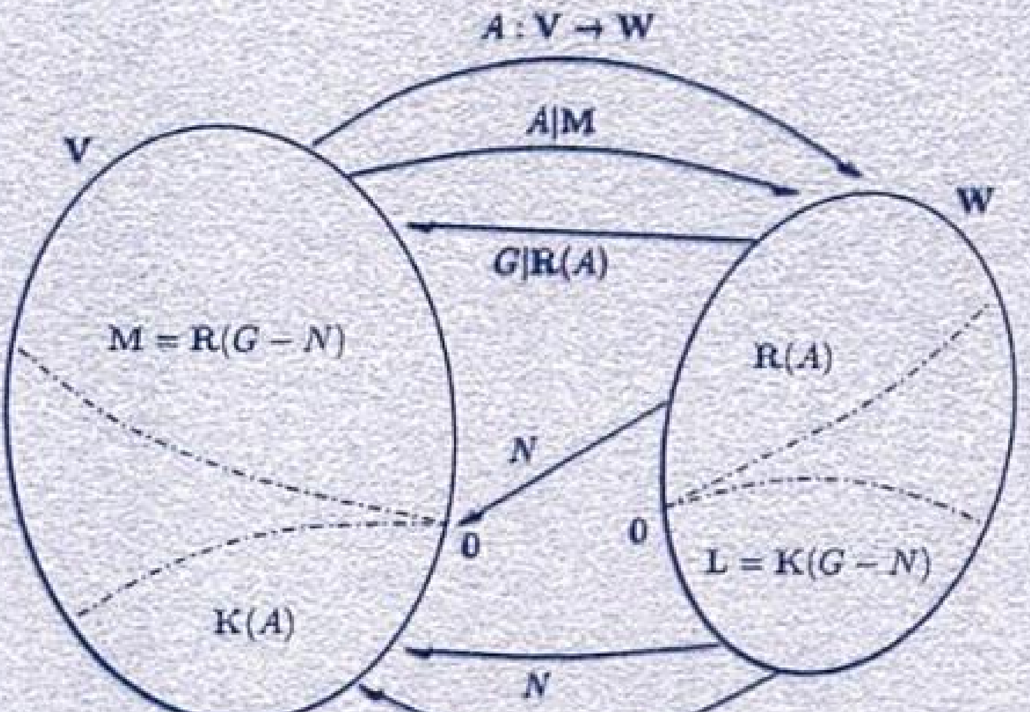
Matrix Algebra and Its Applications to Statistics and Econometrics

C. Radhakrishna Rao
M. Bhaskara Rao



World Scientific

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Rao, C. Radhakrishna (Calyampudi Radhakrishna), 1920–
Matrix algebra and its applications to statistics and econometrics / C. Radhakrishna Rao and M. Bhaskara Rao.
p. cm.
Includes bibliographical references and index.
ISBN 9810232683 (alk. paper)
1. Matrices. 2. Statistics. 3. Econometrics. I. Bhaskara Rao, M.
QA188.R36 1998
512.9'434--dc21 98-5596
CIP

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

First published 1998
Reprinted 2001, 2004

Printed in Singapore by Utopia Press Pte Ltd

CONTENTS

Preface vii
Notation xi

CHAPTER 1. VECTOR SPACES

1.1 Rings and Fields 1
1.2 Mappings 14
1.3 Vector Spaces 16
1.4 Linear Independence and Basis of a Vector Space 19
1.5 Subspaces 24
1.6 Linear Equations 29
1.7 Dual Space 35
1.8 Quotient Space 41
1.9 Projective Geometry 42

CHAPTER 2. UNITARY AND EUCLIDEAN SPACES

2.1 Inner Product 51
2.2 Orthogonality 56
2.3 Linear Equations 66
2.4 Linear Functionals 71
2.5 Semi-inner Product 76
2.6 Spectral Theory 83
2.7 Conjugate Bilinear Functionals and Singular Value Decomposition 101

CHAPTER 3. LINEAR TRANSFORMATIONS AND MATRICES

3.1 Preliminaries 107
3.2 Algebra of Transformations 110
3.3 Inverse Transformations 116
3.4 Matrices 120

CHAPTER 4. CHARACTERISTICS OF MATRICES

4.1 Rank and Nullity of a Matrix 128
4.2 Rank and Product of Matrices 131
4.3 Rank Factorization and Further Results 136
4.4 Determinants 142
4.5 Determinants and Minors 146

CHAPTER 5. FACTORIZATION OF MATRICES

5.1 Elementary Matrices 157
5.2 Reduction of General Matrices 160
5.3 Factorization of Matrices with Complex Entries 166
5.4 Eigenvalues and Eigenvectors 177
5.5 Simultaneous Reduction of Two Matrices 184
5.6 A Review of Matrix Factorizations 188

CHAPTER 6. OPERATIONS ON MATRICES

6.1 Kronecker Product 193
6.2 The Vec Operation 200
6.3 The Hadamard-Schur Product 203
6.4 Khatri-Rao Product 216
6.5 Matrix Derivatives 223

CHAPTER 7. PROJECTORS AND IDEMPOTENT OPERATORS

7.1 Projectors 239
7.2 Invariance and Reducibility 245
7.3 Orthogonal Projection 248
7.4 Idempotent Matrices 250
7.5 Matrix Representation of Projectors 256

CHAPTER 8. GENERALIZED INVERSES

8.1 Right and Left Inverses 264
8.2 Generalized Inverse (g-inverse) 265
8.3 Geometric Approach: LMN-inverse 282
8.4 Minimum Norm Solution 288

8.5 Least Squares Solution 289
8.6 Minimum Norm Least Squares Solution 291
8.7 Various Types of g-inverses 292
8.8 G-inverses Through Matrix Approximations 296
8.9 Gauss-Markov Theorem 300

CHAPTER 9. MAJORIZATION
9.1 Majorization 303
9.2 A Gallery of Functions 307
9.3 Basic Results 308

CHAPTER 10. INEQUALITIES FOR EIGENVALUES
10.1 Monotonicity Theorem 322
10.2 Interlace Theorems 328
10.3 Courant-Fischer Theorem 332
10.4 Poincaré Separation Theorem 337
10.5 Singular Values and Eigenvalues 339
10.6 Products of Matrices, Singular Values, and Horn's Theorem 340
10.7 Von Neumann's Theorem 342

CHAPTER 11. MATRIX APPROXIMATIONS
11.1 Norm on a Vector Space 361
11.2 Norm on Spaces of Matrices 363
11.3 Unitarily Invariant Norms 374
11.4 Some Matrix Optimization Problems 383
11.5 Matrix Approximations 388
11.6 M, N-invariant Norm and Matrix Approximations 394
11.7 Fitting a Hyperplane to a Set of Points 398

CHAPTER 12. OPTIMIZATION PROBLEMS IN STATISTICS AND ECONOMETRICS
12.1 Linear Models 403
12.2 Some Useful Lemmas 403
12.3 Estimation in a Linear Model 406
12.4 A Trace Minimization Problem 409
12.5 Estimation of Variance 413

