{2k} \\ \vdots & & & \\ Y_{k1} & Y_{k2} & \dots & Y_{kk} \end{bmatrix} = Z \quad (4.9)$$

and since

$$\left[\frac{\partial y_{ij}}{\partial x_{rs}} \right] = \frac{\partial Y}{\partial x_{rs}}$$

we can write

$$\frac{\partial |Y|}{\partial x_{rs}} = \text{tr} \left(\frac{\partial Y'}{x_{rs}} Z \right). \quad (4.10)$$

We use (4.10) to evaluate $\partial |Y| / \partial x_{11}$, $\partial |Y| / \partial x_{12}$, ..., $\partial |Y| / \partial x_{mn}$ and then use (4.7) to construct

$$\frac{\partial |Y|}{\partial X}.$$

Example 4.4

Given the matrix $X = [x_{ij}]$ of order (2×2) evaluate $\partial |X| / \partial X$,

- (i) when all the components x_{ij} of X are independent
- (ii) when X is a symmetric matrix.

Solution

(i) In the notation of (4.10), we have

$$Y = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

so that $\partial Y / \partial x_{rs} = E_{rs}$ (for notation see (1.4)).

As

$$Z = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix},$$

we use the result of Example (1.4) to write (4.10) as

$$\frac{\partial |Y|}{\partial x_{rs}} = (\text{vec } E_{rs})' \text{vec } Z.$$

So that, for example

$$\frac{\partial |Y|}{\partial x_{11}} = [1 \ 0 \ 0 \ 0] \begin{bmatrix} X_{11} \\ X_{21} \\ X_{12} \\ X_{22} \end{bmatrix} = X_{11}$$

and

$$\frac{\partial |Y|}{\partial x_{12}} = [0 \ 0 \ 1 \ 0] \begin{bmatrix} X_{11} \\ X_{21} \\ X_{12} \\ X_{22} \end{bmatrix} = X_{12} \text{ and so on.}$$

Hence

$$\begin{aligned} \frac{\partial |Y|}{\partial X} &= \frac{\partial |X|}{\partial X} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \\ &= |X|(X^{-1})' \quad (\text{See [18] p. 124}). \end{aligned}$$

(ii) This time

$$Y = \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix}$$

hence

$$\frac{\partial |Y|}{\partial x_{11}} = E_{11}, \quad \frac{\partial |Y|}{\partial x_{12}} = E_{12} + E_{21} \quad \text{and so on.}$$

(See the introduction to Chapter 4 for explanation of the notation.)

It follows that

$$\frac{\partial |Y|}{\partial x_{12}} = \frac{\partial |Y|}{\partial x_{21}} = [0 \ 1 \ 1 \ 0] \begin{bmatrix} X_{11} \\ X_{21} \\ X_{12} \\ X_{22} \end{bmatrix} = X_{21} + X_{12} = 2X_{12}$$

(Since $X_{12} = X_{21}$)

hence

$$\frac{\partial |Y|}{\partial X} = \begin{bmatrix} X_{11} & 2X_{12} \\ 2X_{21} & X_{22} \end{bmatrix} = 2 \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} - \begin{bmatrix} X_{11} & 0 \\ 0 & X_{22} \end{bmatrix}.$$

The above results can be generalised to a matrix X of order $(n \times n)$.

We obtain, in the symmetric matrix case

$$\frac{\partial |X|}{\partial X} = 2[X_{ij}] - \text{diag}\{X_{ii}\}.$$

We defer the discussion of differentiating other scalar matrix functions to Chapter 5.

4.5 THE DERIVATIVE OF A MATRIX WITH RESPECT TO ONE OF ITS ELEMENTS AND CONVERSELY

In this section we will generalise the concepts discussed in the previous section. We again consider a matrix

$$X = [x_{ij}] \text{ or order } (m \times n).$$

The derivative of the matrix X relative to one of its elements x_{rs} (say), is obviously (see (3.1))

$$\frac{\partial X}{\partial x_{rs}} = E_{rs} \quad (4.11)$$

where E_{rs} is the elementary matrix of order $(m \times n)$ (the order of X) defined in section 1.2.

It follows immediately that

$$\frac{\partial X'}{\partial x_{rs}} = E'_{rs}. \quad (4.12)$$

A more complicated situation arises when we consider a product of the form

$$Y = AXB \quad (4.13)$$

where

$$X = [x_{ij}] \text{ is of order } (m \times n)$$

$$A = [a_{ij}] \text{ is of order } (l \times m)$$

$$B = [b_{ij}] \text{ is of order } (n \times q)$$

and

$$Y = [y_{ij}] \text{ is of order } (l \times q).$$

A and B are assumed independent of X .

Our aim is to find the rule for obtaining the derivatives

$$\frac{\partial Y}{\partial x_{rs}} \quad \text{and} \quad \frac{\partial y_{ij}}{\partial x}$$

where x_{rs} is a typical element of X and y_{ij} is a typical element of Y .

We will first obtain the (i, j) th element y_{ij} in (4.13) as a function of the elements of X .

We can achieve this objective in a number of different ways. For example, we can use (2.13) to write

$$\text{vec } Y = (B' \otimes A) \text{vec } X.$$

From this expression we see that y_{ij} is the (scalar) product of the i th row of

$$[b_{1j}A; b_{2j}A; \dots; b_{nj}A] \text{ and } \text{vec } X,$$

so that

$$y_{ij} = \sum_{p=1}^n \sum_{l=1}^m a_{il} b_{pj} x_{lp} \quad (4.14)$$

From (4.14) we immediately obtain

$$\frac{\partial y_{ij}}{\partial x_{rs}} = a_{ir} b_{sj} \quad (4.15)$$

We can now write the expression for $\partial y_{ij} / \partial X$,

$$\frac{\partial y_{ij}}{\partial X} = \begin{bmatrix} \frac{\partial y_{ij}}{\partial x_{11}} & \frac{\partial y_{ij}}{\partial x_{12}} & \dots & \frac{\partial y_{ij}}{\partial x_{1n}} \\ \frac{\partial y_{ij}}{\partial x_{21}} & \frac{\partial y_{ij}}{\partial x_{22}} & \dots & \frac{\partial y_{ij}}{\partial x_{2n}} \\ \vdots & & & \\ \frac{\partial y_{ij}}{\partial x_{m1}} & \frac{\partial y_{ij}}{\partial x_{m2}} & \dots & \frac{\partial y_{ij}}{\partial x_{mn}} \end{bmatrix} \quad (4.16)$$

Using (4.15), we obtain

$$\frac{\partial y_{ij}}{\partial X} = \begin{bmatrix} a_{i1} b_{1j} & a_{i1} b_{2j} & \dots & a_{i1} b_{nj} \\ a_{i2} b_{1j} & a_{i2} b_{2j} & \dots & a_{i2} b_{nj} \\ \vdots & & & \\ a_{im} b_{1j} & a_{im} b_{2j} & \dots & a_{im} b_{nj} \end{bmatrix} \quad (4.17)$$

We note that the matrix on the right hand side of (4.17) can be expressed as (for notation see (1.5)(1.13)(1.16) and (1.17))

$$\begin{aligned} \frac{\partial y_{ij}}{\partial X} &= \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{im} \end{bmatrix} [b_{1j} b_{2j} \dots b_{nj}] \\ &= A_i \cdot B_j' \\ &= A' e_i e_j' B'. \end{aligned}$$

So that

$$\frac{\partial y_{ij}}{\partial X} = A'E_{ij}B' \quad (4.18)$$

where E_{ij} is an elementary matrix of order $(l \times q)$ the order of the matrix Y .

We also use (4.14) to obtain an expression for $\partial Y/\partial x_{rs}$.

$$\frac{\partial Y}{\partial x_{rs}} = \begin{bmatrix} \frac{\partial y_{1j}}{\partial x_{rs}} \\ \frac{\partial y_{2j}}{\partial x_{rs}} \\ \vdots \\ \frac{\partial y_{lj}}{\partial x_{rs}} \end{bmatrix} \quad (r, s \text{ fixed, } i, j \text{ variable } 1 \leq i \leq l, 1 \leq j \leq q)$$

that is

$$\frac{\partial Y}{\partial x_{rs}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{rs}} & \frac{\partial y_{12}}{\partial x_{rs}} & \cdots & \frac{\partial y_{1q}}{\partial x_{rs}} \\ \frac{\partial y_{21}}{\partial x_{rs}} & \frac{\partial y_{22}}{\partial x_{rs}} & \cdots & \frac{\partial y_{2q}}{\partial x_{rs}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{l1}}{\partial x_{rs}} & \frac{\partial y_{l2}}{\partial x_{rs}} & \cdots & \frac{\partial y_{lq}}{\partial x_{rs}} \end{bmatrix} = \sum_{i,j} E_{ij} \frac{\partial y_{ij}}{\partial x_{rs}} \quad (4.19)$$

where E_{ij} is an elementary matrix of order $(l \times q)$.

We again use (4.15) to write

$$\begin{aligned} \begin{bmatrix} \frac{\partial y_{ij}}{\partial x_{rs}} \end{bmatrix} &= \begin{bmatrix} a_{1r}b_{s1} & a_{1r}b_{s2} & \cdots & a_{1r}b_{sq} \\ a_{2r}b_{s1} & a_{2r}b_{s2} & \cdots & a_{2r}b_{sq} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mr}b_{s1} & a_{mr}b_{s2} & \cdots & a_{mr}b_{sq} \end{bmatrix} \\ &= \begin{bmatrix} a_{1r} \\ a_{2r} \\ \vdots \\ a_{mr} \end{bmatrix} [b_{s1} \ b_{s2} \ \cdots \ b_{sq}] \\ &= A_{\cdot r} B_{s \cdot}' = A e_r e_s' B. \end{aligned}$$

So that

$$\frac{\partial (AXB)}{\partial x_{rs}} = A E_{rs} B \quad (4.20)$$

where E_{rs} is an elementary matrix of order $(m \times n)$, the order of the matrix X .

Example 4.5

Find the derivative $\partial Y/\partial x_{rs}$, given

$$Y = AX'B$$

where the order of the matrices A , X and B is such that the product on the right hand side is defined.

Solution

By the method used above to obtain the derivative $\partial/\partial x_{rs} (AXB)$, we find

$$\frac{\partial}{\partial x_{rs}} (AX'B) = AE'_{rs}B.$$

Before continuing with further examples we need a rule for determining the derivative of a product of matrices.

Consider

$$Y = UV \quad (4.21)$$

where $U = [u_{ij}]$ is of order $(m \times n)$ and $V = [v_{ij}]$ is of order $(n \times l)$ and both U and V are functions of a matrix X .

We wish to determine

$$\frac{\partial Y}{\partial x_{rs}} \quad \text{and} \quad \frac{\partial y_{ij}}{\partial X}.$$

The (i,j) th element of (4.21) is

$$y_{ij} = \sum_{p=1}^n u_{ip} v_{pj} \quad (4.22)$$

hence

$$\frac{\partial y_{ij}}{\partial x_{rs}} = \sum_{p=1}^n \frac{\partial u_{ip}}{\partial x_{rs}} v_{pj} + \sum_{p=1}^n u_{ip} \frac{\partial v_{pj}}{\partial x_{rs}}. \quad (4.23)$$

For fixed r and s , (4.23) is the (i,j) th element of the matrix $\partial Y/\partial x_{rs}$ of order $(m \times l)$ the same as the order of the matrix Y .

On comparing both the terms on the right hand side of (4.23) with (4.22), we can write

$$\frac{\partial(UV)}{\partial x_{rs}} = \frac{\partial U}{\partial x_{rs}} V + U \frac{\partial V}{\partial x_{rs}} \quad (4.24)$$

as one would expect.

On the other hand, when fixing (i, j) , (4.23) is the (r, s) th element of the matrix $\partial y_{ij}/\partial X$, which is of the same order as the matrix X , that is

$$\frac{\partial y_{ij}}{\partial X} = \sum_{p=1}^n \frac{\partial u_{ip}}{\partial X} v_{pj} + \sum_{p=1}^n u_{ip} \frac{\partial v_{pj}}{\partial X} \quad (4.25)$$

We will make use of the result (4.24) in some of the subsequent examples.

Example 4.6

Let $X = [x_{rs}]$ be a non-singular matrix. Find the derivative $\partial Y/\partial x_{rs}$, given

- (i) $Y = AX^{-1}B$, and
- (ii) $Y = X'AX$

Solution

(i) Using (4.24) to differentiate

$$YY^{-1} = I,$$

we obtain

$$\frac{\partial Y}{\partial x_{rs}} Y^{-1} + Y \frac{\partial Y^{-1}}{\partial x_{rs}} = 0,$$

hence

$$\frac{\partial Y}{\partial x_{rs}} = -Y \frac{\partial Y^{-1}}{\partial x_{rs}} Y.$$

But by (4.20)

$$\frac{\partial Y^{-1}}{\partial x_{rs}} = \frac{\partial}{\partial x_{rs}} (B^{-1}XA^{-1}) = B^{-1}E_{rs}A^{-1}$$

so that

$$\begin{aligned} \frac{\partial Y}{\partial x_{rs}} &= \frac{\partial}{\partial x_{rs}} (AX^{-1}B) = -AX^{-1}BB^{-1}E_{rs}A^{-1}AX^{-1}B \\ &= -AX^{-1}E_{rs}X^{-1}B. \end{aligned}$$

(ii) Using (4.24), we obtain

$$\begin{aligned} \frac{\partial Y}{\partial x_{rs}} &= \frac{\partial X'}{\partial x_{rs}} AX + X' \frac{\partial (AX)}{\partial x_{rs}} \\ &= E'_{rs}AX + X'AE_{rs} \quad (\text{by (4.12) and (4.20)}). \end{aligned}$$

Both (4.18) and (4.20) were derived from (4.15) which is valid for all i, j and r, s , defined by the orders of the matrices involved.

The First Transformation Principle

It follows that (4.18) is a transformation of (4.20) and conversely. To obtain (4.18) from (4.20) we replace A by A' , B by B' and E_{rs} by E_{ij} (careful, E_{rs} and E_{ij} may be of different orders).

The interesting point is that although (4.18) and (4.20) were derived for constant matrices A and B , the above transformation is independent of the status of the matrices and is valid even when A and B are functions of X .

Example 4.7

Find the derivative of $\partial y_{ij}/\partial X$, given

- (i) $Y = AX'B$,
- (ii) $Y = AX^{-1}B$, and
- (iii) $Y = X'AX$

where $X = [x_{ij}]$ is a nonsingular matrix.

Solution

(i) Let $W = X'$, then

$$Y = AWB \text{ so that by (4.20) } \frac{\partial Y}{\partial w_{rs}} = AE_{rs}B$$

hence

$$\frac{\partial y_{ij}}{\partial w} = A'E_{ij}B'.$$

But

$$\frac{\partial y_{ij}}{\partial X} = \frac{\partial y_{ij}}{\partial w'} = \left(\frac{\partial y_{ij}}{\partial w} \right)'$$

hence

$$\frac{\partial y_{ij}}{\partial X} = BE'_{ij}A$$

(ii) From Example 4.6(i)

$$\frac{\partial Y}{\partial x_{rs}} = -AX^{-1}E_{rs}X^{-1}B.$$

Let $A_1 = AX^{-1}$ and $B_1 = X^{-1}B$, then

$$\frac{\partial Y}{\partial x_{rs}} = -A_1E_{rs}B_1$$

so that

$$\frac{\partial y_{ij}}{\partial X} = -A'_1E_{ij}B'_1 = -(X^{-1})'A'E_{ij}B'(X^{-1})'.$$

(iii) From Example 4.6(ii)

$$\frac{\partial Y}{\partial x_{rs}} = E'_{rs}AX + X'AE_{rs}.$$

Let $A_1 = I, B_1 = AX, A_2 = X'A$ and $B_2 = I$, then

$$\frac{\partial X}{\partial x_{rs}} = A_1 E'_{rs} B_1 + A_2 E_{rs} B_2.$$

The second term on the right hand side is in standard form. The first term is in the form of the solution to Example 4.5 for which the derivative $\partial y_{ij}/\partial X$ was found in (i) above, hence

$$\begin{aligned}\frac{\partial y_{ij}}{\partial X} &= B_1 E'_{ij} A_1 + A'_2 E_{ij} B'_2 \\ &= AX E'_{ij} + A' X E_{ij}.\end{aligned}$$

It is interesting to compare this last result with the example in section 4.2 when we considered the scalar $y = x'Ax$.