12.6 The Method of MINQUE: A Prologue 415
12.7 Variance Components Models and Unbiased Estimation 416
12.8 Normality Assumption and Invariant Estimators 419
12.9 The Method of MINQUE 422
12.10 Optimal Unbiased Estimation 425
12.11 Total Least Squares 428

CHAPTER 13. QUADRATIC SUBSPACES
13.1 Basic Ideas 433
13.2 The Structure of Quadratic Subspaces 438
13.3 Commutators of Quadratic Subspaces 442
13.4 Estimation of Variance Components 443

CHAPTER 14. INEQUALITIES WITH APPLICATIONS IN STATISTICS
14.1 Some Results on nnd and pd Matrices 449
14.2 Cauchy-Schwartz and Related Inequalities 454
14.3 Hadamard Inequality 456
14.4 Hölder's Inequality 457
14.5 Inequalities in Information Theory 458
14.6 Convex Functions and Jensen's Inequality 459
14.7 Inequalities Involving Moments 461
14.8 Kantorovich Inequality and Extensions 462

CHAPTER 15. NON-NEGATIVE MATRICES
15.1 Perron-Frobenius Theorem 467
15.2 Leontief Models in Economics 477
15.3 Markov Chains 481
15.4 Genetic Models 485
15.5 Population Growth Models 489

CHAPTER 16. MISCELLANEOUS COMPLEMENTS
16.1 Simultaneous Decomposition of Matrices 493
16.2 More on Inequalities 494
16.3 Miscellaneous Results on Matrices 497
16.4 Toeplitz Matrices 501
16.5 Restricted Eigenvalue Problem 506

16.6 Product of Two Raleigh Quotients 507
16.7 Matrix Orderings and Projection 508
16.8 Soft Majorization 509
16.9 Circulants 511
16.10 Hadamard Matrices 514
16.11 Miscellaneous Exercises 515

REFERENCES 519

INDEX 529

$k_1 + k_2 + \ldots + k_r = n$. Suppose

$$\begin{aligned}
\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = \ldots = \mathrm{Var}(Y_{k_1}) &= \sigma_1^2, \\
\mathrm{Var}(Y_{k_1+1}) = \mathrm{Var}(Y_{k_1+2}) = \ldots = \mathrm{Var}(Y_{k_1+k_2}) &= \sigma_2^2, \\
\ldots \qquad \ldots \qquad \ldots & \\
\mathrm{Var}(Y_{k_1+\ldots+k_{r-1}+1}) = \mathrm{Var}(Y_{k_1+\ldots+k_{r-1}+2}) = \ldots = \mathrm{Var}(Y_n) &= \sigma_r^2.
\end{aligned}$$

The mean and variances are all unknown. Develop MINQUE's of the variances.

6.4.2 Let Y_{ij}, $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$ be pairwise uncorrelated random variables with the following structure.

$$E(Y_{ij}) = \alpha_i + \beta_j \quad \text{for all } i \text{ and } j,$$
$$\mathrm{Var}(Y_{ij}) = \sigma_i^2, \; j = 1, 2, \ldots, q \quad \text{and for all } i.$$

The variances, α_i's and β_j's are all unknown. Show that the residuals are given by

$$\hat{\epsilon}_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}, \; i = 1, 2, \ldots, p \quad \text{and} \quad j = 1, 2, \ldots, q,$$

where $\bar{Y}_{i.} = \frac{1}{q}\sum_{j=1}^{q} Y_{ij}, \bar{Y}_{.j} = \frac{1}{p}\sum_{i=1}^{p} Y_{ij}$, and $\bar{Y}_{..} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} Y_{ij}$. Show that the MINQUE of σ_1^2 is given by

$$a\left(\sum_{j=1}^{q} \hat{\epsilon}_{ij}^2\right) + b\left(\sum_{i=1}^{p}\sum_{j=1}^{q} \hat{\epsilon}_{ij}^2\right),$$

where $a = [(p-1)(q-2)]^{-1}$ and $b = -[(p-1)(q-1)(q-2)]^{-1}$.

6.5. Matrix Derivatives

Suppose f is a real valued function of mn variables x_{ij}, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Suppose these variables are arranged in the form of a matrix $X = (x_{ij})$ of order $m \times n$. Assume that the partial derivatives of f exist with respect to each of its variables. The *matrix* derivative $\partial f/\partial X$ of f with respect to X is a matrix of order $m \times n$ given by

$$\frac{\partial f}{\partial X} = \left(\frac{\partial f}{\partial x_{ij}}\right),$$

i.e., the (i,j)-th element of $\partial f/\partial X$ is $\partial f/\partial X_{ij}$. If $n=1$, X is a column vector and it is denoted by x with components $x_1, x_2, \ldots, x_m$. The corresponding derivative $\partial f/\partial x$ is called the *vector* derivative of f with respect to x. More generally, suppose $F=(f_{ij})$ is a matrix function of a matrix variable X. What we mean by this is that each entry f_{ij} of the matrix F is a real valued function of the variables in X. Let F be of order $p \times q$ and X of order $m \times n$. Assume that each of the entries of F has partial derivatives with respect to all the variables in X. The *matrix* derivative $\partial F/\partial X$ of F with respect to X is defined to be the matrix

$$\frac{\partial F}{\partial X} = \left(\frac{\partial f_{ij}}{\partial X}\right) \tag{6.5.1}$$

of order $pm \times qn$ broken up into pq partitions or compartments strung along p rows and q columns. Each partition of the matrix derivative is of order $m \times n$. As an illustration, suppose F is of order 2×4 and X is of order 3×2. The matrix (6.5.1) comports itself as

$$\frac{\partial F}{\partial X} = \left[\begin{array}{cc|cc|cc|cc}
\frac{\partial f_{11}}{\partial x_{11}} & \frac{\partial f_{11}}{\partial x_{12}} & \frac{\partial f_{12}}{\partial x_{11}} & \frac{\partial f_{12}}{\partial x_{12}} & \frac{\partial f_{13}}{\partial x_{11}} & \frac{\partial f_{13}}{\partial x_{12}} & \frac{\partial f_{14}}{\partial x_{11}} & \frac{\partial f_{14}}{\partial x_{12}} \\
\frac{\partial f_{11}}{\partial x_{21}} & \frac{\partial f_{11}}{\partial x_{22}} & \frac{\partial f_{12}}{\partial x_{21}} & \frac{\partial f_{12}}{\partial x_{22}} & \frac{\partial f_{13}}{\partial x_{21}} & \frac{\partial f_{13}}{\partial x_{22}} & \frac{\partial f_{14}}{\partial x_{21}} & \frac{\partial f_{14}}{\partial x_{22}} \\
\frac{\partial f_{11}}{\partial x_{31}} & \frac{\partial f_{11}}{\partial x_{32}} & \frac{\partial f_{12}}{\partial x_{31}} & \frac{\partial f_{12}}{\partial x_{32}} & \frac{\partial f_{13}}{\partial x_{31}} & \frac{\partial f_{13}}{\partial x_{32}} & \frac{\partial f_{14}}{\partial x_{31}} & \frac{\partial f_{14}}{\partial x_{32}} \\
\hdashline
\frac{\partial f_{21}}{\partial x_{11}} & \frac{\partial f_{21}}{\partial x_{12}} & \frac{\partial f_{22}}{\partial x_{11}} & \frac{\partial f_{22}}{\partial x_{12}} & \frac{\partial f_{23}}{\partial x_{11}} & \frac{\partial f_{23}}{\partial x_{12}} & \frac{\partial f_{24}}{\partial x_{11}} & \frac{\partial f_{24}}{\partial x_{12}} \\
\frac{\partial f_{21}}{\partial x_{21}} & \frac{\partial f_{21}}{\partial x_{22}} & \frac{\partial f_{22}}{\partial x_{21}} & \frac{\partial f_{22}}{\partial x_{22}} & \frac{\partial f_{23}}{\partial x_{21}} & \frac{\partial f_{23}}{\partial x_{22}} & \frac{\partial f_{24}}{\partial x_{21}} & \frac{\partial f_{24}}{\partial x_{22}} \\
\frac{\partial f_{21}}{\partial x_{31}} & \frac{\partial f_{21}}{\partial x_{32}} & \frac{\partial f_{22}}{\partial x_{31}} & \frac{\partial f_{22}}{\partial x_{32}} & \frac{\partial f_{23}}{\partial x_{31}} & \frac{\partial f_{23}}{\partial x_{32}} & \frac{\partial f_{24}}{\partial x_{31}} & \frac{\partial f_{24}}{\partial x_{32}}
\end{array}\right]$$

There is some criticism mooted against the way the partial derivatives are strung out in $\partial F/\partial X$. Suppose the matrix function is the identity function, i.e., $F(X) = X$, or equivalently, $f_{ij}(X) = x_{ij}$ for all i and j. If we want to use the matrix of partial derivatives to build the Jacobian of the transformation, the entity $\partial F/\partial X$ is in for a disappointment. Suppose X is of order 2×3 and $F(X) = X$. Then $\partial f/\partial X = ((\text{vec}I_2) \otimes I_3)'$, which is of order 4×9. It is clear that the rank of the matrix $\partial F/\partial X$ is one. The Jacobian of the transformation $F(X) = X$ is I_0. The derivative $\partial F/\partial X$ is nowhere near the Jacobian. Even the order of the matrix $\partial F/\partial X$ is wrong for the Jacobian. To ameliorate the standard definition of the matrix derivative (6.5.1) to meet the needs of the Jacobian, one could define the matrix derivative ${}^*\partial F/\partial X$ of F of order $p \times q$ with respect to X of order $m \times n$ as

$$\frac{{}^*\partial F}{\partial X} = \frac{\partial \text{vec} F}{\partial (\text{vec} X)'}, \tag{6.5.2}$$

which is of order $pq \times mn$. In order to work out the new matrix of partial derivatives, to begin with, one has to stack the variables in X column by column in one long vector, takes it transpose, stack the entries of F column by column into one long column vector, and then take the partial derivatives of each and every entry with respect to $(\text{vec}X)'$. Note that the order of $\text{vec}F$ is $pq \times 1$ and that of $(\text{vec}X)' = 1 \times mn$. Consequently, the order of the matrix (6.5.2) is $pq \times mn$. For example, the case of F with order 2×4 and X of order 3×2 gives rise to the following matrix derivative in its new incarnation:

$$\frac{{}^*\partial F}{\partial X} = \begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \frac{\partial f_{11}}{\partial x_{21}} & \frac{\partial f_{11}}{\partial x_{31}} & \frac{\partial f_{11}}{\partial x_{12}} & \frac{\partial f_{11}}{\partial x_{22}} & \frac{\partial f_{11}}{\partial x_{32}} \\ \frac{\partial f_{21}}{\partial x_{11}} & \frac{\partial f_{21}}{\partial x_{21}} & \frac{\partial f_{21}}{\partial x_{31}} & \frac{\partial f_{21}}{\partial x_{12}} & \frac{\partial f_{21}}{\partial x_{22}} & \frac{\partial f_{21}}{\partial x_{32}} \\ \frac{\partial f_{12}}{\partial x_{11}} & \frac{\partial f_{12}}{\partial x_{21}} & \frac{\partial f_{12}}{\partial x_{31}} & \frac{\partial f_{12}}{\partial x_{12}} & \frac{\partial f_{12}}{\partial x_{22}} & \frac{\partial f_{12}}{\partial x_{32}} \\ \frac{\partial f_{22}}{\partial x_{11}} & \frac{\partial f_{22}}{\partial x_{21}} & \frac{\partial f_{22}}{\partial x_{31}} & \frac{\partial f_{22}}{\partial x_{12}} & \frac{\partial f_{22}}{\partial x_{22}} & \frac{\partial f_{22}}{\partial x_{32}} \\ \frac{\partial f_{13}}{\partial x_{11}} & \frac{\partial f_{13}}{\partial x_{21}} & \frac{\partial f_{13}}{\partial x_{31}} & \frac{\partial f_{13}}{\partial x_{12}} & \frac{\partial f_{13}}{\partial x_{22}} & \frac{\partial f_{13}}{\partial x_{32}} \\ \frac{\partial f_{23}}{\partial x_{11}} & \frac{\partial f_{23}}{\partial x_{21}} & \frac{\partial f_{23}}{\partial x_{31}} & \frac{\partial f_{23}}{\partial x_{12}} & \frac{\partial f_{23}}{\partial x_{22}} & \frac{\partial f_{23}}{\partial x_{32}} \\ \frac{\partial f_{14}}{\partial x_{11}} & \frac{\partial f_{14}}{\partial x_{21}} & \frac{\partial f_{14}}{\partial x_{31}} & \frac{\partial f_{14}}{\partial x_{12}} & \frac{\partial f_{14}}{\partial x_{22}} & \frac{\partial f_{14}}{\partial x_{32}} \\ \frac{\partial f_{24}}{\partial x_{11}} & \frac{\partial f_{24}}{\partial x_{21}} & \frac{\partial f_{24}}{\partial x_{31}} & \frac{\partial f_{24}}{\partial x_{12}} & \frac{\partial f_{24}}{\partial x_{22}} & \frac{\partial f_{24}}{\partial x_{32}} \end{bmatrix}$$

As one can see, the partial derivatives in ${}^*\partial F/\partial X$ are set out in the style of evaluating the Jacobian of a transformation. The entries of ${}^*\partial F/\partial X$ are simply a rearrangement of the entries of $\partial F/\partial X$. More precisely, in the special case, first we form transposes of vecs of each partition of

$\partial F/\partial X$ as

$$\begin{bmatrix} \left(\operatorname{vec}\frac{\partial f_{11}}{\partial X}\right)' & \left(\operatorname{vec}\frac{\partial f_{12}}{\partial X}\right)' & \left(\operatorname{vec}\frac{\partial f_{13}}{\partial X}\right)' & \left(\operatorname{vec}\frac{\partial f_{14}}{\partial X}\right)' \\ \left(\operatorname{vec}\frac{\partial f_{21}}{\partial X}\right)' & \left(\operatorname{vec}\frac{\partial f_{22}}{\partial X}\right)' & \left(\operatorname{vec}\frac{\partial f_{23}}{\partial X}\right)' & \left(\operatorname{vec}\frac{\partial f_{24}}{\partial X}\right)' \end{bmatrix},$$

and then treating each vec as a single entity arrange them in vec form in order to obtain ${}^*\partial F/\partial X$. For a more precise relationship, see Complement 6.5.3.