In this special case when the matrix X has only one column, the elementary matrix which is of the same order as Y , becomes

$$E_{ij} = E'_{ij} = 1.$$

Hence

$$\frac{\partial y_{ij}}{\partial X} = \frac{\partial y}{\partial x} = Ax + A'x$$

which is the result obtained in section 4.2 (see (4.4)).

Conversely using the above techniques we can also obtain the derivatives of the matrix equivalents of the other equations in the table (4.4).

Example 4.8

Find

$$\frac{\partial Y}{\partial x_{rs}} \quad \text{and} \quad \frac{\partial y_{ij}}{\partial X}$$

when

(i) $Y = AX$, and

(ii) $Y = X'X$.

Solution

(i) With $B = I$, apply (4.20)

$$\frac{\partial Y}{\partial x_{rs}} = AE_{rs}.$$

The transformation principle results in

$$\frac{\partial y_{ij}}{\partial X} = A' E_{ij}.$$

(ii) This is a special case of Example 4.6(ii) in which $A = I$.

We have found the solution

$$\frac{\partial Y}{\partial x_{rs}} = E'_{rs} X + X' E_{rs}$$

and (Solution to Example 4.7(iii))

$$\frac{\partial y_{ij}}{\partial X} = X E'_{ij} + X E_{ij}.$$

4.6 THE DERIVATIVES OF THE POWERS OF A MATRIX

Our aim in this section is to obtain the rules for determining

$$\frac{\partial Y}{\partial x_{rs}} \quad \text{and} \quad \frac{\partial y_{ij}}{\partial X}$$

when

$$Y = X^n.$$

Using (4.24) when $U = V = X$ so that

$$Y = X^2$$

we immediately obtain

$$\frac{\partial Y}{\partial x_{rs}} = E'_{rs} X + X E_{rs}$$

and, applying the first transformation principle,

$$\frac{\partial y_{ij}}{\partial X} = E_{ij} X' + X' E_{ij}.$$

It is instructive to repeat this exercise with

$$U = X^2 \quad \text{and} \quad V = X$$

so that

$$Y = X^3.$$

We obtain

$$\frac{\partial Y}{\partial x_{rs}} = E'_{rs} X^2 + X E'_{rs} X + X^2 E_{rs}$$

and

$$\frac{\partial y_{ij}}{\partial X} = E_{ij} (X')^2 + X' E_{ij} X' + (X')^2 E_{ij}.$$

More generally, it can be proved by induction, that for

$$Y = X^n$$

$$\frac{\partial Y}{\partial x_{rs}} = \sum_{k=0}^{n-1} X^k E_{rs} X^{n-k-1} \quad (4.26)$$

where by definition $X^0 = I$, and

$$\frac{\partial y_{ij}}{\partial X} = \sum_{k=1}^{n-1} (X')^k E_{ij} (X')^{n-k-1} \quad (4.27)$$

Example 4.9

Using the result (4.26), obtain $\partial Y / \partial x_{rs}$ when

$$Y = X^{-n}$$

Solution

Using (4.24) on both sides of

$$X^{-n} X^n = I$$

we find

$$\frac{\partial(X^{-n})}{\partial x_{rs}} X^n + X^{-n} \frac{\partial(X^n)}{\partial x_{rs}} = 0$$

so that

$$\frac{\partial(X^{-n})}{\partial x_{rs}} = -X^{-n} \frac{\partial(X^n)}{\partial x_{rs}} X^{-n}.$$

Now making use of (4.26), we conclude that

$$\frac{\partial(X^{-n})}{\partial x_{rs}} = -X^{-n} \left[\sum_{k=0}^{n-1} X^k E_{rs} X^{n-k-1} \right] X^{-n}.$$

Problems for Chapter 4

(1) Given

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}, \quad Y = \begin{bmatrix} x & e^{2x} \\ 1 & x^{-1} \\ 2x^2 & \sin x \end{bmatrix}$$

and $y = 2x_{11}x_{22} - x_{21}x_{13}$, calculate

$$\frac{\partial y}{\partial X} \quad \text{and} \quad \frac{\partial Y}{\partial x}.$$

(2) Given

$$X = \begin{bmatrix} \sin x & x \\ \cos x & e^x \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} \sin x & e^x \\ e^x & x \end{bmatrix}$$

evaluate

$$\frac{\partial |X|}{\partial X}$$

by

- (a) a direct method
(b) use of a derivative formula.

(3) Given

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \quad \text{and} \quad Y = X'X,$$

use a direct method to evaluate

$$(a) \frac{\partial Y}{\partial x_{21}} \quad \text{and} \quad (b) \frac{\partial y_{13}}{\partial X}$$

(4) Obtain expressions for

$$\frac{\partial Y}{\partial x_{rs}} \quad \text{and} \quad \frac{\partial y_{ij}}{\partial X}$$

when

$$(a) Y = XAX \quad \text{and} \quad (b) Y = X'AX'.$$

(5) Obtain an expression for $\partial |AXB| / \partial x_{rs}$. It is assumed AXB is non-singular.(6) Evaluate $\partial Y / \partial x_{rs}$ when

$$(a) Y = X(X')^2 \quad \text{and} \quad (b) Y = (X')^2 X.$$

CHAPTER 5

Further Development of Matrix Calculus including an Application of Kronecker Products

5.1 INTRODUCTION

In Chapter 4 we discussed rules for determining the derivatives of a vector and then the derivatives of a matrix.

But it will be remembered that when Y is a matrix, then $\text{vec } Y$ is a vector. This fact, together with the closely related Kronecker product techniques discussed in Chapter 2 will now be exploited to derive some interesting results.

Also we explore further the derivatives of some scalar functions with respect to a matrix first considered in the previous chapter.

5.2 DERIVATIVES OF MATRICES AND KRONECKER PRODUCTS

In the previous chapter we have found $\partial y_{ij}/\partial X$ when

$$Y = AXB \quad (5.1)$$

where $Y = [y_{ij}]$, $A = [a_{ij}]$, $X = [x_{ij}]$ and $B = [b_{ij}]$.

We now obtain $(\partial \text{vec } Y)/(\partial \text{vec } X)$ for (5.1). We can write (5.1) as

$$y = Px \quad (5.2)$$

where $y = \text{vec } Y$, $x = \text{vec } X$ and $P = B' \otimes A$.

By (4.1), (4.4) and (2.10)

$$\frac{\partial y}{\partial x} = P' = (B' \otimes A)' = B \otimes A'. \quad (5.3)$$

The corresponding result for the equation

$$Y = AX'B \quad (5.4)$$

is not so simple.

The problem is that when we write (5.4) in the form of (5.2), we have this time

$$y = Pz \quad (5.5)$$

where $z = \text{vec } X'$

We can find (see (2.25)) a permutation matrix U such that

$$\text{vec } X' = U \text{vec } X \quad (5.6)$$

in which case (5.5) becomes

$$y = PUX$$

so that

$$\frac{\partial y}{\partial x} = (PU)' = U'(B \otimes A') \quad (5.7)$$

It is convenient to write

$$U'(B \otimes A') = (B \otimes A')_{(n)} \quad (5.8)$$

U' is seen to premultiply the matrix $(B \otimes A')$. Its effect is therefore to rearrange the rows of $(B \otimes A')$.

In fact the first and every subsequent n th row of $(B \otimes A')$ form the first consecutive m rows of $(B \otimes A')_{(n)}$. The second and every subsequent n th row form the next m consecutive rows of $(B \otimes A')_{(n)}$ and so on.

A special case of this notation is for $n = 1$, then

$$(B \otimes A')_{(1)} = B \otimes A' \quad (5.9)$$

Now, returning to (5.5), we obtain, by comparison with (5.3)

$$\frac{\partial y}{\partial x} = (B \otimes A')_{(n)} \quad (5.10)$$

Example 5.1

Obtain $(\partial \text{vec } Y)/(\partial \text{vec } X)$, given $X = [x_{ij}]$ of order $(m \times n)$, when

- (i) $Y = AX$, (ii) $Y = XA$, (iii) $Y = AX'$ and (iv) $Y = X'A$.

Solution

Let $y = \text{vec } Y$ and $x = \text{vec } X$.

- (i) Use (5.3) with $B = I$

$$\frac{\partial y}{\partial x} = I \otimes A'.$$

(ii) Use (5.3)

$$\frac{\partial y}{\partial x} = A \otimes I.$$

(iii) Use (5.10)

$$\frac{\partial y}{\partial x} = (I \otimes A')_{(n)}.$$

(iv) Use (5.10)

$$\frac{\partial y}{\partial x} = (A \otimes I)_{(n)}.$$

5.3 THE DETERMINATION OF $(\partial \text{vec } X)/(\partial \text{vec } Y)$ FOR MORE COMPLICATED EQUATIONS

In this section we wish to determine the derivative $(\partial \text{vec } Y)/(\partial \text{vec } X)$ when, for example,

$$Y = X'AX \quad (5.11)$$

where X is of order $(m \times n)$.

Since Y is a matrix of order $(n \times n)$, it follows that $\text{vec } Y$ and $\text{vec } X$ are vectors of order nn and nm respectively.

With the usual notation

$$Y = [y_{ij}], \quad X = [x_{ij}]$$

we have, by definition (4.1),

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{11}} & \cdots & \frac{\partial y_{nn}}{\partial x_{11}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{21}} & \cdots & \frac{\partial y_{nn}}{\partial x_{21}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_{11}}{\partial x_{mn}} & \frac{\partial y_{21}}{\partial x_{mn}} & \cdots & \frac{\partial y_{nn}}{\partial x_{mn}} \end{bmatrix}. \quad (5.12)$$

But by definition (4.19),

the first row of the matrix (5.12) is $\left(\text{vec } \frac{\partial Y}{\partial x_{11}} \right)'$,

the second row of the matrix (5.12) is $\left(\text{vec } \frac{\partial Y}{\partial x_{21}} \right)'$, etc.

We can therefore write (5.12) as

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X} = \left[\text{vec } \frac{\partial Y}{\partial x_{11}} : \text{vec } \frac{\partial Y}{\partial x_{21}} : \dots : \text{vec } \frac{\partial Y}{\partial x_{mn}} \right]' \quad (5.13)$$

We now use the solution to Example (4.6) where we had established that

$$\text{when } Y = X'AX, \text{ then } \frac{\partial Y}{\partial x_{rs}} = E'_{rs}AX + X'AE_{rs}. \quad (5.14)$$

It follows that

$$\begin{aligned} \text{vec } \frac{\partial Y}{\partial x_{rs}} &= \text{vec } E'_{rs}AX + \text{vec } X'AE_{rs} \\ &= (X'A' \otimes I) \text{vec } E'_{rs} + (I \otimes X'A) \text{vec } E_{rs} \end{aligned} \quad (5.15)$$

(using (2.13)).

Substituting (5.15) into (5.13) we obtain

$$\begin{aligned} \frac{\partial \text{vec } Y}{\partial \text{vec } X} &= [(X'A' \otimes I)[\text{vec } E'_{11} : \text{vec } E'_{21} : \dots : \text{vec } E'_{mn}]]' \\ &\quad + [(I \otimes X'A)[\text{vec } E_{11} : \text{vec } E_{21} : \dots : \text{vec } E_{mn}]]' \\ &= [\text{vec } E'_{11} : \text{vec } E'_{21} : \dots : \text{vec } E'_{mn}]'(AX \otimes I) \\ &\quad + [\text{vec } E_{11} : \text{vec } E_{21} : \text{vec } E_{mn}]'(I \otimes A'X) \end{aligned} \quad (5.16)$$

(by (2.10)).

The matrix

$$[\text{vec } E_{11} : \text{vec } E_{21} : \dots : \text{vec } E_{mn}]$$

is the unit matrix I of order $(mn \times mn)$.

Using (2.23) we can write (5.16) as

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X} = U'(AX \otimes I) + (I \otimes A'X).$$

That is

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X} = (AX \otimes I)_{(0)} + (I \otimes A'X). \quad (5.17)$$

In the above calculations we have used the derivative $\partial Y/\partial x_{rs}$ to obtain $(\partial \text{vec } Y)/(\partial \text{vec } X)$.

The Second Transformation Principle-

Only slight modifications are needed to generalise the above calculations and show that whenever

$$\frac{\partial Y}{\partial x_{rs}} = AE'_{rs}B + CE'_{rs}D$$

where A , B , C and D may be functions of X , then

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X} = B \otimes A' + (D \otimes C')_{(n)} \quad (5.18)$$

We will refer to the above result as the second transformation principle.

Example 5.2

Find

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X}$$

when

$$(i) Y = X'X \quad (ii) Y = AX^{-1}B$$

Solution

Let $y = \text{vec } Y$ and $x = \text{vec } X$.

(i) From Example 4.8

$$\frac{\partial Y}{\partial x_{rs}} = E'_{rs}X + X'E_{rs}$$

Now use the second transformation principle, to obtain

$$\frac{\partial y}{\partial x} = I \otimes X + (X \otimes I)_{(n)}$$

(ii) From Example 4.6

$$\frac{\partial Y}{\partial x_{rs}} = -AX^{-1}E_{rs}X^{-1}B$$

hence

$$\frac{\partial y}{\partial x} = -(X^{-1}B) \otimes (X^{-1})'A'$$

Hopefully, using the above results for matrices, we should be able to rediscover results for the derivatives of vectors considered in Chapter 4.

For example let X be a column vector x then

$$Y = X'X \text{ becomes } y = x'x \text{ (} y \text{ is a scalar).}$$

The above result for $\partial y / \partial x$ becomes

$$\frac{\partial y}{\partial x} = (I \otimes x) + (x \otimes I)_{(1)}.$$

But the unit vectors involved are of order $(n \times 1)$ which, for the one column vector X is (1×1) . Hence

$$\begin{aligned} \frac{\partial y}{\partial x} &= 1 \otimes x + x \otimes 1 \quad (\text{use (5.9)}) \\ &= x + x = 2x \end{aligned}$$

which is the result found in (4.4).

5.4 MORE ON DERIVATIVES OF SCALAR FUNCTIONS WITH RESPECT TO A MATRIX

In section 4.4 we derived a formula, (4.10), which is useful when evaluating $\partial|Y|/\partial X$ for a large class of scalar matrix functions defined by Y .

Example 5.3

Evaluate the derivatives

$$(i) \frac{\partial \log |X|}{\partial X} \text{ and } (ii) \frac{\partial |X|^r}{\partial X}.$$

Solution

(i) We have

$$\frac{\partial}{\partial x_{rs}} (\log |X|) = \frac{1}{|X|} \frac{\partial |X|}{\partial x_{rs}}.$$

From Example 4.4,

$$\frac{\partial |X|}{\partial X} = |X|(X^{-1})' \quad (\text{non-symmetric case}).$$

Hence

$$\frac{\partial \log |X|}{\partial X} = (X^{-1})'.$$

$$(ii) \quad \frac{\partial |X|^r}{\partial x_{rs}} = r |X|^{r-1} \frac{\partial |X|}{\partial x_{rs}}.$$

Hence

$$\frac{\partial |X|^r}{\partial X} = r|X|^{r-1}(X^{-1})'.$$

Traces of matrices form an important class of scalar matrix functions covering a wide range of applications, particularly in statistics in the formulation of least squares and various optimisation problems.