There is a minor conflict between the standard practice of writing the vector derivative and the version (6.5.2) in the case of a scalar valued function f of a vector variable x. On one hand, $\partial f/\partial x$ is a column vector and on the other hand, ${}^*\partial f/\partial x = \partial f/\partial x' = (\partial f/\partial x)'$, which is a row vector. When we provide a list of some derivatives of some standard functions, we follow the standard form of arranging the partial derivatives. The formulas for the modified form can be jotted down in a simple manner.

A critical result which is useful in deriving some formulas for matrix derivatives is the following. Let f be scalar valued function of a matrix variable X of order $m \times n$. Let Y be a constant matrix of order $m \times n$. Assume that f is differentiable, i.e., all its partial derivatives exist and are continuous. Then the directional derivative of f in the direction of Y as defined by (6.5.3) exists and

$$\lim_{t\to 0}\frac{f(X+tY)-f(X)}{t} = \operatorname{tr}(Y'\frac{\partial f}{\partial X}). \tag{6.5.3}$$

In some problems, it may be relatively easy to evaluate the limit on the left-hand side of (6.5.3). Once we know what it is, $\partial f/\partial X$ can be figured out from (6.5.3). As an example, consider the following problem. Let A be a matrix of order $m \times m$ and for x in R^m, let $f(x) = x'Ax$. Observe that for any constant vector y,

$$\begin{aligned}\lim_{t\to 0}\frac{f(x+ty)-f(x)}{t} &= \lim_{t\to 0}\frac{x'Ax + t^2y'Ay + ty'Ax + tx'Ay - x'Ax}{t} \\ &= y'Ax + x'Ay = y'(A+A')x = \operatorname{tr}(y'\frac{\partial f}{\partial x}).\end{aligned}$$

But $\mathrm{tr}(y'(\partial f/\partial x)) = y'(\partial f/\partial x)$. Hence $\partial f/\partial x = (A + A')x$. This can also be obtained by a straightforward evaluation of the vector derivative.

The formula (6.5.3) is also useful in deriving some identities involving matrix derivatives. Some of them are jotted down below.

P 6.5.1 Let f and g be two differentiable real valued functions of a matrix variable X. Then the following are valid:

(1) $\frac{\partial fg}{\partial X} = f\frac{\partial g}{\partial X} + g\frac{\partial f}{\partial X}$.

(2) $\frac{\partial (f/g)}{\partial X} = \frac{1}{g}\frac{\partial f}{\partial X} - \frac{f}{g^2}\frac{\partial g}{\partial X}$ provided g is not zero.

(3) For a scalar valued function f of a matrix valued function $H = (h_{ij})$ of a matrix variable X,

$$\frac{\partial f(H)}{\partial H} = \sum_i \sum_j \frac{\partial f}{\partial h_{ij}} \frac{\partial h_{ij}}{\partial X}.$$

We now focus on vector derivatives. All our functions f are real valued functions defined on the vector space R^m.

P 6.5.2 (1) If $f(x) = a'x$ for some constant vector $a \in R^m$, then $\frac{\partial f}{\partial x} = a$.

(2) If $f(x) = x'x$, then $\frac{\partial f}{\partial x} = 2x$.

(3) If $f(x) = x'Ax$ for some constant matrix A of order $m \times m$ with real entries, then $\frac{\partial f}{\partial x} = (A + A')x$.

(4) If $f(x) = x'Ax$ for some constant symmetric matrix A of order $m \times m$ with real entries, then $\frac{\partial f}{\partial x} = 2Ax$.

We seek two important applications of these results. Let A be a non-negative definite matrix of order $m \times m$, B a matrix of order $r \times m$, and p a column vector of order $r \times 1$, all with constant real entries. The objective is to minimize the function f given by $f(x) = x'Ax$, $x \in R^m$, subject to the restriction that $Bx = p$. Introduce the vector λ of Lagrange multipliers of order $r \times 1$ and consider the function

$$L(x, \lambda) = x'Ax + 2\lambda'(Bx - p),\ x \in R^m, \lambda \in R^r.$$

The stationary values of the function L are obtained by setting separately the vector derivatives of L with respect to x and λ equal to zero. Using **P 6.5.2**, we have

$$\frac{\partial L}{\partial x} = 2Ax + 2B'\lambda = 0, \frac{\partial L}{\partial \lambda} = 2(Bx - p) = 0.$$

These equations which are linear in x and λ can be rewritten as

$$\begin{bmatrix} A & B' \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ p \end{bmatrix}.$$

Solving these equations is another story. If $\text{rank}(B) = r$ and A is positive definite, then the system of equations admits a unique solution. From the equations, the following series of computations follow:

$$x = -A^{-1}B'\lambda, \; Bx = -BA^{-1}B'\lambda = p;$$
$$\lambda = -(BA^{-1}B')^{-1}p, \; x = A^{-1}B'(BA^{-1}B')^{-1}p.$$

This type of optimization problem arises in Linear Models. Suppose Y is a random vector of m components whose distribution could be any one of the distributions indexed by a finite dimensional parameter $\theta \in R^r$. Suppose under each $\theta \in R^r$, Y has the same dispersion matrix A but the expected value is given by $E_\theta Y = B'\theta$ for some known matrix B of order $r \times m$. (The expected value of each component of Y is a known linear combination of the components of θ.) One of the important problems in Linear Models is to estimate a linear function $p'\theta$ of the parameter vector unbiasedly with minimum variance, where the vector p of order $r \times 1$ is known. In order to make the estimation problem simple, we seek only linear functions of the data Y which are unbiased estimators of $p'\theta$ and in this collection of estimators we search for one with minimum variance. One can show that a linear function $x'Y$ of the data Y is unbiased for $p'\theta$ if $Bx = p$. For such x, the variance of $x'Y$ is $x'Ax$. Now the objective is to minimize $x'Ax$ over all x but subject to the condition $Bx = p$. If B is of rank r and A is of full rank, then the linear unbiased estimator of $p'\theta$ with minimum variance (Best Linear Unbiased Estimator with the acronym BLUE) is given by

$$x'Y = p'(BA^{-1}B')^{-1}BA^{-1}Y.$$

Let us look at another optimization problem. Let A be a symmetric and B a positive definite matrix with real entries. Let $f(x) = x'Ax$ and $g(x) = x'Bx$, $x \in R^m$. We would like to determine the stationary values of the function $f(x)/g(x)$, $x \in R^m - \{0\}$. We equate the vector derivative of this ratio with respect to x to zero. Using **P 6.5.2**, we have

$$\frac{\partial(f/g)}{\partial x} = \frac{2}{x'Bx}Ax - \frac{2x'Ax}{(x'Bx)^2}Bx = 0.$$