Having discussed the evaluation of the derivative $\partial Y/\partial x_{rs}$ for various products of matrices, we can now apply these results to the evaluation of the derivative

$$\frac{\partial(\text{tr } Y)}{\partial X}.$$

We first note that

$$\frac{\partial(\text{tr } Y)}{\partial X} = \left[\frac{\partial(\text{tr } Y)}{\partial x_{rs}} \right] \quad (5.19)$$

where the bracket on the right hand side of (5.19) denotes, (as usual) a matrix of the same order as X , defined by its (r, s) th element.

As a consequence of (5.19) or perhaps more clearly seen from the definition (4.7), we note that on transposing X , we have

$$\frac{\partial(\text{tr } Y)}{\partial X'} = \left(\frac{\partial(\text{tr } Y)}{\partial X} \right)'. \quad (5.20)$$

Another, and possibly an obvious property of a trace is found when considering the definition of $\partial Y/\partial x_{rs}$ (see (4.19)).

Assuming that $Y = [y_{ij}]$ is of order $(n \times n)$

$$\begin{aligned} \text{tr } \frac{\partial Y}{\partial x_{rs}} &\equiv \frac{\partial y_{11}}{\partial x_{rs}} + \frac{\partial y_{22}}{\partial x_{rs}} + \dots + \frac{\partial y_{nn}}{\partial x_{rs}} \\ &= \frac{\partial}{\partial x_{rs}} (y_{11} + y_{22} + \dots + y_{nn}). \end{aligned}$$

Hence,

$$\text{tr } \frac{\partial Y}{\partial x_{rs}} = \frac{\partial(\text{tr } Y)}{\partial x_{rs}}. \quad (5.21)$$

Example 5.4

Evaluate

$$\frac{\partial \text{tr}(AX)}{\partial X}.$$

Solution

$$\begin{aligned}
 \frac{\partial \operatorname{tr}(AX)}{\partial x_{rs}} &= \operatorname{tr} \frac{\partial (AX)}{\partial x_{rs}} && \text{by (5.21)} \\
 &= \operatorname{tr}(AE_{rs}) && \text{by Example (4.8)} \\
 &= \operatorname{tr}(E'_{rs}A') && \text{since } \operatorname{tr} Y = \operatorname{tr} Y' \\
 &= (\operatorname{vec} E_{rs})' (\operatorname{vec} A') && \text{by Example (1.4).}
 \end{aligned}$$

Hence,

$$\frac{\partial \operatorname{tr}(AX)}{\partial X} = A'.$$

As we found in the previous chapter we can use the derivative of the trace of one product to obtain the derivative of the trace of a different product.

Example 5.5

Evaluate

$$\frac{\partial \operatorname{tr}(AX')}{\partial X}.$$

Solution

From the previous result

$$\frac{\partial \operatorname{tr}(BX)}{\partial X} = \frac{\partial \operatorname{tr}(X'B')}{\partial X} = B'.$$

Let $A' = B$ in the above equation, it follows that

$$\frac{\partial \operatorname{tr}(X'A)}{\partial X} = \frac{\partial \operatorname{tr}(A'X)}{\partial X} = A.$$

The derivatives of traces of more complicated matrix products can be found similarly.

Example 5.6

Evaluate

$$\frac{\partial (\operatorname{tr} Y)}{\partial Y},$$

when

$$(i) \ Y = X'AX$$

$$(ii) \ Y = X'AXB$$

Solution

It is obvious that (i) follows from (ii) when $B = I$.

(ii) $Y = X_1 B$ where $X_1 = X'AX$.

$$\begin{aligned}\frac{\partial Y}{\partial x_{rs}} &= \frac{\partial X_1}{\partial x_{rs}} B \\ &= E'_{rs} AXB + X'AE_{rs}B \quad (\text{by Example 4.6})\end{aligned}$$

Hence,

$$\begin{aligned}\text{tr}\left(\frac{\partial Y}{\partial x_{rs}}\right) &= \text{tr}(E'_{rs}AXB) + \text{tr}(X'AE_{rs}B) \\ &= \text{tr}(E'_{rs}AXB) + \text{tr}(E'_{rs}A'XB') \\ &= (\text{vec } E_{rs})' \text{vec}(AXB) + (\text{vec } E_{rs})' \text{vec}(A'XB').\end{aligned}$$

It follows that

$$\frac{\partial(\text{tr } Y)}{\partial X} = AXB + A'XB'.$$

(i) Let $B = I$ in the above equation, we obtain

$$\frac{\partial(\text{tr } Y)}{\partial X} = AX + A'X = (A + A')X.$$

5.5 THE MATRIX DIFFERENTIAL

For a scalar function $f(x)$ where $x = [x_1 \ x_2 \ \dots \ x_n]'$, the differential df is defined as

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i. \quad (5.23)$$

Corresponding to this definition we define the matrix differential dX for the matrix $X = [x_{ij}]$ of order $(m \times n)$ to be

$$dX = \begin{bmatrix} dx_{11} & dx_{12} & \dots & dx_{1n} \\ dx_{21} & dx_{22} & \dots & dx_{2n} \\ \vdots & & & \\ dx_{m1} & dx_{m2} & \dots & dx_{mn} \end{bmatrix}. \quad (5.24)$$

The following two results follow immediately:

$$d(\alpha X) = \alpha(dX) \quad (\text{where } \alpha \text{ is a scalar}) \quad (5.25)$$

$$d(X + Y) = dX + dY. \quad (5.26)$$

Consider now $X = [x_{ij}]$ of order $(m \times n)$ and $Y = [y_{ij}]$ of order $(n \times p)$.

$$XY = [\sum_j x_{ij} y_{jk}],$$

hence

$$\begin{aligned} d(XY) &= d\left[\sum_j x_{ij} y_{jk}\right] \\ &= \left[\sum_j d(x_{ij} y_{jk})\right] \\ &= \left[\sum_j (dx_{ij}) y_{jk}\right] + \left[\sum_j x_{ij} (dy_{jk})\right]. \end{aligned}$$

It follows that

$$d(XY) = (dX)Y + X(dY). \quad (5.27)$$

Example 5.7

Given $X = [x_{ij}]$ a nonsingular matrix, evaluate

$$(i) \ d|X|, \quad (ii) \ d(X^{-1})$$

Solution

(i) By (5.23)

$$\begin{aligned} d|X| &= \sum_{i,j} \frac{\partial |X|}{\partial x_{ij}} (dx_{ij}) \\ &= \sum_{i,j} X_{ij} (dx_{ij}) \end{aligned}$$

since $(\partial |X|)/(\partial x_{ij}) = X_{ij}$, the cofactor of x_{ij} in $|X|$.

By an argument similar to the one used in section 4.4, we can write

$$d|X| = \text{tr} \{Z'(dX)\} \quad (\text{compare with (4.10)})$$

where $Z = [X_{ij}]$.

Since $Z' = |X|X^{-1}$, we can write

$$d|X| = |X| \text{tr} \{X^{-1}(dX)\}.$$

(ii) Since

$$X^{-1}X = I$$

we use (5.27) to write

$$d(X^{-1})X + X^{-1}(dX) = 0.$$

Hence

$$d(X^{-1}) = -X^{-1}(dX)X^{-1}$$

(compare with Example 4.6).

Notice that if X is a symmetric matrix, then

$$X = X'$$

and

$$(dX)' = dX. \quad (5.28)$$

Problems for Chapter 5

(1) Consider

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \quad \text{and} \quad Y = AX'.$$

Use a direct method to evaluate

$$\frac{\partial \operatorname{vec} Y}{\partial \operatorname{vec} X}$$

and verify (5.10).

(2) Obtain

$$\frac{\partial \operatorname{vec} Y}{\partial \operatorname{vec} X},$$

when

$$(i) Y = AX'B \quad \text{and} \quad (ii) Y = \frac{1}{2} X^2.$$

(3) Find expressions for

$$\frac{\partial \operatorname{tr} Y}{\partial X},$$

when

$$(a) Y = AXB, \quad (b) Y = X^2 \quad \text{and} \quad (c) Y = XX'.$$

(4) Evaluate

$$\frac{\partial \operatorname{tr} Y}{\partial X}$$

when

$$(a) Y = X^{-1}, \quad (b) Y = AX^{-1}B, \quad (c) Y = X^n \quad \text{and} \quad (d) Y = e^X.$$

(5) (a) Use the direct method to obtain expressions for the matrix differential dY when

$$(i) Y = AX, \quad (ii) Y = X'X \quad \text{and} \quad (iii) Y = X^2.$$

(b) Find dY when

$$Y = AXBX.$$

CHAPTER 6

The Derivative of a Matrix with respect to a Matrix

6.1 INTRODUCTION

In the previous two chapters we have defined the derivative of a matrix with respect to a scalar and the derivative of a scalar with respect to a matrix. We will now generalise the definitions to include the derivative of a matrix with respect to a matrix. The author has adopted the definition suggested by Vetter [31], although other definitions also give rise to some useful results.

6.2 THE DEFINITIONS AND SOME RESULTS

Let $Y = [y_{ij}]$ be a matrix of order $(p \times q)$. We have defined (see (4.19)) the derivative of Y with respect to a scalar x_{rs} , it is the matrix $[\partial y_{ij}/\partial x_{rs}]$ of order $(p \times q)$.

Let $X = [x_{rs}]$ be a matrix of order $(m \times n)$ we generalise (4.19) and define the derivative of Y with respect to X , denoted by

$$\frac{\partial Y}{\partial X}$$

as the partitioned matrix whose (r, s) th partition is

$$\frac{\partial Y}{\partial x_{rs}}$$

in other words

$$\frac{\partial Y}{\partial X} = \begin{bmatrix} \frac{\partial Y}{\partial x_{11}} & \frac{\partial Y}{\partial x_{12}} & \dots & \frac{\partial Y}{\partial x_{1n}} \\ \frac{\partial Y}{\partial x_{21}} & \frac{\partial Y}{\partial x_{22}} & \dots & \frac{\partial Y}{\partial x_{2n}} \\ \vdots & & & \\ \frac{\partial Y}{\partial x_{m1}} & \frac{\partial Y}{\partial x_{m2}} & \dots & \frac{\partial Y}{\partial x_{mn}} \end{bmatrix} = \sum_{r,s} E_{rs} \otimes \frac{\partial Y}{\partial x_{rs}} \quad (6.1)$$

The right hand side of (6.1) following from the definitions (1.4) and (2.1) where E_{rs} is of order $(m \times n)$, the order of the matrix X .

It is seen that $\partial Y / \partial X$ is a matrix of order $(mp \times nq)$.

Example 6.1

Consider

$$Y = \begin{bmatrix} x_{11} x_{12} x_{22} & e^{x_{11} x_{22}} \\ \sin(x_{11} + x_{12}) & \log(x_{11} + x_{21}) \end{bmatrix}$$

and

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix},$$

Evaluate

$$\frac{\partial Y}{\partial X}.$$

Solution

$$\frac{\partial Y}{\partial x_{11}} = \begin{bmatrix} x_{12} x_{22} & x_{22} e^{x_{11} x_{22}} \\ \cos(x_{11} + x_{12}) & \frac{1}{(x_{11} + x_{21})} \end{bmatrix},$$

$$\frac{\partial Y}{\partial x_{12}} = \begin{bmatrix} x_{11} x_{22} & 0 \\ \cos(x_{11} + x_{12}) & 0 \end{bmatrix},$$

$$\frac{\partial Y}{\partial x_{21}} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{x_{11} + x_{21}} \end{bmatrix}, \quad \frac{\partial Y}{\partial x_{22}} = \begin{bmatrix} x_{11} x_{12} & x_{11} e^{x_{11} x_{22}} \\ 0 & 0 \end{bmatrix}.$$

$$\frac{\partial Y}{\partial X} = \begin{bmatrix} x_{12} x_{22} & x_{22} e^{x_{11} x_{22}} & x_{11} x_{22} & 0 \\ \cos(x_{11} + x_{12}) & \frac{1}{x_{11} + x_{21}} & \cos(x_{11} + x_{12}) & 0 \\ 0 & 0 & x_{11} x_{12} & x_{11} e^{x_{11} x_{22}} \\ 0 & \frac{1}{x_{11} + x_{21}} & 0 & 0 \end{bmatrix}$$

Example 6.2

Given the matrix $X = [x_{ij}]$ of order $(m \times n)$, evaluate $\partial X / \partial X$ when

- (i) All elements of X are independent
- (ii) X is a symmetric matrix (of course in this case $m = n$).

Solution

(i) By (6.1)

$$\frac{\partial X}{\partial X} = \sum_{r,s} E_{rs} \otimes E_{rs} = \bar{U} \quad (\text{see (2.26)})$$

(ii)

$$\frac{\partial X}{\partial x_{rs}} = E_{rs} + E_{sr} \quad \text{for } r \neq s$$

$$\frac{\partial X}{\partial x_{rr}} = E_{rr} \quad \text{for } r = s$$

We can write the above as;

$$\frac{\partial X}{\partial x_{rs}} = E_{rs} + E_{sr} - \delta_{rs} E_{rr}.$$

Hence,

$$\begin{aligned} \frac{\partial X}{\partial X} &= \sum_{r,s} E_{rs} \otimes E_{rs} + \sum_{r,s} E_{rs} \otimes E_{sr} - \delta_{rs} \sum_{r,s} E_{sr} \otimes E_{rr} \\ &= \bar{U} + U - \sum E_{rr} \otimes E_{rr} \quad (\text{see (2.24) and (2.26)}) \end{aligned}$$

Example 6.3

Evaluate and write out in full $\partial X' / \partial X$ given

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}.$$

Solution

By (6.1) we have

$$\begin{aligned} \frac{\partial X'}{\partial X} &= E_{rs} \otimes E'_{rs} \\ &= U. \end{aligned}$$

Hence

$$\frac{\partial X'}{\partial X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

From the definition (6.1) we obtain

$$\begin{aligned} \left(\frac{\partial Y}{\partial X} \right)' &= \left(\sum_{r,s} E_{rs} \otimes \frac{\partial Y}{\partial x_{rs}} \right)' \\ &= \sum_{r,s} E'_{rs} \otimes \left(\frac{\partial Y}{\partial x_{rs}} \right)' \quad \text{by (2.10)} \\ &= \sum_{r,s} E'_{rs} \otimes \frac{\partial Y'}{\partial x_{rs}} \quad \text{from (4.19).} \end{aligned}$$

It follows that

$$\left(\frac{\partial Y}{\partial X} \right)' = \frac{\partial Y'}{\partial X'}. \quad (6.2)$$

6.3 PRODUCT RULES FOR MATRICES

We shall first obtain a rule for the derivative of a product of matrices with respect to a matrix, that is to find an expression for

$$\frac{\partial(XY)}{\partial Z}$$

where the order of the matrices are as indicated

$$X(m \times n), \quad Y(n \times v), \quad Z(p \times q).$$

By (4.24) we write

$$\frac{\partial(XY)}{\partial z_{rs}} = \frac{\partial X}{\partial z_{rs}} Y + X \frac{\partial Y}{\partial z_{rs}}$$

where $Z = [z_{rs}]$.

If E_{rs} is an elementary matrix of order $(p \times q)$, we make use of (6.1) to write

$$\begin{aligned} \frac{\partial(XY)}{\partial Z} &= \sum_{r,s} E_{rs} \otimes \left[\frac{\partial X}{\partial z_{rs}} Y + X \frac{\partial Y}{\partial z_{rs}} \right] \\ &= \sum_{r,s} E_{rs} \otimes \frac{\partial X}{\partial z_{rs}} Y + \sum_{r,s} E_{rs} \otimes X \frac{\partial Y}{\partial z_{rs}} \\ &= \sum_{r,s} E_{rs} I_q \otimes \frac{\partial X}{\partial z_{rs}} Y + \sum_{r,s} I_p E_{rs} \otimes X \frac{\partial Y}{\partial z_{rs}} \end{aligned}$$

(where I_q and I_p are unit matrices of order $(q \times q)$ and $(p \times p)$ respectively)

$$= \sum_{r,s} \left(E_{rs} \otimes \frac{\partial X}{\partial z_{rs}} \right) (I_q \otimes Y) + \sum_{r,s} (I_p \otimes X) \left(E_{rs} \otimes \frac{\partial Y}{\partial z_{rs}} \right) \quad (\text{by 2.11})$$

finally, by (6.1)

$$\frac{\partial (XY)}{\partial Z} = \frac{\partial X}{\partial Z} (I_q \otimes Y) + (I_p \otimes X) \frac{\partial Y}{\partial Z} \quad (6.3)$$

Example 6.4

Find an expression for

$$\frac{\partial X^{-1}}{\partial X}.$$

Solution

Using (6.3) on

$$X X^{-1} = I,$$

we obtain

$$\frac{\partial (X X^{-1})}{\partial X} = \frac{\partial X}{\partial X} (I \otimes X^{-1}) + (I \otimes X) \frac{\partial X^{-1}}{\partial X} = 0$$

hence

$$\begin{aligned} \frac{\partial X^{-1}}{\partial X} &= -(I \otimes X)^{-1} \frac{\partial X}{\partial X} (I \otimes X^{-1}) \\ &= -(I \otimes X^{-1}) \bar{U} (I \otimes X^{-1}) \end{aligned}$$

(by Example 6.2 and (2.12)).

Next we determine a rule for the derivative of a Kronecker product of matrices with respect to a matrix, that is an expression for

$$\frac{\partial (X \otimes Y)}{\partial Z}$$

The order of the matrix Y is not now restricted, we will consider that it is $(u \times v)$. On representing $X \otimes Y$ by its (i,j) th partition $[x_{ij} Y]$ ($i = 1, 2, \dots, m$, $k = 1, 2, \dots, n$), we can write

$$\frac{\partial (X \otimes Y)}{\partial z_{rs}} = \frac{\partial}{\partial z_{rs}} [x_{ij} Y]$$

where (r, s) are fixed

$$\begin{aligned} &= \left[\frac{\partial x_{ij}}{\partial z_{rs}} Y \right] + \left[x_{ij} \frac{\partial Y}{\partial z_{rs}} \right] \\ &= \frac{\partial X}{\partial z_{rs}} \otimes Y + X \otimes \frac{\partial Y}{\partial z_{rs}}. \end{aligned}$$

Hence by (6.1)

$$\frac{\partial(X \otimes Y)}{\partial Z} = \sum_{r,s} E_{rs} \otimes \frac{\partial X}{\partial z_{rs}} \otimes Y + \sum_{r,s} E_{rs} \otimes X \otimes \frac{\partial Y}{\partial z_{rs}}$$

where E_{rs} is of order $(p \times q)$

$$= \frac{\partial X}{\partial Z} \otimes Y + \sum_{r,s} E_{rs} \otimes \left(X \otimes \frac{\partial Y}{\partial z_{rs}} \right).$$

The summation on the right hand side is not $X \otimes \partial Y / \partial Z$ as may appear at first sight, nevertheless it can be put into a more convenient form, as a product of matrices. To achieve this aim we make repeated use of (2.8) and (2.11)

$$\begin{aligned} \sum_{r,s} E_{rs} \otimes \left(X \otimes \frac{\partial Y}{\partial z_{rs}} \right) &= \sum_{r,s} [I_p E_{rs} I_q] \otimes \left[U_1 \left(\frac{\partial Y}{\partial z_{rs}} \otimes X \right) U_2 \right] \\ &\quad \text{by (2.14)} \\ &= \sum_{r,s} \left[(I_p E_{rs}) \otimes \left(U_1 \frac{\partial Y}{\partial z_{rs}} \otimes X \right) \right] [I_q \otimes U_2] \quad \text{by (2.11)} \\ &= \sum_{r,s} [I_p \otimes U_1] \left[E_{rs} \otimes \frac{\partial Y}{\partial z_{rs}} \otimes X \right] [I_q \otimes U_2] \quad \text{by (2.11)}. \end{aligned}$$

Hence

$$\frac{\partial(X \otimes Y)}{\partial Z} = \frac{\partial X}{\partial Z} \otimes Y + [I_p \otimes U_1] \left[\frac{\partial Y}{\partial Z} \otimes X \right] [I_q \otimes U_2] \quad (6.4)$$

where U_1 and U_2 are permutation matrices of orders $(mu \times mu)$ and $(nv \times nv)$ respectively.

We illustrate the use of equation (6.4) with a simple example.

Example 6.5

$A = [a_{ij}]$ and $X = [x_{ij}]$ are matrices, each of order (2×2) . Use

- (i) Equation (6.4), and
- (ii) a direct method to evaluate

$$\frac{\partial(A \otimes X)}{\partial X}.$$

Solution

(i) In this example (6.4) becomes

$$\frac{\partial(A \otimes X)}{\partial X} = [I \otimes U_1] \left[\frac{\partial X}{\partial X} \otimes A \right] [I \otimes U_2]$$

where I is the unit matrix of order (2×2) and

$$U_1 = U_2 = \sum E_{rs} \otimes E'_{rs} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since

$$\frac{\partial X}{\partial X} = \bar{U} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

only a simple calculation is necessary to obtain the result. It is found that

$$\frac{\partial(A \otimes X)}{\partial X} = \begin{bmatrix} a_{11} & 0 & a_{12} & 0 & 0 & a_{11} & 0 & a_{12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & a_{22} & 0 & 0 & a_{21} & 0 & a_{22} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{11} & 0 & a_{12} & 0 & 0 & a_{11} & 0 & a_{12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & a_{22} & 0 & 0 & a_{21} & 0 & a_{22} \end{bmatrix}.$$

(ii) We evaluate

$$Y = A \otimes X = \begin{bmatrix} a_{11}x_{11} & a_{11}x_{12} & a_{12}x_{11} & a_{12}x_{12} \\ a_{11}x_{21} & a_{11}x_{22} & a_{12}x_{21} & a_{12}x_{22} \\ a_{21}x_{11} & a_{21}x_{12} & a_{22}x_{11} & a_{22}x_{12} \\ a_{21}x_{21} & a_{21}x_{22} & a_{22}x_{21} & a_{22}x_{22} \end{bmatrix}$$

and then make use of (6.1) to obtain the above result.

6.4 THE CHAIN RULE FOR THE DERIVATIVE OF A MATRIX WITH RESPECT TO A MATRIX

We wish to obtain an expression for

$$\frac{\partial Z}{\partial X}$$

where the matrix Z is a matrix function of a matrix X , that is

$$Z = Y(X)$$

where

$$X = [x_{ij}] \text{ is of order } (m \times n)$$

$$Y = [y_{ij}] \text{ is of order } (u \times v)$$

$$Z = [z_{ij}] \text{ is of order } (p \times q)$$

By definition in (6.1)

$$\frac{\partial Z}{\partial X} = \sum_{r,s} E_{rs} \otimes \frac{\partial Z}{\partial x_{rs}} \quad \begin{matrix} r = 1, 2, \dots, m \\ s = 1, 2, \dots, n \end{matrix}$$

where E_{rs} is an elementary matrix of order $(m \times n)$,

$$= \sum_{r,s} E_{rs} \otimes \sum_{i,j} E_{ij} \frac{\partial z_{ij}}{\partial x_{rs}} \quad \begin{matrix} i = 1, 2, \dots, u \\ j = 1, 2, \dots, q \end{matrix}$$

where E_{ij} is of order $(p \times q)$

As in section 4.3, we use the chain rule to write

$$\frac{\partial z_{ij}}{\partial x_{rs}} = \sum_{\alpha, \beta} \frac{\partial z_{ij}}{\partial y_{\alpha\beta}} \cdot \frac{\partial y_{\alpha\beta}}{\partial x_{rs}} \quad \begin{matrix} \alpha = 1, 2, \dots, u \\ \beta = 1, 2, \dots, v \end{matrix}$$

Hence

$$\begin{aligned} \frac{\partial Z}{\partial X} &= \sum_{r,s} E_{rs} \otimes \left[\sum_{i,j} E_{ij} \sum_{\alpha, \beta} \frac{\partial z_{ij}}{\partial y_{\alpha\beta}} \cdot \frac{\partial y_{\alpha\beta}}{\partial x_{rs}} \right] \\ &= \sum_{r,s} \sum_{\alpha, \beta} E_{rs} \frac{\partial y_{\alpha\beta}}{\partial x_{rs}} \otimes \sum_{i,j} E_{ij} \frac{\partial z_{ij}}{\partial y_{\alpha\beta}} \quad (\text{by 2.5}) \\ &= \sum_{\alpha, \beta} \frac{\partial y_{\alpha\beta}}{\partial X} \otimes \frac{\partial Z}{\partial y_{\alpha\beta}} \quad (\text{by (4.7) and (4.19)}) . \end{aligned}$$

If I_n and I_p are unit matrices of orders $(n \times n)$ and $(p \times p)$ respectively, we can write the above as

$$\frac{\partial Z}{\partial X} = \sum_{\alpha, \beta} \left(\frac{\partial y_{\alpha\beta}}{\partial X} I_n \right) \otimes \left(I_p \frac{\partial Z}{\partial y_{\alpha\beta}} \right).$$

Hence, by (2.11)

$$\frac{\partial Z}{\partial X} = \sum_{\alpha, \beta} \left(\frac{\partial y_{\alpha\beta}}{\partial X} \otimes I_p \right) \left(I_n \otimes \frac{\partial Z}{\partial y_{\alpha\beta}} \right) \quad (6.5)$$

Equation (6.5) can be written in a more convenient form, avoiding the summation, if we define an appropriate notation, a generalisation of the previous one.

Since

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1v} \\ y_{21} & y_{22} & \dots & y_{2v} \\ \vdots & \vdots & \ddots & \vdots \\ y_{u1} & y_{u2} & \dots & y_{uv} \end{bmatrix}$$

than $(\text{vec } Y)' = [y_{11} \ y_{21} \ \dots \ y_{uv}]$.

We will write the partitioned matrix

$$\left[\frac{\partial y_{11}}{\partial X} \otimes I_p \ : \ \frac{\partial y_{21}}{\partial X} \otimes I_p \ : \ \dots \ : \ \frac{\partial y_{uv}}{\partial X} \otimes I_p \right]$$

as

$$\frac{\partial [y_{11} \ y_{21} \ \dots \ y_{uv}]}{\partial X} \otimes I_p$$

or as

$$\frac{\partial (\text{vec } Y)'}{\partial X} \otimes I_p.$$

Similarly, we write the partitioned matrix

$$\begin{bmatrix} I_n \otimes \frac{\partial Z}{\partial y_{11}} \\ \hline I_n \otimes \frac{\partial Z}{\partial y_{21}} \\ \hline \vdots \\ \hline I_n \otimes \frac{\partial Z}{\partial y_{uv}} \end{bmatrix} \quad \text{as} \quad \begin{bmatrix} I_n \otimes \frac{\partial Z}{\partial \text{vec } Y} \end{bmatrix}$$

We can write the sum (6.5) in the following order

$$\begin{aligned} \frac{\partial Z}{\partial X} = & \left[\frac{\partial y_{11}}{\partial X} \otimes I_p \right] \left[I_n \otimes \frac{\partial Z}{\partial y_{11}} \right] + \left[\frac{\partial y_{21}}{\partial X} \otimes I_p \right] \left[I_n \otimes \frac{\partial Z}{\partial y_{21}} \right] + \dots \\ & + \left[\frac{\partial y_{uv}}{\partial X} \otimes I_p \right] \left[I_n \otimes \frac{\partial Z}{\partial y_{uv}} \right]. \end{aligned}$$

We can write this as a (partitioned) matrix product

$$\frac{\partial Z}{\partial X} = \left[\frac{\partial y_{11}}{\partial X} \otimes I_p : \frac{\partial y_{21}}{\partial X} \otimes I_p : \dots : \frac{\partial y_{uv}}{\partial X} \otimes I_p \right] \begin{bmatrix} I_n \otimes \frac{\partial Z}{\partial y_{11}} \\ \hline I_n \otimes \frac{\partial Z}{\partial y_{21}} \\ \hline \vdots \\ \hline I_n \otimes \frac{\partial Z}{\partial y_{uv}} \end{bmatrix}.$$

Finally, using the notations defined above, we have

$$\frac{\partial Z}{\partial X} = \left[\frac{\partial [\text{vec } Y]'}{\partial Z} \otimes I_p \right] \left[I_n \otimes \frac{\partial Z}{\partial \text{vec } Y} \right] \quad (6.6)$$

We consider a simple example to illustrate the application of the above formula. The example can also be solved by evaluating the matrix Z in terms of the components of the matrix X and then applying the definition in (6.1).

Example 6.6

Given the matrix $A = [a_{ij}]$ and $X = [x_{ij}]$ both of order (2×2) , evaluate

$$\partial Z / \partial X$$

where $Z = Y'Y$ and $Y = AX$.

- (i) Using (6.6)
- (ii) Using a direct method.

Solution

(i) For convenience write (6.6) as

$$\frac{\partial Z}{\partial X} = QR$$

where

$$Q = \left[\frac{\partial [\text{vec } Y]'}{\partial X} \otimes I_p \right] \quad \text{and} \quad R = \left[I_n \otimes \frac{\partial Z}{\partial \text{vec } Y} \right].$$

The product of Q and R is the derivative we have been asked to evaluate

$$QR = \begin{bmatrix} 2a_{11}y_{11} + 2a_{21}y_{21} & a_{11}y_{12} + a_{21}y_{22} & 0 & a_{11}y_{11} + a_{21}y_{21} \\ a_{11}y_{12} + a_{21}y_{22} & 0 & a_{11}y_{11} + a_{21}y_{21} & 2a_{11}y_{12} + 2a_{21}y_{22} \\ 2a_{12}y_{11} + 2a_{22}y_{21} & a_{12}y_{12} + a_{22}y_{22} & 0 & a_{12}y_{11} + a_{22}y_{21} \\ a_{12}y_{12} + a_{22}y_{22} & 0 & a_{12}y_{11} + a_{22}y_{21} & 2a_{12}y_{12} + 2a_{22}y_{22} \end{bmatrix}$$

(ii) By a simple extension of the result of Example 4.6(b) we find that when

$$Z = X'A'AX$$

$$\frac{\partial Z}{\partial x_{rs}} = E'_{rs}A'AX + X'A'AE_{rs}$$

$$= E'_{rs}A'Y + Y'AE_{rs}$$

where $Y = AX$.

By (6.1) and (2.11)

$$\frac{\partial Z}{\partial X} = \sum_{r,s} (E_{rs} \otimes E'_{rs}) (I \otimes A'Y) + \sum_{r,s} (I \otimes Y'Z) (E_{rs} \otimes E_{rs}).$$

Since the matrices involved are all of order (2×2)

$$\sum E_{rs} \otimes E'_{rs} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\sum E_{rs} \otimes E_{rs} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

On substitution and multiplying out in the above expression for $\partial Z/\partial X$, we obtain the same matrix as in (i).

Problems for Chapter 6

(1) Evaluate $\partial Y/\partial X$ given

$$Y = \begin{bmatrix} \cos(x_{12} + x_{22}) & x_{11}x_{21} \\ e^{x_{11}x_{12}} & x_{12}x_{22} \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

(2)

The elements of the matrix $X = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix}$

are all independent. Use a direct method to evaluate $\partial X / \partial X$.

(3)

Given a non-singular matrix $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$

use a direct method to obtain

$$\frac{\partial X^{-1}}{\partial X}$$

and verify the solution to Example 6.4.

(4) The matrices $A = [a_{ij}]$ and $X = [x_{ij}]$ are both of order (2×2) , X is non-singular. Use a direct method to evaluate

$$\frac{\partial(A \otimes X^{-1})}{\partial X}$$

CHAPTER 7

Some Applications of Matrix Calculus

7.1 INTRODUCTION

As in Chapter 3, where a number of applications of the Kronecker product were considered, in this chapter a number of applications of matrix calculus are discussed. The applications have been selected from a number considered in the published literature, as indicated in the Bibliography at the end of this book.