This equation leads to the equation

$$(A - \lambda B)x = 0,$$

where $\lambda = x'Ax/x'Bx$. Thus the stationary value x in $R^m - \{0\}$ of the ratio of quadratic forms has to satisfy the equation $(A - \lambda B)x = 0$ for some λ. (But λ will be automatically equal to $x'Ax/x'Bx$. Why?) A non-zero solution to the equation exists if the determinant $|A - \lambda B| = 0$. This determinantal equations has exactly m roots. Thus the stationary values of the ratio of the quadratic forms of interest are at most m in number.

We now focus on matrix derivatives. The function f is a real valued function of a matrix variable X of order $m \times m$. The domain of definition of f need not be the space of all matrices. For the determinant function, we will consider the collection $\mathbf{M}_m(ns)$ of all non-singular matrices of order $m \times m$ with real entries. This set is an open subset of the collection of all matrices of order $m \times m$ in its usual topology. The set $\{X \in \mathbf{M}_m(ns) : |X| > 0\}$ is also an open set and we will consider functions having this set as their domain. Differentiability of the determinant function $|X|$ of X in its domain should pose no problems.

P 6.5.3 (1) If $f(X) = |X|, X \in \mathbf{M}_m(ns)$, then $\partial f/\partial X = |X|(X^{-1})'$.
(2) If $f(X) = \log|X|, |X| > 0$, then $\frac{\partial f}{\partial X} = (X^{-1})'$.
(3) If $f(X) = |X|^r, |X| > 0$ fixed, then $\partial f/\partial X = r|X|^r|(X^{-1})'$.

PROOF. (1) We use (6.5.3). Let $X = (x_{ij}) \in \mathbf{M}_m(ns)$. Let $Y = (y_{ij})$ be any arbitrary matrix of order $m \times m$. For small values of t, $X+tY$ will be non-singular. Let us embark on finding the determinant of $X + tY$. Let $X^c = (x^{ij})$ be the matrix of cofactors of X. After expanding $|X + tY|$ and omitting terms of the order t^2, we have

$$|X + tY| = |X| + t\sum_{i=1}^{m}\sum_{j=1}^{m} y_{ij}x^{ij} = |X| + t\ \mathrm{tr}(Y'X^c).$$

Consequently,

$$\lim_{t\to 0}\frac{|X+tY|-|X|}{t}=\operatorname{tr}(Y'X^c)=\operatorname{tr}(Y'\frac{\partial f}{\partial X})$$
$$\text{i.e., } \frac{\partial f}{\partial X}=X^c=|X|(X^{-1})'.$$

This completes the proof. The proofs of (2) and (3) are now trivial.

The case of symmetric matrices requires some caution. The space $\mathbf{M}_m(s)$ of all symmetric matrices of order $m\times m$ is no longer an m^2-dimensional vector space. In fact, it is an $m(m+1)/2$-dimensional vector space. Now we consider the subset $\mathbf{M}_m(s,ns)$ of all non-singular matrices in $\mathbf{M}_m(s)$. [The letters ns stand for nonsingular.] This subset is an open set in $\mathbf{M}_m(s)$ in its usual topology. The determinant function on this subset is under focus. As a simple example, consider the case of $m=2$. Any matrix X in $\mathbf{M}_m(s,ns)$ is of the form,

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

with the determinant $x_{11}x_{22}-x_{12}^2\neq 0$. Observe that

$$\frac{\partial|X|}{\partial X}=\begin{bmatrix}\frac{\partial|X|}{\partial x_{11}} & \frac{\partial|X|}{\partial x_{12}}\\ \frac{\partial|X|}{\partial x_{12}} & \frac{\partial|X|}{\partial x_{22}}\end{bmatrix}=\begin{bmatrix} x_{22} & -2x_{12}\\ -2x_{12} & x_{11}\end{bmatrix}$$

$$=2\begin{bmatrix} x_{22} & -x_{12}\\ -x_{12} & x_{11}\end{bmatrix}-\begin{bmatrix} x_{22} & 0\\ 0 & x_{11}\end{bmatrix}=|X|[2X^{-1}-\operatorname{diag}(X^{-1})].$$

This formula holds in general too. Before we jot it down let us discuss the problem of taking derivatives of functions whose domain of definition is the set of all symmetric matrices.

Let f be a scalar valued function of a matrix variable X. It is clear that $\partial f/\partial X'=(\partial f/\partial X)'$. Let f be a scalar valued function of a matrix variable X, where X is symmetric. What we need is a formula analogous to (6.5.3) operational in the case of a symmetric argument. We do have a direct formula which in conjunction with (6.5.3) can be used to solve the symmetric problem. The formula for X symmetric is

$$\frac{\partial f}{\partial X}=\left\{\frac{\partial f(Y)}{\partial Y}+\frac{\partial f(Y)}{\partial Y'}-\operatorname{diag}\left(\frac{\partial f(Y)}{\partial Y}\right)\right\}|_{Y=X}. \tag{6.5.4}$$

In working out the derivative $\partial f(Y)/\partial Y$, the function $f(\cdot)$ is pretended to have been defined on the class of all matrices Y, i.e., all the components of Y are regarded as independent variables, and then the derivative formed. Let us illustrate the mechanics of this formula with a simple example. Let $f(X) = |X|$, where X is of order 2×2, $|X| \neq 0$, and X symmetric. Regard $f(\cdot)$ as a function of $Y = (y_{ij})$, where Y is of order 2×2 and $|Y| \neq 0$. More precisely, $f(Y) = |Y| = y_{11}y_{22} - y_{12}y_{21}$. Note that

$$\frac{\partial f(Y)}{\partial Y} = \begin{bmatrix} \frac{\partial |Y|}{\partial y_{11}} & \frac{\partial |Y|}{\partial y_{12}} \\ \frac{\partial |Y|}{\partial y_{21}} & \frac{\partial |Y|}{\partial y_{22}} \end{bmatrix} = \begin{bmatrix} y_{22} & -y_{21} \\ -y_{12} & y_{11} \end{bmatrix},$$