These problems were originally intended for the expert, but by expansion and simplification it is hoped that they will now be appreciated by the general reader.

7.2 THE PROBLEMS OF LEAST SQUARES AND CONSTRAINED OPTIMISATION IN SCALAR VARIABLES

In this section we consider, very briefly, the Method of Least Squares to obtain a curve or a line of 'best fit', and the Method of Lagrange Multipliers to obtain an extremum of a function subject to constraints.

For the least squares method we consider a set of data

$$(x_i, y_i) \quad i = 1, 2, \dots, n \quad (7.1)$$

and a relationship, usually a polynomial function

$$y = f(x). \quad (7.2)$$

For each x_i , we evaluate $f(x_i)$ and the residual or the deviation

$$e_i = y_i - f(x_i). \quad (7.3)$$

The method depends on choosing the unknown parameters, the polynomial coefficients when $f(x)$ is a polynomial, so that the sum of the squares of the residuals is a minimum, that is

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (7.4)$$

is a minimum.

In particular, when $f(x)$ is a linear function

$$y = a_0 + a_1 x$$

$S(a_0, a_1)$ is a minimum when

$$\frac{\partial S}{\partial a_0} = 0 = \frac{\partial S}{\partial a_1}. \quad (7.5)$$

These two equations, known as normal equations, determine the two unknown parameters a_0 and a_1 which specify the line of 'best fit' according to the principle of least squares.

For the second method we wish to determine the extremum of a continuously differentiable function

$$f(x_1, x_2, \dots, x_n) \quad (7.6)$$

whose n variables are constrained by m equations of the form

$$g_i(x_1, x_2, \dots, x_n) = 0, \quad i = 1, 2, \dots, m \quad (7.7)$$

The method of Lagrange Multipliers depends on defining an augmented function

$$f^* = f + \sum_{i=1}^m \mu_i g_i \quad (7.8)$$

where the μ_i are known as Lagrange multipliers.

The extreme of $f(x)$ is determined by solving the system of the $(m + n)$ equations

$$\begin{aligned} \frac{\partial f^*}{\partial x_r} &= 0 & r &= 1, 2, \dots, n \\ g_i &= 0 & i &= 1, 2, \dots, m \end{aligned} \quad (7.9)$$

for the m parameters $\mu_1, \mu_2, \dots, \mu_m$ and the n variables x determining the extremum.

Example 7.1

Given a matrix $A = [a_{ij}]$ of order (2×2) determine a symmetric matrix $X = [x_{ij}]$ which is a best approximation to A by the criterion of least squares.

Solution

Corresponding to (7.3) we have

$$E = A - X$$

where $E = [e_{ij}]$ and $e_{ij} = a_{ij} - x_{ij}$.

The criterion of least squares for this example is to minimise

$$S = \sum_{i,j} e_{ij}^2 = \sum (a_{ij} - x_{ij})^2$$

which is the equivalent of (7.6) above.

The constraint equation is

$$x_{12} - x_{21} = 0$$

and the augmented function is

$$f^* = \sum (a_{ij} - x_{ij})^2 + \mu(x_{12} - x_{21}) = 0$$

$$\frac{\partial f^*}{\partial x_{11}} = -2(a_{11} - x_{11}) = 0$$

$$\frac{\partial f^*}{\partial x_{12}} = -2(a_{12} - x_{12}) + \mu = 0$$

$$\frac{\partial f^*}{\partial x_{21}} = -2(a_{21} - x_{21}) - \mu = 0$$

$$\frac{\partial f^*}{\partial x_{22}} = -2(a_{22} - x_{22}) = 0$$

This system of 5 equations (including the constraint) leads to the solution

$$\mu = a_{12} - x_{21}$$

$$x_{11} = a_{11}, \quad x_{22} = a_{22}, \quad x_{12} = x_{21} = \frac{1}{2}(a_{12} + a_{21}).$$

Hence

$$\begin{aligned} X &= \begin{bmatrix} a_{11} & \frac{a_{12} + a_{21}}{2} \\ \frac{a_{12} + a_{21}}{2} & a_{22} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \frac{1}{2} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \\ &= \frac{1}{2}(A + A') \end{aligned}$$

7.3 PROBLEM 1 – MATRIX CALCULUS APPROACH TO THE PROBLEMS OF LEAST SQUARES AND CONSTRAINED OPTIMISATION

If we can express the residuals in the form of a matrix E , as in Example 7.1, then the sum of the residuals squared is

$$S = \text{tr } E'E \quad (7.10)$$

The criterion of the least squares method is to minimise (7.10) with respect to the parameters involved.

The constrained optimisation problem then takes the form of finding the matrix X such that the scalar matrix function

$$S = f(X)$$

is minimised subject to constraints on X in the form of

$$G(X) = 0 \quad (7.11)$$

where $G = [g_{ij}]$ is a matrix of order $(s \times t)$ where s and t are dependent on the number of constraints g_{ij} involved.

As for the scalar case, we use Lagrange multipliers to form an augmented matrix function $f^*(X)$.

Each constraint g_{ij} is associated with a parameter (Lagrange multiplier) μ_{ij} .

Since
$$\sum \mu_{ij} g_{ij} = \text{tr } U'G$$

where

$$U = [\mu_{ij}]$$

we can write the augmented scalar matrix function as

$$f^*(X) = \text{tr } E'X + \text{tr } U'G \quad (7.12)$$

which is the equivalent to (7.8). To find the optimal X , we must solve the system of equations

$$\frac{\partial f^*}{\partial X} = 0. \quad (7.13)$$

Problem

Given a non-singular matrix $A = [a_{ij}]$ of order $(n \times n)$ determine a matrix $X = [x_{ij}]$ which is a least squares approximation to A

- (i) when X is a symmetric matrix
- (ii) when X is an orthogonal matrix.

Solution

(i) The problem was solved in Example 7.1 when A and X are of order (2×2) . With the terminology defined above, we write

$$E = A - X$$

$$G(X) = X - X' = 0$$

so that G and hence U are both of order $(n \times n)$.

Equation (7.12) becomes

$$\begin{aligned} f^* &= \operatorname{tr} [A' - X'] [A - X] + \operatorname{tr} U' [X - X'] \\ &= \operatorname{tr} A'A - \operatorname{tr} A'X - \operatorname{tr} X'A + \operatorname{tr} X'X + \operatorname{tr} U'X - \operatorname{tr} U'X'. \end{aligned}$$

We now make use of the results, in modified form if necessary, of Examples 5.4 and 5.5, we obtain

$$\begin{aligned} \frac{\partial f^*}{\partial X} &= -2A + 2X + U - U' \\ &= 0 \quad \text{for} \quad X = A + \frac{U - U'}{2}. \end{aligned}$$

Then

$$X' = A' + \frac{U' - U}{2}$$

and since $X = X'$, we finally obtain

$$X = \frac{1}{2}(A + A').$$

(ii) This time

$$G(X) = X'X - I = 0.$$

Hence

$$f^* = \operatorname{tr} [A' - X'] [A - X] + \operatorname{tr} U' [XX' - I]$$

so that

$$\begin{aligned} \frac{\partial f^*}{\partial X} &= -2A + 2X + X[U + U'] \\ &= 0 \quad \text{for} \quad X = A - X \left[\frac{U + U'}{2} \right]. \end{aligned}$$

Premultiplying by X' and using the condition

$$X'X = I$$

we obtain

$$X'A = I + \frac{U + U'}{2}$$

and on transposing

$$A'X = I + \frac{U + U'}{2}.$$

Hence

$$A'X = X'A. \tag{7.14}$$

If a solution to (7.14) exists, there are various ways of solving this matrix equation.

For example with the help of (2.13) and Example (2.7) we can write it as

$$[(I \otimes A') - (A' \otimes I)U]x = 0 \quad (7.15)$$

where U is a permutation matrix (see (2.24)) and

$$x = \text{vec } X.$$

We have now reduced the matrix equation into a system of homogeneous equations which can be solved by a standard method.

If a non-trivial solution to (7.15) does exist, it is not unique. We must scale it appropriately for X to be orthogonal.

There may, of course, be more than one linearly independent solution to (7.15). We must choose the solution corresponding to X being an orthogonal matrix.

Example 7.2

Given

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

find the orthogonal matrix X which is the least squares best approximation to A .

Solution

$$[I \otimes A'] = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & 1 \end{bmatrix} \quad \text{and} \quad [A' \otimes I]U = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}.$$

Equation (7.15) can now be written as

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 2 & 1 & -1 & 1 \\ -2 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} x = 0.$$

There are 3 non-trivial (linearly independent) solutions, (see [18] p. 131). They are

$$x = [1 \ -2 \ 1 \ 1]', \quad x = [1 \ 1 \ 2 \ -1]' \quad \text{and} \quad x = [2 \ -3 \ 3 \ 2]'$$

Only the last solution leads to an orthogonal matrix X , it is

$$X = \frac{1}{\sqrt{13}} \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}.$$

7.4 PROBLEM 2 – THE GENERAL LEAST SQUARES PROBLEM

The linear regression problem presents itself in the following form:

N samples from a population are considered. The i th sample consists of an observation from a variable Y and observations from variables X_1, X_2, \dots, X_n (say).

We assume a linear relationship between the variables. If the variables are measured from zero, the relationship is of the form

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in} + e_i. \quad (7.16)$$

If the observations are measured from their means over the N samples, then

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in} + e_i \quad (i = 1, 2, \dots, N) \quad (7.17)$$

$b_0, b_1, b_2, \dots, b_n$ are estimated parameters and e_i is the corresponding residual.

In matrix notation we can write the above equations as

$$y = Xb + e \quad (7.18)$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

and

$$X = \begin{bmatrix} 1 & x_{12} & \dots & x_{1n} \\ 1 & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N2} & \dots & x_{Nn} \end{bmatrix} \quad \text{or} \quad X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix}.$$

As already indicated, the 'goodness of fit' criterion is the minimisation with respect to the parameters b of the sum of the squares of the residuals, which in this case is

$$S = e'e = (y' - b'X')(y - Xb).$$

Making use of the results in table (4.4), we obtain

$$\begin{aligned} \frac{\partial(e'e)}{\partial b} &= -(y'X)' - X'y + (X'Xb + X'Xb) \\ &= -2X'y + 2X'Xb \\ &= 0 \quad \text{for} \quad X'X\hat{b} = X'y. \end{aligned} \quad (7.19)$$

where \hat{b} is the least squares estimate of b .

If $(X'X)$ is non-singular, we obtain from (7.19)

$$\hat{b} = (X'X)^{-1}X'y. \quad (7.20)$$

We can write (7.19) as

$$X'(y - X\hat{b}) = 0$$

or

$$X'e = 0 \quad (7.21)$$

which is the matrix form of the normal equations defined in section 7.2.

Example 7.3

Obtain the normal equations for a least squares approximation when each sample consists of one observation from Y and one observation from

- (i) a random variable X
- (ii) two random variables X and Z .

Solution

$$(i) \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix},$$

hence

$$X'[y - X\hat{b}] = \begin{bmatrix} \Sigma y_i - \hat{b}_1 N - \hat{b}_2 \Sigma x_i \\ \Sigma x_i y_i - \hat{b}_1 \Sigma x_i - \hat{b}_2 \Sigma x_i^2 \end{bmatrix}.$$

So that the normal equations are

$$\Sigma y_i = \hat{b}_1 N + \hat{b}_2 \Sigma x_i$$

and

$$\Sigma x_i y_i = \hat{b}_1 \Sigma x_i + \hat{b}_2 \Sigma x_i^2.$$

(ii) In this case

$$X = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & z_N \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix}.$$

The normal equations are

$$\Sigma y_i = \hat{b}_1 N + \hat{b}_2 \Sigma x_i + \hat{b}_3 \Sigma z_i$$

$$\Sigma x_i y_i = \hat{b}_1 \Sigma x_i + \hat{b}_2 \Sigma x_i^2 + \hat{b}_3 \Sigma x_i z_i$$

and

$$\Sigma x_i z_i = \hat{b}_1 \Sigma z_i + \hat{b}_2 \Sigma x_i z_i + \hat{b}_3 \Sigma z_i^2.$$

7.5 PROBLEM 3 – MAXIMUM LIKELIHOOD ESTIMATE OF THE MULTIVARIATE NORMAL

Let $X_i (i = 1, 2, \dots, n)$ be n random variables each having a normal distribution with mean μ_i and standard deviation σ_i , that is

$$X_i = n(\mu_i, \sigma_i). \quad (7.22)$$

The joint probability density function (p.d.f.) of the n random variables is

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} \exp \left(\frac{-\frac{1}{2}(x - \mu)' V^{-1} (x - \mu)}{2} \right) \quad (7.23)$$

where

$$-\infty < x_i < \infty \quad (i = 1, 2, \dots, n)$$

and

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_{nn} \end{bmatrix}$$

is the covariance matrix.

$$\mu' = [\mu_1, \mu_2, \dots, \mu_n], \quad x' = [x_1, x_2, \dots, x_n]$$

and

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j \quad (i \neq j)$$

$$\sigma_{ii} = \sigma_i^2$$

are the covariances of the random variables.

ρ_{ij} is the correlation coefficient between X_i and X_j . The covariance matrix V is symmetric and positive definite.

Equation (7.23) is called a **multivariate normal p.d.f.** Maximum likelihood estimates have certain properties (for example, they are asymptotically efficient) which makes them very useful in estimation and hypothesis testing problems.

For a sample of N observations from the multivariate normal distribution (7.23) the likelihood function is

$$L = \frac{1}{(2\pi)^{nN/2} |V|^{N/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (x_i - \mu)' V^{-1} (x_i - \mu) \right\}$$

so that

$$\log L = C - \frac{N}{2} \log |V| - \frac{1}{2} \sum_{i=1}^N (x_i - \mu)' V^{-1} (x_i - \mu) \quad (7.24)$$

where C is a constant.

(a) The maximum likelihood estimate of μ

On expanding the last term of (7.24), we obtain

$$-\frac{1}{2} \sum_{i=1}^N \{x_i' V^{-1} x_i - \mu' V^{-1} x_i - x_i' V^{-1} \mu + \mu' V^{-1} \mu\}.$$

With the help of table (4.4) and using the result

$$(x_i' V^{-1})' = V^{-1} x_i \quad (\text{since } V \text{ is symmetric})$$

we differentiate with respect to μ , to obtain

$$\begin{aligned} \frac{\partial \log L}{\partial \mu} &= V^{-1} \sum_{i=1}^N (x_i - \mu) \\ &= 0 \quad \text{when} \quad \hat{\mu} = \frac{\sum x_i}{N} = \bar{x}. \end{aligned}$$

Hence the maximum likelihood estimate of μ is $\mu = \bar{x}$, the sample mean.

(b) The maximum likelihood estimate of V .

We note the following results:

$$(1) \quad \sum_{i=1}^N y_i' V^{-1} y_i = \text{tr}(Y' V^{-1} Y) = \text{tr}(Y Y' V^{-1})$$

where $Y = [y_1 y_2 \dots y_N]$

and $y_i = x_i - \mu \quad (i = 1, 2, \dots, N).$

V^{-1} is a symmetric matrix.