$$\frac{\partial f(Y)}{\partial Y'} = \begin{bmatrix} \frac{\partial |Y|}{\partial y_{11}} & \frac{\partial |Y|}{\partial y_{21}} \\ \frac{\partial |Y|}{\partial y_{12}} & \frac{\partial |Y|}{\partial y_{22}} \end{bmatrix} = \begin{bmatrix} y_{22} & -y_{12} \\ -y_{21} & y_{11} \end{bmatrix},$$

$$\operatorname{diag}\left(\frac{\partial f(Y)}{\partial Y}\right) = \begin{bmatrix} \frac{\partial |Y|}{\partial y_{11}} & 0 \\ 0 & \frac{\partial |Y|}{\partial y_{22}} \end{bmatrix} = \begin{bmatrix} y_{22} & 0 \\ 0 & y_{11} \end{bmatrix},$$

and for X symmetric,

$$\frac{\partial f}{\partial X} = \left\{\frac{\partial f(Y)}{\partial Y} + \frac{\partial f(Y)}{\partial Y'} - \operatorname{diag}\left(\frac{\partial f(Y)}{\partial Y}\right)\right\}\Big|_{Y=X}$$

$$= \begin{bmatrix} x_{22} & -x_{12} \\ -x_{12} & x_{11} \end{bmatrix} + \begin{bmatrix} x_{22} & -x_{12} \\ -x_{12} & x_{11} \end{bmatrix} - \begin{bmatrix} x_{22} & 0 \\ 0 & x_{11} \end{bmatrix}$$

$$= |X|[2X^{-1} - \operatorname{diag}(X^{-1})].$$

P 6.5.4 (1) If $f(X) = |X|$, $X \in \mathbf{M}_m(s, ns)$, then

$$\frac{\partial f}{\partial X} = |X|[2X^{-1} - \operatorname{diag}(X^{-1})].$$

(2) If $f(X) = \log|X|,\ X \in \mathbf{M}_m(s, ns), |X| > 0$, then

$$\frac{\partial f}{\partial X} = [2X^{-1} - \operatorname{diag}(X^{-1})].$$

(3) If $f(X) = |X|^r,\ X \in \mathbf{M}_m(s, ns), |X| > 0$, then

$$\frac{\partial f}{\partial X} = r|X|^r[2X^{-1} - \operatorname{diag}(X^{-1})].$$

We will now outline some useful formulas on matrix derivatives. Let U and V be two matrix functions of a matrix variable X, where $U = (u_{ij})$ and $V = (v_{ij})$ are of orders $p \times q$ and X is of order $m \times n$. Applying **P 6.5.1** (1) to each term $u_{ij}(X)v_{ji}(X)$, we deduce

$$\begin{aligned} &\frac{\partial}{\partial X}\operatorname{tr}(U(X)V(X)) \\ &= \frac{\partial}{\partial X}\operatorname{tr}(U(X)V(Y))|_{Y=X} + \frac{\partial}{\partial X}\operatorname{tr}(U(Y)V(X))|_{Y=X}. \end{aligned} \tag{6.5.5}$$

Instead of the trace function dealt in (6.5.5), we could deal with any scalar valued function f of $U(X)$ and $V(X)$. Accordingly, we have

$$\begin{aligned} \frac{\partial}{\partial X} f(U(X), V(X)) = &\left[\frac{\partial}{\partial U} f(U, V)\right]\left[\frac{\partial}{\partial X} U(X)\right] \\ &+ \left[\frac{\partial}{\partial V} f(U, V)\right]\left[\frac{\partial}{\partial X} V(X)\right]. \end{aligned} \tag{6.5.6}$$

Using (6.5.5) or (6.5.6), one can establish the validity of the following proposition.

P 6.5.5 (1) Let $U(X)$ be a matrix valued function of a matrix variable X, where $U(X)$ is of order $p \times p$, non-singular, and X is of order $m \times n$. Then

$$\frac{\partial}{\partial X}\operatorname{tr}(U^{-1}(X)) = -\frac{\partial}{\partial X}\operatorname{tr}(U^{-2}(Y)U(X))|_{Y=X}.$$

(2) Let A be a constant matrix of order $p \times p$ and $U(X)$ a matrix valued function of a matrix argument X, where $U(X)$ is of order $p \times p$, non-singular, and X is of order $m \times n$. Then

$$\frac{\partial}{\partial X}\operatorname{tr}(U^{-1}(X)A) = -\frac{\partial}{\partial X}\operatorname{tr}(U^{-1}(Y)AU^{-1}(Y)U(X))|_{Y=X}.$$

(3) Let A and B be constant matrices each of order $m \times m$ and $f(X) = \text{tr}(AX^{-1}B)$, $X \in \mathbf{M}_m(ns)$. Then

$$\frac{\partial f}{\partial X} = -(X^{-1}BAX^{-1})'.$$

(4) Let $U(X)$ be a matrix valued function of a matrix variable X, where $U(X)$ is of order $p \times p$, non-singular, and X is of order $m \times n$. Then

$$\frac{\partial}{\partial X}|U(X)| = |U(X)|\frac{\partial}{\partial X}\text{tr}(U^{-1}(Y)U(X))|_{Y=X}.$$

(5) Let A be a constant matrix of order $m \times m$ and $f(X) = |AX|$, X is of order $m \times m$ and AX non-singular. Then

$$\begin{aligned}\frac{\partial f}{\partial X} &= |AX|\frac{\partial}{\partial X}\text{tr}((AY)^{-1}AX)|_{Y=X} \\ &= |AX|((AX)^{-1}A)'.\end{aligned}$$

At the beginning of this section, we toyed with another idea of writing the matrix of partial derivatives. More precisely, let $F(X)$ be a matrix valued function of a matrix variable X. We defined

$$\frac{{}^*\partial F}{\partial X} = \frac{\partial \text{vec}(F)}{\partial (\text{vec}X)'}.$$

Even though the entries of ${}^*\partial F/\partial X$ are simply a rearrangement of the entries of $\partial F/\partial X$, it is useful to compile ${}^*\partial F/\partial X$ for some standard functions F of X. This is what we do in the following proposition. All these results can be derived from first principles.

P 6.5.6 (1) Let $F(X) = AX$, where A is a constant matrix of order $p \times m$ and X of order $m \times n$. Then

$$\frac{{}^*\partial F}{\partial X} = I_n \otimes A.$$

(2) Let $F(X) = XB$, where B is a constant matrix of order $n \times q$ and X of order $m \times n$. Then

$$\frac{{}^*\partial F}{\partial X} = B' \otimes I_m.$$

(3) Let $F(X) = AXB$, where A and B are constant matrices of orders $p \times m$ and $n \times q$, respectively, and X of order $m \times n$. Then

$$\frac{{}^*\partial F}{\partial X} = B' \otimes A.$$

(4) Let $F(X) = AX'B$, where A and B are constant matrices of orders $p \times n$ and $m \times q$, respectively, and X of order $m \times n$. Then

$$\frac{{}^*\partial F}{\partial X} = (A \otimes B')P,$$

where P is the permutation matrix which transforms the vector $\text{vec}(X)$ into $\text{vec}(X')$, i.e., $\text{vec}(X') = P\text{vec}(X)$.