(2) By Example 5.3, but taking account of the symmetry of V^{-1} (see Example 4.4)

$$\frac{\partial \log |V^{-1}|}{\partial V^{-1}} = 2V - \text{diag}\{V\}.$$

(3) If X is a symmetric matrix

$$\frac{\partial \text{tr}(AX)}{\partial X} = A + A' - \text{diag}\{A\}.$$

Let $A = YY'$ and $X = V^{-1}$, then

$$\frac{\partial \text{tr}(YY' V^{-1})}{\partial V^{-1}} = 2YY' - \text{diag}\{YY'\}.$$

We now write (7.24) as

$$\log L = C + \frac{N}{2} \log V^{-1} - \frac{1}{2} \text{tr}(YY'V^{-1}).$$

Differentiating $\log L$ with respect to V^{-1} , using the estimate $\mu = \bar{x}$, and the results (2) and (3) above, we obtain

$$\frac{\partial \log L}{\partial V^{-1}} = \frac{N}{2} [2V - \text{diag}\{V\}] - YY' + \frac{1}{2} \text{diag}\{YY'\}.$$

Let $Q = NV - YY'$, then

$$\begin{aligned} \frac{\partial \log L}{\partial V^{-1}} &= Q - \frac{1}{2} \text{diag}\{Q\} \\ &= 0 \quad \text{when} \quad 2Q = \text{diag}\{Q\}. \end{aligned}$$

Since Q is symmetric, the only solution to the above equation is

$$Q = 0.$$

It follows that the maximum likelihood estimate of V is

$$\hat{V} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})'}{N}.$$

7.6 PROBLEM 4 – EVALUATION OF THE JACOBIANS OF SOME TRANSFORMATIONS

The interest in Jacobians arises from their importance particularly with reference to a change of variables in multiple integration.

In terms of scalars, the problem presents itself in the following way.

We consider a multiple integral of a subset R of an n -dimensional space

$$\int_R f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots, dx_n. \quad (7.25)$$

where f is a piecewise continuous function in R .

We consider a one to one transformation which maps R onto a subset T

$$y_1 = \mu_1(x), \quad y_2 = \mu_2(x), \quad \dots, \quad y_n = \mu_n(x) \quad (7.26)$$

and the inverse transformation

$$x_1 = w_1(y), \quad x_2 = w_2(y), \quad \dots, \quad x_n = w_n(y) \quad (7.27)$$

where

$$x' = [x_1, x_2, \dots, x_n] \quad \text{and} \quad y' = [y_1, y_2, \dots, y_n].$$

Assuming the first partial derivations of the inverse transformation (7.27) to be continuous, (7.25) can be expressed as

$$\int_T f[w_1(y), w_2(y), \dots, w_n(y)] |J| dy_1 dy_2 \dots dy_n \quad (7.28)$$

where $|J|$ can be written as

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_1}{\partial y_n} & \frac{\partial x_2}{\partial y_n} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|, \quad (7.29)$$

subject to $|J|$ not vanishing identically in T .

Example 7.4

Let

$$I = 2 \int_R \exp \{-2x_1 + 3x_2\} dx_1 dx_2$$

$$0 < x_1 < \infty, \quad 0 < x_2 < \infty.$$

Consider the transformation

$$y_1 = 2x_1 - x_2$$

$$y_2 = x_2.$$

Write down the integral corresponding to (7.28).

Solution

We are given

$$R = \{(x_1, x_2): 0 < x_1 < \infty, 0 < x_2 < \infty\}.$$

The above transformation (corresponding to (7.26)) results in the following inverse transformation (7.27)

$$x_1 = \frac{1}{2}(y_1 + y_2)$$

$$x_2 = y_2$$

which defines

$$T = \{(y_1, y_2): y_2 > 0, y_2 > -y_1, -\infty < y_1 < \infty\},$$

and by (7.29)

$$|J| = \begin{vmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{vmatrix} = \frac{1}{2}.$$

Hence

$$\begin{aligned} I &= \int_T f\left[\frac{1}{2}(y_1 + y_2), y_2\right] dy_1 dy_2 \\ &= \int_T \exp(-y_1 + 2y_2) dy_1 dy_2. \end{aligned}$$

Our main interest in this section is to evaluate Jacobians when the transformation corresponding to (7.26) is expressed in matrix form, for example as

$$Y = AXB \quad (7.30)$$

where A , X and B are all assumed to be of order $(n \times n)$.

As in section 5.2 (see (5.1) and (5.2)) we can write (7.6) as

$$y = Px \quad (7.31)$$

where $y = \text{vec } Y$, $x = \text{vec } X$ and $P = B' \otimes A$.

In this case

$$\frac{\partial y}{\partial x} = B \otimes A'$$

and

$$\frac{\partial x}{\partial y} = [B \otimes A']^{-1} = B^{-1} \otimes (A')^{-1} \quad \text{by (2.12).}$$

It follows that

$$|J| = \left| \frac{\partial \text{vec } Y}{\partial \text{vec } X} \right|^{-1} = |B|^{-n} |A|^{-n} \quad (\text{by Property X, p. 27}) \quad (7.32)$$

Example 7.5

Consider the transformation

$$Y = AXB$$

where

$$A = \begin{bmatrix} 2 & -4 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Find the Jacobian of this transformation

- (i) By a direct method
- (ii) Using (7.32).

Solution

(i) We have

$$X = A^{-1}YB^{-1} = \frac{1}{2} \begin{bmatrix} 3y_1 + 4y_2 - 3y_3 - 4y_4 & -3y_1 - 4y_2 + 6y_3 + 8y_4 \\ y_1 + 2y_2 - y_3 - 2y_4 & -y_1 - 2y_2 + 2y_3 + 4y_4 \end{bmatrix}$$

so that

$$\frac{\partial x}{\partial y} = \left(\frac{1}{2}\right)^4 \begin{vmatrix} 3 & 1 & -3 & -1 \\ 4 & 2 & 4 & -2 \\ 0 & 0 & 3 & 1 \\ -4 & -2 & 8 & 4 \end{vmatrix} = \frac{1}{4}.$$

(ii) $|A| = 2$, $|B| = 1$ hence $|J| = \frac{1}{4}$.

Similarly, we can use the theory developed in this book to evaluate the Jacobians of many other transformations.

Example 7.6

Evaluate the Jacobian associated with the following transformation

(i) $Y = X^{-1}$

(ii) $Y = X^2$.

Solution

(i) From Example 5.2

$$\frac{\partial y}{\partial x} = -X^{-1} \otimes (X^{-1})'$$

so that

$$\frac{\partial x}{\partial y} = -X \otimes X'.$$

Hence

$$J = \text{mod} \left| \frac{\partial y}{\partial x} \right| = |X \otimes X'| = |X|^{-n} |X|^{-n} = |X|^{-2n}.$$

(ii) From section 4.6

$$\frac{\partial Y}{\partial x_{rs}} = E_{rs}X + E_{rs},$$

so that by the 2nd transformation principle (see section 5.3)

$$\frac{\partial y}{\partial x} = X \otimes I + I \otimes X'$$

and

$$J = |X \otimes I + I \otimes X'|^{-1}$$

7.7 PROBLEM 5 – TO FIND THE DERIVATIVE OF AN EXPONENTIAL MATRIX WITH RESPECT TO A MATRIX

Since we make use of the spectral decomposition of an exponential matrix, we now discuss this technique briefly.

Assume that the matrix $Q = [q_{ij}]$ of order $(n \times n)$ has eigenvalues

$$\lambda_1, \lambda_2, \dots, \lambda_n$$

(not necessarily distinct) and corresponding linearly independent eigenvectors

$$x_1, x_2, \dots, x_n.$$

The eigenvectors of Q' are

$$y_1, y_2, \dots, y_n.$$

These two sets of eigenvectors have the property

$$x'_i y_j = 0 \quad \text{or (equivalently)} \quad y'_i x_j = 0 \quad (i \neq j) \quad (7.33)$$

and can be normalised so that

$$x'_i y_i = 1 \quad \text{or} \quad y'_i x_i = 1 \quad (i = 1, 2, \dots, n). \quad (7.34)$$

Sets of eigenvectors $\{x_i\}$ and $\{y_i\}$ having the properties (7.33) and (7.34) are said to be properly normalised.

It is well known (see [18] p. 227) that

$$\exp(Qt) = P \operatorname{diag} \{e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}\} P^{-1}$$

where P is the modal matrix of Q , that is the matrix

$$P = [x_1 : x_2 : \dots : x_n]. \quad (7.35)$$

It follows from (7.33), (7.34) and (7.35) that

$$P^{-1} = [y_1, y_2, \dots, y_n]' = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix}. \quad (7.36)$$

Hence

$$\begin{aligned} \exp(Qt) &= [x_1 \ x_2 \ \dots \ x_n] \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix} \\ &= x_1 y'_1 \exp(\lambda_1 t) + x_2 y'_2 \exp(\lambda_2 t) + \dots + x_n y'_n \exp(\lambda_n t), \end{aligned}$$

that is

$$\exp(Qt) = \sum_{i=1}^n x_i y_i' \exp(\lambda_i t) . \quad (7.37)$$

The right hand side of (7.37) is known as the **spectral representation** (or **spectral decomposition**) of the exponential matrix $\exp(Qt)$.

We consider a very simple illustrative example.

Example 7.7

Find the spectral representation of the matrix $\exp(Qt)$, where

$$Q = \begin{bmatrix} 3 & 4 \\ -2 & -3 \end{bmatrix} .$$

Solution

$$\begin{aligned} \lambda_1 = 1, \lambda_2 = -1; \quad x_1' &= [2 \ -1], \quad x_2' = [1 \ -1] \\ y_1' &= [1 \ 1], \quad y_2' = [-1 \ -2] . \end{aligned}$$

By (7.37)

$$\begin{aligned} \exp(Qt) &= \begin{bmatrix} 2 \\ -1 \end{bmatrix} [1 \ 1] \exp(t) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [-1 \ -2] \exp(-t) \\ &= \begin{bmatrix} 2 & 2 \\ -1 & -1 \end{bmatrix} \exp(t) + \begin{bmatrix} -1 & -2 \\ 1 & 3 \end{bmatrix} \exp(-t) . \end{aligned}$$

Although we have considered matrices having real eigenvalues, and eigenvectors having real elements, the spectral decomposition (7.37) is also valid for complex elements as can be shown by a slight modification of the above exposition.

By the use of (2.17), that is of the result

$$\exp(I \otimes Q) = I \otimes \exp(Q)$$

we generalise the result (7.37) to

$$\exp(I \otimes Q)t = \sum (I \otimes x_i y_i') \exp(\lambda_i t) . \quad (7.38)$$

We now consider the main problem, to obtain an expression for

$$\partial \Phi / \partial Z$$

where

$$\Phi(t) = \exp(Qt) , \quad (7.39)$$

so that

$$\Phi(0) = I , \quad (7.40)$$

$$\frac{d}{dt} \Phi = Q \Phi , \quad (7.41)$$

and

$$Z = [z_{ij}] \text{ is a matrix of order } (r \times s).$$

The matrix Q is assumed to be a function of Z , that is $Q(Z)$. Making use of the result (6.5), we can write

$$\frac{d}{dt} \frac{\partial \Phi}{\partial Z} = \frac{\partial(Q\Phi)}{\partial Z} = \frac{\partial Q}{\partial Z} (I \otimes \Phi) + (I \otimes Q) \frac{\partial \Phi}{\partial Z} \quad (7.42)$$

and from (7.40)

$$\frac{\partial \Phi}{\partial Z}(0) = 0. \quad (7.43)$$

We next make use of a generalisation of a well known result (see [19] p. 68);
Given

$$\frac{d}{dt} X = RX + BU$$

and

$$X(0) = 0,$$

then

$$X = \int_0^t \exp \{R(t-\tau)\} BU(\tau) d\tau. \quad (7.44)$$

For

$$X = \frac{\partial \Phi}{\partial Z}, \quad R = I \otimes Q, \quad B = \frac{\partial Q}{\partial Z} \quad \text{and} \quad U = I \otimes \Phi$$

the solution to (7.42) subject to (7.43) becomes

$$\begin{aligned} \frac{\partial \Phi}{\partial Z} &= \int_0^t \exp \{I \otimes Q(t-\tau)\} \frac{\partial Q}{\partial Z} [I \otimes \Phi(\tau)] d\tau \\ &= \int_0^t \sum_{i,j} (I \otimes x_i y_i') \exp(\lambda_i(t-\tau)) \frac{\partial Q}{\partial Z} [I \otimes x_j y_j'] \exp(\lambda_j \tau) d\tau \\ &\quad \text{(by 7.37 and 7.38)} \\ &= \sum_{i,j} (I \otimes x_i y_i') \frac{\partial Q}{\partial Z} (I \otimes x_j y_j') \exp(\lambda_i t) \int_0^t \exp((\lambda_j - \lambda_i)\tau) d\tau. \end{aligned}$$

Hence,

$$\frac{\partial \Phi}{\partial Z} = \sum_{i,j} I \otimes x_i y_i' \frac{\partial Q}{\partial Z} (I \otimes x_j y_j') \exp(\lambda_i t) f_{ij}(t) \quad (7.45)$$

where

$$f_{ij}(t) = t \quad \text{if} \quad \lambda_i = \lambda_j$$

and

$$f_{ij}(t) = (1/(\lambda_j - \lambda_i)) [\exp(\lambda_j - \lambda_i)t - 1] \quad \text{if} \quad \lambda_i \neq \lambda_j.$$

Solution to Problems

CHAPTER 1

$$(1) \quad AB = \begin{bmatrix} A_{1 \cdot B \cdot 1} & A_{1 \cdot B \cdot 2} & A_{1 \cdot B \cdot 3} \\ A_{2 \cdot B \cdot 1} & A_{2 \cdot B \cdot 2} & A_{2 \cdot B \cdot 3} \\ A_{3 \cdot B \cdot 1} & A_{3 \cdot B \cdot 2} & A_{3 \cdot B \cdot 3} \\ A_{4 \cdot B \cdot 1} & A_{4 \cdot B \cdot 2} & A_{4 \cdot B \cdot 3} \end{bmatrix}.$$

(2) (a) The k th column of AE_{ik} is the i th column of A , all other columns are zero.

(b) The i th row of $E_{ik}A$ is the k th row of A , all other rows are zero.

$$AE_{ik} = A e_i e_k' = A_{i \cdot} e_k'$$

$$E_{ik}A = e_i e_k' A = e_i A_{k \cdot}'.$$

$$(3) \quad \begin{aligned} \text{tr } ABC &= \sum_i e_i' ABC e_i = \sum_i (e_i' A) B (C e_i) \\ &= \sum_i A_{i \cdot}' B C_{\cdot i}. \end{aligned}$$

$$(4) \quad \begin{aligned} \text{tr } AE_{ij} &= \sum_k e_k' E_{ij} e_k = \sum_{k,r,s} e_k' a_{rs} E_{rs} E_{ij} e_k \\ &= \sum_{k,r,s} a_{rs} e_k' e_r e_s' e_i e_j' e_k \\ &= \sum_{k,r,s} a_{rs} \delta_{kr} \delta_{si} \delta_{jk} = a_{ji}. \end{aligned}$$

$$\begin{aligned}
 (5) \quad A &= \sum_{i,j} a_{ij} A_{ij} = \sum_{i,j} \text{tr}(B E_{ij} \delta_{ij}) E_{ij} \\
 &= \sum_{i,j,k} \mathbf{e}_k' B E_{ij} \mathbf{e}_k E_{ii} = \sum_{i,j,k} \mathbf{e}_k' B \mathbf{e}_i \delta_{ik} E_{ii} \\
 &= \sum_i \mathbf{e}_i' B \mathbf{e}_i E_{ii} = \sum_i b_{ii} E_{ii} = \text{diag}\{B\}.
 \end{aligned}$$

CHAPTER 2

- (1) Since U is an orthogonal matrix, the result follows.
More formally,

$$\begin{aligned}
 &\sum_{r,s} [E_{rs}(m \times n) \otimes E_{sr}(n \times m)] [E_{sr}(n \times m) \otimes E_{rs}(m \times n)] \\
 &= \sum_{r,s} [E_{rs}(m \times n) E_{sr}(n \times m)] \otimes [E_{sr}(n \times m) E_{rs}(m \times n)] \\
 &= \sum_{r,s} [\delta_{ss} E_{rr}(m \times m)] \otimes [\delta_{rr} E_{ss}(n \times n)] \\
 &= \sum_r E_{rr}(m \times m) \otimes \left[\sum_s E_{ss}(n \times n) \right] \\
 &= I_m \otimes I_n = I_{mn}, \text{ the result follows.}
 \end{aligned}$$

$$(3) \quad (a) \quad A \otimes B = \begin{bmatrix} -2 & 2 & -1 & 1 \\ 4 & 0 & 2 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix}, \quad B \otimes A = \begin{bmatrix} -2 & -1 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 4 & 2 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}$$

$$(b) \quad U_1 = U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

- (4) See [18] p. 228 for methods of calculating matrix exponentials.

$$(a) \quad \exp(A) = \begin{bmatrix} 2e - e^{-1} & 2(e - e^{-1}) \\ e^{-1} - e & 2e^{-1} - e \end{bmatrix}$$

$$\begin{aligned}
 \exp(A \otimes I) &= \begin{bmatrix} 2e - e^{-1} & 0 & 2(e - e^{-1}) & 0 \\ 0 & 2e - e^{-1} & 0 & 2(e - e^{-1}) \\ -(e - e^{-1}) & 0 & -e + 2e^{-1} & 0 \\ 0 & -(e - e^{-1}) & 0 & -e + 2e^{-1} \end{bmatrix} \\
 \exp(A) \otimes I &= \begin{bmatrix} 2e - e^{-1} & 0 & 2(e - e^{-1}) & 0 \\ 0 & 2e - e^{-1} & 0 & 2(e - e^{-1}) \\ e^{-1} - e & 0 & 2e^{-1} - e & 0 \\ 0 & e^{-1} - e & 0 & 2e^{-1} - e \end{bmatrix}.
 \end{aligned}$$

Hence $\exp(A) \otimes I = \exp(A \otimes I)$.

$$\begin{aligned}
 (5) \text{ (a)} \quad A^{-1} &= \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}, \quad B^{-1} = \frac{1}{2} \begin{bmatrix} -4 & 2 \\ 3 & -1 \end{bmatrix}, \quad \text{so that} \\
 A^{-1} \otimes B^{-1} &= \frac{1}{2} \begin{bmatrix} -4 & 2 & -4 & 2 \\ 3 & -1 & 3 & -1 \\ 4 & -2 & 3 & -1 \\ -3 & 1 & -6 & 2 \end{bmatrix}.
 \end{aligned}$$

$$\begin{aligned}
 \text{(b) As} \quad A \otimes B &= \begin{bmatrix} 2 & 4 & 1 & 2 \\ 6 & 8 & 3 & 4 \\ -1 & -2 & -1 & -2 \\ -3 & -4 & -3 & -4 \end{bmatrix}, \quad \text{it follows that} \\
 (A \otimes B)^{-1} &= \begin{bmatrix} -2 & 1 & -2 & 1 \\ 3/2 & -1/2 & 3/2 & -1/2 \\ 2 & -1 & 4 & -2 \\ -3/2 & 1/2 & -3 & -4 \end{bmatrix}.
 \end{aligned}$$

This verifies (2.12)

$$(6) \text{ (a) For } A; \lambda_1 = -1, \lambda_2 = 2, x'_1 = [1 \ 4] \text{ and } x'_2 = [1 \ 1].$$

$$\text{For } B; \mu_1 = 1, \mu_2 = 4, y'_1 = [1 \ -1] \text{ and } y'_2 = [1 \ 2].$$

$$\text{(b) } A \otimes B = \begin{bmatrix} 6 & 3 & -2 & -1 \\ 6 & 9 & -2 & -3 \\ 8 & 4 & -4 & -2 \\ 8 & 12 & -4 & -6 \end{bmatrix} = E \text{ (say)}.$$

Solution to Problems

$$C(\lambda) = |\lambda I - E| = \lambda^4 - 5\lambda^3 - 30\lambda^2 + 40\lambda - 30 \\ = (\lambda + 1)(\lambda + 4)(\lambda - 2)(\lambda - 8).$$

Hence the eigenvalues of E are

$$\{-1, -4, 2, 8\} = \{\lambda_1 \mu_1, \lambda_1 \mu_2, \lambda_2 \mu_1, \lambda_2 \mu_2\}.$$

The corresponding eigenvectors of E are:

$$\begin{bmatrix} 1 \\ -1 \\ 4 \\ -4 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 5 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}.$$

(c) This verifies Property IX

(7) For some non-singular P and Q

$$A = P^{-1}CP \quad \text{and} \quad B = Q^{-1}DQ.$$

Hence

$$\begin{aligned} A \otimes B &= P^{-1}CP \otimes Q^{-1}DQ \\ &= (P^{-1} \otimes Q^{-1})(CP \otimes DQ) \quad \text{by (2.11)} \\ &= (P \otimes Q)^{-1}(C \otimes D)(P \otimes Q) \quad \text{by (2.12) and (2.11)} \\ &= R^{-1}(C \otimes D)R \end{aligned}$$

where

$$R = P \otimes Q.$$

The result follows.

CHAPTER 4

$$(1) \quad \frac{\partial y}{\partial X} = \begin{bmatrix} 2x_{22} & 0 & -x_{21} \\ -x_{13} & 2x_{11} & 0 \end{bmatrix}, \quad \frac{\partial Y}{\partial x} = \begin{bmatrix} 1 & 2e^{2x} \\ 0 & -x^{-2} \\ 4x & \cos x \end{bmatrix}$$

$$(2)(a) \quad |X| = x \sin x - \exp(2x)$$

$$\frac{\partial |X|}{\partial X} = \begin{bmatrix} e^x & -\cos x \\ -x & \sin x \end{bmatrix}.$$

$$|X| = \exp(x) \sin x - x \cos x,$$

$$\frac{\partial |X|}{\partial X} = \begin{bmatrix} x & -2e^{2x} \\ -2e^{2x} & \sin x \end{bmatrix}.$$

$$(b) \quad \frac{\partial |X|}{\partial X} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} e^x & -\cos x \\ -x & \sin x \end{bmatrix}.$$

$$\begin{aligned} \frac{\partial |X|}{\partial X} &= 2[X_{ii}] - \text{diag}\{X_{ii}\} \\ &= \begin{bmatrix} 2x & -2e^x \\ -2e^x & 2 \sin x \end{bmatrix} - \begin{bmatrix} x & 0 \\ 0 & \sin x \end{bmatrix}. \end{aligned}$$

$$(3) \quad Y = \begin{bmatrix} x_{11}^2 + x_{21}^2 & x_{11}x_{12} + x_{21}x_{22} & x_{11}x_{13} + x_{21}x_{23} \\ x_{12}x_{11} + x_{22}x_{21} & x_{12}^2 + x_{22}^2 & x_{12}x_{13} + x_{22}x_{23} \\ x_{13}x_{11} + x_{23}x_{21} & x_{13}x_{12} + x_{23}x_{22} & x_{13}^2 + x_{23}^2 \end{bmatrix},$$

hence

$$(a) \quad \frac{\partial Y}{\partial x_{21}} = \begin{bmatrix} 2x_{21} & x_{22} & x_{23} \\ x_{22} & 0 & 0 \\ x_{23} & 0 & 0 \end{bmatrix},$$

From Example 4.8

$$\frac{\partial Y}{\partial x_{21}} = E'_{21}X + X'E_{21} =$$

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} + \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

(b) Since $y_{13} = x_{11}x_{13} + x_{21}x_{23}$

$$\frac{\partial y_{13}}{\partial X} = \begin{bmatrix} x_{13} & 0 & x_{11} \\ x_{23} & 0 & x_{21} \end{bmatrix} =$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which is the result in Example 4.8.

116

Solution to Problems

$$(4) (a) \quad \frac{\partial Y}{\partial x_{rs}} = E_{rs}AX + XAE_{rs},$$

$$\frac{\partial y_{ij}}{\partial X} = E_{ij}X'A' + A'X'E_{ij}.$$

$$(b) \quad \frac{\partial Y}{\partial x_{rs}} = E'_{rs}AX' + X'AE'_{rs},$$

$$\frac{\partial y_{ij}}{\partial X} = AX'E'_{ij} + E'_{ij}X'A'.$$

(5) By (4.10)

$$\begin{aligned} \frac{\partial |Y|}{\partial x_{rs}} &= \text{tr} \{ |Y|(Y^{-1})'B'E'_{rs}A' \} \\ &= |Y| \text{tr} \{ A'(Y^{-1})'B'E'_{rs} \} \\ &= |Y| (\text{vec } E_{rs})' \text{vec} [A'(Y^{-1})'B'] \\ &= |AXB| z_{rs} \end{aligned}$$

where

$$[z_{rs}] = Z = A'[(AXB)^{-1}]'B'.$$

(6) (a) Since

$$\frac{\partial (XX')}{\partial x_{rs}} = E_{rs}X' + XE'_{rs},$$

$$\frac{\partial Y}{\partial x_{rs}} = E_{rs}(X')^2 + XE'_{rs}X' + XX'E'_{rs}$$

$$(b) \quad \frac{\partial Y}{\partial x_{rs}} = E'_{rs}X'X + X'E'_{rs}X + (X')^2E_{rs}.$$

CHAPTER 5

(1) Since

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \end{bmatrix} = \begin{bmatrix} a_{11}x_{11} + a_{12}x_{12} \\ a_{21}x_{11} + a_{22}x_{12} \\ a_{11}x_{21} + a_{12}x_{22} \\ a_{21}x_{21} + a_{22}x_{22} \end{bmatrix},$$

$$\frac{\partial \text{vec } Y}{\partial \text{vec } X} = \begin{bmatrix} a_{11} & a_{21} & 0 & 0 \\ 0 & 0 & a_{11} & a_{21} \\ a_{12} & a_{22} & 0 & 0 \\ 0 & 0 & a_{12} & a_{22} \end{bmatrix}.$$

$$(2)(a) \quad \frac{\partial \text{vec } Y}{\partial \text{vec } X} = (B \otimes A')_{(n)} \quad \text{by (5.18)}$$

$$(b) \quad \frac{\partial \text{vec } Y}{\partial \text{vec } X} = X \otimes I + I \otimes X'.$$

$$(3)(a) \quad \frac{\partial \text{tr } Y}{\partial x_{rs}} = \text{tr } A E_{rs} B = \text{tr } E'_{rs} A' B' = (\text{vec } E_{rs})' (\text{vec } A' B'),$$

hence

$$\frac{\partial \text{tr } Y}{\partial X} = A' B'.$$

$$(b) \quad \frac{\partial \text{tr } Y}{\partial x_{rs}} = 2 \text{tr } E'_{rs} X',$$

hence

$$\frac{\partial \text{tr } Y}{\partial X} = 2X'.$$

$$(c) \quad \frac{\partial \text{tr } Y}{\partial x_{rs}} = 2 \text{tr } E'_{rs} X,$$

hence

$$\frac{\partial \text{tr } Y}{\partial X} = 2X.$$

$$(4)(a) \quad \frac{\partial \text{tr } Y}{\partial x_{rs}} = -\text{tr } X^{-1} E'_{rs} X^{-1} = -\text{tr } E'_{rs} (X^{-2})',$$

hence

$$\frac{\partial \text{tr } Y}{\partial X} = (X^{-2})'.$$

$$(b) \quad \frac{\partial \text{tr } Y}{\partial x_{rs}} = -\text{tr } A X^{-1} E'_{rs} X^{-1} B,$$

hence

$$\frac{\partial \text{tr } Y}{\partial X} = -(X^{-1} B A X^{-1})'.$$

$$(c) \quad \frac{\partial \operatorname{tr} Y}{\partial x_{rs}} = \operatorname{tr} E_{rs} X^{n-1} + \operatorname{tr} X E_{rs} X^{n-2} + \dots + \operatorname{tr} X^{n-1} E_{rs}$$

hence

$$\frac{\partial \operatorname{tr} Y}{\partial x_{rs}} = n(X^{n-1})'$$

$$(d) \quad \exp(X) = I + X + \frac{1}{2!} X^2 + \frac{1}{3!} X^3 + \dots$$

hence by the result (c) above

$$\frac{\partial \operatorname{tr} Y}{\partial X} = \exp(X').$$

$$(5)(a)(i) \quad dY = \begin{bmatrix} a_{11} dx_{11} + a_{12} dx_{21} & a_{11} dx_{12} + a_{12} dx_{22} \\ a_{21} dx_{11} + a_{22} dx_{21} & a_{21} dx_{12} + a_{22} dx_{22} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} dx_{11} & dx_{12} \\ dx_{21} & dx_{22} \end{bmatrix} = A(dX).$$

$$(ii) \quad dY = \begin{bmatrix} 2x_{11} dx_{11} + 2x_{21} dx_{21} & x_{11} dx_{12} + x_{12} dx_{11} + x_{21} dx_{22} + x_{22} dx_{21} \\ x_{11} dx_{12} + x_{12} dx_{11} + x_{21} dx_{22} + x_{22} dx_{21} & 2x_{12} dx_{12} + 2x_{22} dx_{22} \end{bmatrix}$$

$$= \begin{bmatrix} dx_{11} & dx_{21} \\ dx_{12} & dx_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} + \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \begin{bmatrix} dx_{11} & dx_{12} \\ dx_{21} & dx_{22} \end{bmatrix}$$

$$= (dX)'X + X'(dX).$$

(iii)

$$dY = \begin{bmatrix} 2x_{11} dx_{11} + x_{12} dx_{21} + x_{21} dx_{12} & x_{11} dx_{12} + x_{12} dx_{11} + x_{21} dx_{22} + x_{22} dx_{12} \\ x_{11} dx_{21} + x_{21} dx_{11} + x_{22} dx_{21} + x_{21} dx_{22} & x_{21} dx_{12} + x_{12} dx_{21} + 2x_{22} dx_{22} \end{bmatrix}$$

$$= \begin{bmatrix} x_{11} dx_{11} + x_{12} dx_{21} & x_{11} dx_{12} + x_{12} dx_{22} \\ x_{21} dx_{11} + x_{22} dx_{21} & x_{21} dx_{12} + x_{22} dx_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} x_{11} dx_{11} + x_{21} dx_{12} & x_{12} dx_{11} + x_{22} dx_{12} \\ x_{11} dx_{21} + x_{21} dx_{21} & x_{12} dx_{21} + x_{22} dx_{22} \end{bmatrix}$$

$$= X(dX) + (dX)X.$$

- (b) Write $Y = UV$ where $U = AX$ and $V = BX$,
 then $dY = U(dV) + (dU)V$
 $= AXB(dX) + A(dX)BX.$

CHAPTER 6

$$(1) \quad \frac{\partial Y}{\partial X} = \begin{bmatrix} -\sin(x_{12} + x_{22}) & x_{21} & 0 & 0 \\ x_{12} e^{x_{11} x_{12}} & 0 & x_{11} e^{x_{11} x_{12}} & x_{22} \\ 0 & x_{11} & -\sin(x_{12} + x_{22}) & 0 \\ 0 & 0 & 0 & x_{12} \end{bmatrix}$$

$$(2) \quad \frac{\partial X}{\partial x_{11}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \frac{\partial X}{\partial x_{12}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \frac{\partial X}{\partial x_{13}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{and so on,}$$

hence by (6.1)

$$\frac{\partial X}{\partial X} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \bar{U}.$$

$$(3) \text{ Since } X^{-1} = \frac{1}{\Delta} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix}$$

where $\Delta = x_{11}x_{22} - x_{12}x_{21}$, we can calculate $\partial X^{-1}/\partial x_{rs}$, for example

$$\frac{\partial X^{-1}}{\partial x_{11}} = \frac{1}{\Delta^2} \begin{bmatrix} -x_{22}^2 & x_{12}x_{22} \\ x_{21}x_{22} & -x_{12}x_{21} \end{bmatrix}.$$

Hence

$$\begin{aligned} \frac{\partial X^{-1}}{\partial X} &= \frac{1}{\Delta^2} \begin{bmatrix} x_{22}^2 & -x_{12}x_{22} & -x_{22}x_{21} & x_{11}x_{22} \\ -x_{21}x_{22} & x_{12}x_{21} & x_{21}^2 & -x_{11}x_{21} \\ -x_{12}x_{22} & x_{12}^2 & x_{12}x_{21} & -x_{11}x_{12} \\ x_{11}x_{22} & -x_{12}x_{11} & -x_{11}x_{21} & x_{11}^2 \end{bmatrix} \\ &= -\frac{1}{\Delta} \begin{bmatrix} x_{22} & -x_{12} & 0 & 0 \\ -x_{21} & x_{11} & 0 & 0 \\ 0 & 0 & x_{22} & -x_{12} \\ 0 & 0 & -x_{21} & x_{11} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{22} & -x_{12} & 0 & 0 \\ -x_{21} & x_{11} & 0 & 0 \\ 0 & 0 & x_{22} & -x_{12} \\ 0 & 0 & -x_{21} & x_{11} \end{bmatrix} \\ &= -(I \otimes X^{-1}) \bar{U} (I \otimes X^{-1}). \end{aligned}$$

(4)

$$A \otimes X^{-1} = \frac{1}{\Delta} \begin{bmatrix} a_{11}x_{22} & -a_{11}x_{12} & a_{12}x_{22} & -a_{12}x_{12} \\ -a_{11}x_{21} & a_{11}x_{11} & -a_{12}x_{21} & a_{12}x_{11} \\ a_{21}x_{22} & -a_{21}x_{12} & a_{22}x_{22} & -a_{22}x_{12} \\ -a_{21}x_{21} & a_{21}x_{11} & -a_{22}x_{21} & a_{22}x_{11} \end{bmatrix}$$

where $\Delta = x_{11}x_{22} - x_{12}x_{21}$.