(5) Let $U(X)$ and $V(X)$ be matrix valued functions of a matrix variable X, where $U(X)$ is of order $p \times q$, $V(X)$ of order $q \times r$, and X of order $m \times n$. Then

$$\frac{{}^*\partial}{\partial X} U(X)V(X) = (V(X) \otimes I_r)' \frac{{}^*\partial}{\partial X} U(X) + (I \otimes U(X)) \frac{{}^*\partial}{\partial X} V(X).$$

(6) Let $F(X) = X'AX$, where A is a constant matrix of order $m \times m$ and X of order $m \times n$. Then

$$\frac{{}^*\partial F}{\partial X} = (X'A' \otimes I_n)P + (I_n \otimes X'A).$$

(7) Let $F(X) = AX^{-1}B$, where A and B are constant matrices of orders $p \times m$ and $m \times q$, respectively, X of order $m \times m$ and non-singular. Then

$$\frac{{}^*\partial F}{\partial X} = -(X^{-1}B)' \otimes (AX^{-1}).$$

(8) Let $U(X)$ and $Z(X)$ be two matrix valued functions of a matrix variable X, where $U(\cdot)$ is of order $p \times q$, $Z(\cdot)$ of order 1×1 and X of order $m \times n$. Let $f(X) = Z(X)U(X)$. Then

$$\frac{{}^*\partial f}{\partial X} = \text{vec}(U(X)) \frac{{}^*\partial Z(X)}{\partial X} + Z(X) \frac{{}^*\partial U(X)}{\partial X}.$$

(9) Let $U(X)$ be a matrix valued function of a matrix variable X, where $U(\cdot)$ is of order $p \times p$ and non-singular, and X is of order $m \times n$. Let $f(X) = [U(X)]^{-1}$. Then

$$\frac{{}^*\partial f}{\partial X} = ((U^{-1}(X))' \otimes U^{-1}(X))\frac{{}^*\partial U(X)}{\partial X}.$$

(10) Let $Y(X)$ be a matrix valued function of a matrix variable X, where $Y(\cdot)$ is of order $p \times q$ and X of order $m \times n$. Let $Z(V)$ be a matrix valued function of a matrix variable V, where $Z(\cdot)$ is of order $r \times s$ and V of order $p \times q$. Let $f(X) = Z(Y(X)),\ X \in \mathbf{M}_{m,n}$. Then

$$\frac{{}^*\partial f}{\partial X} = \left(\left.\frac{{}^*\partial Z(V)}{\partial V}\right|_{V=Y(X)}\right)\left(\frac{{}^*\partial Y(X)}{\partial X}\right).$$

(11) Let $Z(X)$ and $Y(X)$ be two matrix valued functions of a matrix variable X, where $Z(X)$ and $Y(X)$ are of the same order $p \times q$ and X of order $m \times n$. Let $f(X) = Z(X) \cdot Y(X),\ X \in \mathbf{M}_{m,n}$, where the symbol $\cdot$ denotes HS multiplication. Then

$$\frac{{}^*\partial f}{\partial X} = D(Z(X))\frac{{}^*\partial Y(X)}{\partial X} + D(Y(X))\frac{{}^*\partial Z(X)}{\partial X},$$

where for any matrix $Z = (z_{ij})$ of order $p \times q$,

$D(Z) = \text{diag}(z_{11}, z_{12}, \ldots, z_{1q}, z_{21}, z_{22}, \ldots, z_{2q}, \ldots, z_{p1}, z_{p2}, \ldots, z_{pq})$.

(12) Let $Z(X)$ be a matrix valued function of a matrix variable X, where $Z(X)$ is of order $p \times q$ and X of order $m \times n$. Let B be a constant matrix of order $p \times q$ and $f(X) = Z(X) \cdot B,\ X \in \mathbf{S}_{mn}$. Then

$$\frac{{}^*\partial f}{\partial X} = D(B)\frac{{}^*\partial Z(X)}{\partial X}.$$

As has been indicated earlier, the matrix derivative defined as ${}^*\partial f/\partial X$ is very useful in evaluating the Jacobian of a transformation. Suppose $f(X)$ is a matrix valued function of a matrix variable X, where both X and $f(X)$ are of the same order $m \times n$. The Jacobian J of the transformation $f(\cdot)$ is given by

$$J = \left|\frac{{}^*\partial f}{\partial X}\right|_+,$$

where the suffix + indicates the positive value of the determinant of the matrix ${}^*\partial f/\partial X$ of order $mn \times mn$. Suppose $f(X) = AXB$, where A and B are constant non-singular matrices of orders $m \times m$ and $n \times n$, respectively, and $X \in \mathbf{M}_{m,n}$. The Jacobian of the transformation $f(\cdot)$ is given by

$$J = \left| \frac{{}^*\partial f}{\partial X} \right|_+ = |B' \otimes A|_+ = |A|_+^n |B|_+^m.$$

Complements

6.5.1 Let $F(X) = X$ be the identity transformation of the matrix variable X of order $m \times n$. Show that $\partial F/\partial X = (\text{vec}(I_m)) \otimes (\text{vec}(I_n))'$.
6.5.2 Let $F(X) = X$ be the identity transformation of the matrix variable X of order $m \times n$. Show that ${}^*\partial F/\partial X = I$.
6.5.3 Let F be a matrix valued function of order $p \times q$ of a matrix variable $X = (x_{ij})$ of order $m \times n$. Show that

$$\frac{{}^*\partial F}{\partial X} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\text{vec} \frac{\partial F}{\partial x_{ij}} \right) \left(\text{vec} E_{ij} \right)',$$

where E_{ij} is a matrix of order $m \times n$ whose (i,j)-th entry is unity and the rest of its entries zeros.
6.5.4 Let A be a constant matrix of order $m \times n$ and $f(X) = \text{tr}(AX)$, $X \in \mathbf{M}_{m,n}$, the vector space of all matrices X of order $n \times m$. Show that $(\partial f/\partial X) = A'$. If $m = n$ and the domain of definition of f is the collection $\mathbf{M}(s)$ of all $m \times m$ symmetric matrices, show that $\partial f/\partial X = 2A' - \text{diag}(A)$.
6.5.5 Let $f(X) = \text{tr}(X^2)$, $X \in \mathbf{M}_m$, the space of all $m \times m$ matrices. Show that $(\partial f/\partial X) = 2X'$. If the domain of definition of f is the collection of all symmetric matrices, how does the matrix of partial derivatives change?
6.5.6 Let A and B be two constant matrices of orders $m \times m$ and $n \times n$, respectively. Let $f(X) = \text{tr}(X'AXB)$, $X \in \mathbf{M}_{m,n}$, the space of all $m \times n$ matrices, Show that $(\partial f/\partial X) = AXB + A'XB'$. If $m = n$ and the domain of definition of f is the space of all $m \times m$ matrices, show that

$$\frac{\partial f}{\partial X} = AXB + A'XB' + BXA + B'XA' - \text{diag}(AXB + A'XB').$$