We can now calculate

$$\frac{\partial(A \otimes X^{-1})/\partial x_{rs}}{\partial x_{rs}}$$

and form

$$\frac{\partial(A \otimes X^{-1})}{\partial X} = \frac{1}{\Delta} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -a_{11} & 0 & -a_{12} \\ 0 & a_{11} & 0 & a_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{21} & 0 & -a_{22} \\ 0 & a_{21} & 0 & a_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{11} & 0 & a_{12} & 0 \\ -a_{11} & 0 & -a_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{21} & 0 & a_{22} & 0 \\ -a_{21} & 0 & -a_{22} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Tables of Formulae and Derivatives

Table 1

Notation used: $A = [a_{ij}]$, $B = [b_{ij}]$

$$\begin{aligned}
 E_{ij} &= \mathbf{e}_i \mathbf{e}_j' \\
 \delta_{ij} &= \mathbf{e}_i' \mathbf{e}_j = \mathbf{e}_j' \mathbf{e}_i \\
 E_{ij} \mathbf{e}_r &= \delta_{jr} \mathbf{e}_i \\
 E_{ij} E_{rs} &= \delta_{jr} E_{is} \\
 E_{ij} E_{js} E_{sm} &= E_{im} \\
 E_{ij} E_{rs} &= 0 \text{ if } j \neq r \\
 A &= \sum_i \sum_j a_{ij} E_{ij} \\
 A_{.j} &= A \mathbf{e}_j \\
 A_{j.} &= A' \mathbf{e}_j \\
 E_{ij} A E_{rs} &= a_{jr} E_{ij} \\
 \text{tr } AB &= \sum_i \sum_j a_{ij} b_{ji} \\
 \text{tr } AB' &= \text{tr } A'B. \\
 \text{tr } AB &= (\text{vec } A')' \text{vec } B.
 \end{aligned}$$

Table 2

$$\begin{aligned}
 A \otimes B &= [a_{ij}B] \\
 A \otimes (\alpha B) &= \alpha(A \otimes B) \\
 (A + B) \otimes C &= A \otimes C + B \otimes C \\
 A \otimes (B + C) &= A \otimes B + A \otimes C \\
 A \otimes (B \otimes C) &= (A \otimes B) \otimes C \\
 (A \otimes B)' &= A' \otimes B' \\
 (A \otimes B)(C \otimes D) &= AC \otimes BC \\
 (A \otimes B)^{-1} &= A^{-1} \otimes B^{-1} \\
 \text{vec}(AYB) &= (B' \otimes A) \text{vec } Y \\
 |A \otimes B| &= |A|^m |B|^n \text{ when } A \text{ and } B \text{ are of order} \\
 &\quad (n \times n) \text{ and } (m \times m) \text{ respectively} \\
 A \otimes B &= U_1(B \otimes A)U_2, U_1 \text{ and } U_2 \text{ are permutation} \\
 &\quad \text{matrices} \\
 \text{tr}(A \otimes B) &= \text{tr } A \text{ tr } B \\
 A \oplus B &= A \otimes I_m + I_n \otimes B \\
 U &= \sum_r \sum_s E_{rs} \otimes E'_{rs}
 \end{aligned}$$

Table 3

$$\begin{aligned}
 \frac{\partial(Ax)}{\partial x} &= A' \\
 \frac{\partial(x'A)}{\partial x} &= A \\
 \frac{\partial(x'x)}{\partial x} &= 2x \\
 \frac{\partial(x'Ax)}{\partial x} &= Ax + A'x \\
 \frac{\partial z}{\partial x} &= \frac{\partial y}{\partial x} \frac{\partial z}{\partial y}
 \end{aligned}$$

Table 4

$$\frac{\partial f(X)}{\partial X} = \sum \sum E_{ij} \frac{\partial f(X)}{\partial x_{ij}}$$

$$\begin{aligned} \frac{\partial |X|}{\partial X} &= |X|(X^{-1}), \text{ when elements of } X \text{ are} \\ &\quad \text{independent} \\ &= 2[X_{ii}] - \text{diag } \{X_{ii}\}, \text{ when } X \text{ is symmetric.} \end{aligned}$$

$$\frac{\partial X}{\partial x_{rs}} = E_{rs}$$

$$\frac{\partial X'}{\partial x_{rs}} = E'_{rs}$$

$$\frac{\partial (AXB)}{\partial x_{rs}} = AE_{rs}B$$

$$\frac{\partial (AX'B)}{\partial x_{rs}} = AE'_{rs}B$$

$$\frac{\partial (X'A'AX)}{\partial x_{rs}} = E'_{rs}A'AX + X'A'AE_{rs}$$

$$\frac{\partial (AX^{-1}B)}{\partial x_{rs}} = -AX^{-1}E_{rs}X^{-1}B$$

$$\frac{\partial (X'AX)}{\partial x_{rs}} = E'_{rs}AX + X'AE_{rs}$$

$$\frac{\partial (X^n)}{\partial x_{rs}} = \sum_{k=0}^{n-1} X^k E_{rs} X^{n-k-1}$$

$$\frac{\partial (X^{-n})}{\partial x_{rs}} = -X^{-n} \left[\sum_{k=0}^{n-1} X^k E_{rs} X^{n-k-1} X^{-n} \right]$$

Table 5

$$\frac{\partial \operatorname{vec}(AXB)}{\partial \operatorname{vec} X} = B' \otimes A$$

$$\frac{\partial \operatorname{vec}(X'AX)}{\partial \operatorname{vec} X'} = U'(AX \otimes I) + (I \otimes A'X)$$

$$\frac{\partial \operatorname{vec}(AX^{-1}B)}{\partial \operatorname{vec} X} = -(X^{-1}B) \otimes (X^{-1})'A'$$

Table 6

$$\frac{\partial \log |X|}{\partial X} = (X^{-1})'$$

$$\frac{\partial |X|^r}{\partial X} = r|X|^{r-1}(X^{-1})'$$

$$\frac{\partial \operatorname{tr}(AX)}{\partial X} = A'$$

$$\frac{\partial \operatorname{tr}(A'X)}{\partial X} = A$$

$$\frac{\partial \operatorname{tr}(X'AXB)}{\partial X} = AXB + A'XB'$$

$$\frac{\partial \operatorname{tr}(XX')}{\partial X} = 2X$$

$$\frac{\partial \operatorname{tr}(X^n)}{\partial X} = nX^{n-1}$$

$$\frac{\partial \operatorname{tr}(e^X)}{\partial X} = e^X$$

$$\frac{\partial \operatorname{tr}(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})'$$

Table 7

$$\frac{\partial Y}{\partial X} = \Sigma \Sigma E_{rs} \otimes \frac{\partial Y}{\partial x_{rs}}$$

$$\frac{\partial X}{\partial X} = \bar{U} + U - \Sigma E_{rr} \otimes E_{rr} \quad (X \text{ symmetric})$$

$$\frac{\partial X}{\partial X} = \bar{U} \quad (\text{elements of } X \text{ independent})$$

$$\frac{\partial X'}{\partial X} = U$$

$$\frac{\partial (XY)}{\partial Z} = \frac{\partial X}{\partial Z} (I \otimes Y) + (I \otimes X) \frac{\partial Y}{\partial Z}$$

$$\frac{\partial X^{-1}}{\partial X} = -(I \otimes X^{-1}) \bar{U} (I \otimes X^{-1})$$

$$\frac{\partial (X \otimes Y)}{\partial Z} = \frac{\partial X}{\partial Z} \otimes Y + [I \otimes U_1] \left[\frac{\partial Y}{\partial Z} \otimes X \right] [I \otimes U_2]$$

Bibliography

- [1] Anderson, T. W., (1958), *An Introduction to Multivariate Statistical Analysis*, John Wiley.
- [2] Athans, M., (1968), *The Matrix Minimum Principle*, Information and Control, 11, 592-606.
- [3] Athans, M., and Tse, E., (1967), A Direct Derivation of the Optimal Linear Filter Using the Maximum Principle, *IEEE Trans. Auto. Control*, AC-12, No. 6, 690-698.
- [4] Athans M., and Schweppe, F. C., (1965), *Gradient Matrices and Matrix Calculations*, MIT Lincoln Lab. Tech., Note 1965-53, Lexington, Mass.
- [5] Barnett, S., (1973), Matrix Differential Equations and Kronecker Products, *SIAM, J. Appl. Math.*, 24, No. 1.
- [6] Bellman, R., (1960), *Introduction to Matrix Analysis*, McGraw-Hill.
- [7] Bodewig, E., (1959), *Matrix Calculus*, Amsterdam: North Holland Publishing Co.
- [8] Brewer, J. W. (1978), Kronecker Products and Matrix Calculus in System Theory, *IEEE Trans. on Circuits and Systems*, 25, No. 9, 772-781.
- [9] Brewer, J. W., (1977), The Derivatives of the Exponential Matrix with respect to a Matrix, *IEEE Trans. Auto. Control*, 22, 656-657.
- [10] Brewer, J. W., (1979), Derivatives of the Characteristic Polynomial Trace and Determinant with respect to a Matrix, *IEEE Trans. Auto. Control*, 24, 787-790.
- [11] Brewer, J. W., (1977), The Gradient with respect to a Symmetric Matrix, *IEEE Trans. Auto. Control*, 22, 265-267.
- [12] Brewer, J. W., (1977), The Derivative of the Riccati Matrix with respect to a Matrix, *IEEE Trans. Auto. Control*, 22, No. 6, 980-983.
- [13] Conlisk, J. (1969), The Equilibrium Covariance Matrix of Dynamic Econometric Models, *American Stat. Ass. Journal*, No. 64, 277-279.
- [14] Deemer, W. L. and Olkin, I., (1951), The Jacobians of certain Matrix Transformations, *Biometrika*, 30, 345-367.

- [15] Dwyer, P. S. and Macphail, M. S., (1948), Symbolic Matrix Derivatives, *Ann. Math. Statist.*, 19, 517-537.
- [16] Dwyer, P. S., (1967), Some Applications of Matrix Derivatives in Multivariate Analysis, *American Statistical Ass. Journal*, June, pt 62, 607-625.
- [17] Geering, H. P., (1976), On Calculating Gradient Matrices, *IEEE Trans. Auto. Control*, August, 615-616.
- [18] Graham, A., (1979), *Matrix Theory and Applications for Engineers and Mathematicians*, Ellis Horwood.
- [19] Graham, A., and Burghies, D., (1980), *Introduction to Control Theory Including Optimal Control*, Ellis Horwood.
- [20] Lancaster, P., (1970), Explicit Solutions of Linear Matrix Equations, *SIAM Rev.*, 12, No. 4, 544-566.
- [21] MacDuffee, C. C. (1956), *The Theory of Matrices*, Chelsea, New York.
- [22] Neudecker, H. (1969), Some Theorems on Matrix Differentiation with special reference to Kronecker Matrix Products, *J. Amer. Statist. Assoc.*, 64, 953-963.
- [23] Neudecker, H., *A Note of Kronecker Matrix Products and Matrix Equation Systems*.
- [24] Paraskevopoulos, P. N. and King, R. E., (1976), A Kronecker Product approach to Pole assignment by output feedback, *Int. J. Contr.*, 24, No. 3, 325-334.
- [25] Roth, W. E., (1944), On Direct Product Matrices, *Bull. Amer. Math. Soc.*, No. 40, 461-468.
- [26] Schoneemann, P. H., (1965), *On the Formal Differentiation of Traces and Determinants*, Research Memorandum No. 27, University of North Carolina.
- [27] Schweppe, F. C., (1973), *Uncertain Dynamic Systems*, Englewood Cliffs, Prentice Hall.
- [28] Tracy, D. S. and Dwyer, P. S., (1969), Multivariate Maxima and Minima with Matrix Derivatives, *J. Amer. Statist. Assoc.*, 64, 1576-1594.
- [29] Turnbull, H. W., (1927), On Differentiating a Matrix, *Proc. Edinburgh Math. Soc.*, 11, ser. 2, 111-128.
- [30] Turnbull, H. W., (1930/31), 'A Matrix Form of Taylor's Theorem', *Proc. Edinburgh Math. Soc.*, Ser. 2, 33-54.
- [31] Vetter, W. J., (1970), Derivative Operations on Matrices, *IEEE Trans. Auto. Control*, AC-15, 241-244.
- [32] Vetter, W. J., (1971), Correction to 'Derivative Operations on Matrices', *IEEE Trans. Auto. Control*, AC-16, 113.
- [33] Vetter, W. J., (1971), An Extension to Gradient Matrices, *IEEE Trans. Syst. Man. Cybernetics*, SMC-1, 184-186.
- [34] Vetter, W. J., (1973), Matrix Calculus Operations and Taylor Expansions, *SIAM Rev.*, 2, 352-369.
- [35] Vetter, W. J., (1975), Vector Structures and Solutions of Linear Matrix Equations, *Linear Algebra and its Applications*, 10, 181-188.

1

Bibliography

- [W. J., (1971), On Linear Estimates, Minimum Variance and Least-Weighting Matrices, *IEEE Trans. Auto. Control*, AC-16, 265-
- [S. R. J. and Mulholland, R. J., (1980), Kronecker Product Representation or the Solution of the General Linear Matrix Equation, *IEEE Trans. Control*, AC-25, No. 3, 563-564.

Index

C

Chain Rule
 matrix, 88
 vector, 54
characteristic equation, 47
cofactor, 57
column vector, 14
companion form, 47
constrained optimisation, 94, 96

D

decomposition of a matrix, 13
direct product, 21
derivative
 Kronecker product, 70
 matrix, 60, 62, 64, 67, 70, 75, 81
 scalar function, 56, 75
 vector, 52
determinant, 27, 56
deviation, 94

E

Eigenvalues, 27, 30
eigenvectors, 27, 30
elementary matrix, 12, 19
 transpose, 19
exponential matrix, 29, 31, 42, 108

G

gradient matrix, 56

J

Jacobian, 53, 109

K

Kronecker delta, 13
 product, 21, 23, 33, 70, 85
 sum, 30

L

Langrange multipliers, 95
least squares, 94, 96, 100

M

Matrix

 calculus, 51, 94
 companion, 47
 decomposition, 13
 derivative, 37, 60, 62, 67, 70, 75,
 81, 84, 88
 differential, 78
 elementary, 12, 19
 exponential, 29, 31, 42, 108
 gradient, 56
 integral, 37
 orthogonal, 97
 permutation, 23, 28, 32
 product rule, 84
 symmetric, 58, 95, 97
 transition, 42
 maximum likelihood, 102
 mixed product rule, 24
 multivariable system, 45
 multivariate normal, 102

N

normal equations, 95, 101