6.5.7 Let A and B be two constant matrices of the same order $m \times m$ and $f(X) = \text{tr}(XAXB)$, $X \in \mathbf{S}_m$. Show that

$$\frac{\partial f}{\partial X} = A'X'B' + B'X'A'.$$

If the domain of definition of f is the space of all symmetric matrices, show that

$$\frac{\partial f}{\partial X} = A'XB' + B'XA' + AXB + BXA - \text{diag}(A'XB' + B'XA').$$

6.5.8 Let A be a constant matrix of order $m \times m$ and $f(X) = \text{tr}(X'AX)$, $X \in \mathbf{M}_{m,n}$. Show that $(\partial f/\partial X) = (A + A')X$. If $m = n$ and the domain of definition of f is the collection of all symmetric matrices, show that

$$\frac{\partial f}{\partial X} = (A + A')X + X(A + A') - \text{diag}((A + A')X).$$

6.5.9 Let $f(X) = \text{tr}(X^n)$, $X \in \mathbf{M}_m$, $n \geq 1$. Show that $(\partial f/\partial X) = nX^{n-1}$. If the domain of definition of f is the space of all symmetric matrices, show that

$$\frac{\partial f}{\partial X} = 2nX^{n-1} - n\text{diag}(X^{n-1}).$$

6.5.10 Let x and y be two fixed column vectors of orders $m \times 1$ and $n \times 1$, and $f(X) = x'Xy$, $X \in \mathbf{M}_{m,n}$. Show that $(\partial f/\partial X) = xy'$. If $m = n$ and the domain of definition of f is the set of all symmetric matrices, show that $(\partial f/\partial X) = xy' + yx'$.

6.5.11 Let A be a constant matrix of order $m \times m$ and $f(X) = \text{tr}(AX^{-1}), X \in \mathbf{M}_m(ns)$, the set of all non-singular matrices of order $m \times m$. Show that

$$\frac{\partial f}{\partial X} = -(X^{-1}AX^{-1})'.$$

If the domain of definition of f is confined to the collection of all non-singular symmetric matrices, show that

$$\frac{\partial f}{\partial X} = -X^{-1}AX^{-1} - X^{-1}A'X^{-1} + \text{diag}(X^{-1}AX^{-1}).$$

6.5.12 Let $f(X) = |XX'|$, $X \in \mathbf{M}_{m,n}$ and $\text{rank}(X) = m$. Show that

$$\frac{\partial f}{\partial X} = 2|XX'|(XX')^{-1}X.$$

6.5.13 Let a and b be two constant column vectors of orders $m\times 1$ and $n\times 1$, respectively. Determine the matrix derivative of each of the scalar valued functions $f_1(X) = a'Xb$, and $f_2(X) = a'XX'a$, $X \in \mathbf{M}_{m,n}$, the collection of all matrices of order $m \times n$, with respect to X.

6.5.14 Let a be a constant column vector of order $m \times 1$ and $f(X) = a'X^{-1}a$, $X \in \mathbf{M}_m(ns)$, the collection of all $m\times m$ non-singular matrices of order $m \times m$. Determine the matrix derivative of the scalar valued function f with respect to X.

6.5.15 Let p be any positive integer and $f(X) = X^p$, $X \in \mathbf{M}_m$. Show that

$$\frac{{}^*\partial f}{\partial X} = \sum_{j=1}^{p}(X')^{p-j} \otimes X^{j-1}.$$

6.5.16 Find the Jacobian of each of the following transformations, where A and B are constant matrices of order $m \times m$, and $X \in \mathbf{M}_m$.

(1) $f(X) = AX^{-1}B$, X non-singular.
(2) $f(X) = XAX'$.
(3) $f(X) = X'AX$.
(4) $f(X) = XAX$, $X \in \mathbf{M}_m$.
(5) $f(X) = X'AX'$.

Notes: The following papers and books have been consulted for developing the material in this chapter. Hartley, Rao, and Kiefer (1969), Rao and Mitra (1971b), Rao (1973c), Srivastava and Khatri (1979), Rao and Kleffe (1980), Graham (1981), Barnett (1990), Rao (1985a), Rao and Kleffe (1988), Magnus and Neudecker (1991), Liu (1995), among others.

CHAPTER 7

PROJECTORS AND IDEMPOTENT OPERATORS

The notion of an orthogonal projection has been introduced in Section 2.2 in the context of inner product spaces. Look up Definition 2.2.11 and the ensuing discussion. In this chapter, we will introduce projectors in the general context of vector spaces. Under a particular mixture of circumstances, an orthogonal projection is seen to be a special kind of projector. We round up the chapter with some examples and complements.

7.1. Projectors

DEFINITION 7.1.1. Let a vector space $\mathbf{V}$ be the direct sum of two subspaces $\mathbf{V}_1$ and $\mathbf{V}_2, \mathbf{V}_1 \cap \mathbf{V}_2 = \{0\}$, i.e., $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2$. (See **P 1.5.5** and the discussion preceding **P 1.5.7.**) Then any vector x in $\mathbf{V}$ has a unique decomposition $x = x_1 + x_2$ with $x_1 \in \mathbf{V}_1$ and $x_2 \in \mathbf{V}_2$. The transformation $x \to x_1$ is called the projection of x on $\mathbf{V}_1$ along $\mathbf{V}_2$. The operator or map P defined on the vector space $\mathbf{V}$ by $Px = x_1$ is called a projector from the vector space $\mathbf{V}$ to the subspace $\mathbf{V}_1$ along the subspace $\mathbf{V}_2$.

The first thing we would like to point out is that the map P is well-defined. Further, the map P is an onto map from $\mathbf{V}$ to $\mathbf{V}_1$. It is also transparent that the projector P restricted to the subspace $\mathbf{V}_1$ is the identity transformation from $\mathbf{V}_1$ to $\mathbf{V}_1$, i.e., $Px = x$ if $x \in \mathbf{V}_1$. If $\mathbf{V}$ is an inner product space and $x \in \mathbf{V}_1, y \in \mathbf{V}_2$ implies that x and y are orthogonal, i.e., $\mathbf{V}_2$ is the orthogonal complement of $\mathbf{V}_1$, the map P is precisely the orthogonal projection from the space $\mathbf{V}$ to the space $\mathbf{V}_1$ as enunciated in Definition 2.2.11. Suppose $\mathbf{V}_1$ is a subspace of $\mathbf{V}$. There could be any number of subspaces $\mathbf{V}_2$ of $\mathbf{V}$ such that $\mathbf{V}_1 \oplus \mathbf{V}_2 = \mathbf{V}$. Each such subspace $\mathbf{V}_2$ gives a projector P from $\mathbf{V}$ onto $\mathbf{V}_1$ along $\mathbf{V}_2$.