Linear Algebra and Linear Models, Second Edition

R.B. Bapat

Springer

Universitext

Editorial Board
(North America):

S. Axler
F.W. Gehring
K.A. Ribet

Springer
New York
Berlin
Heidelberg
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Universitext

Editors (North America): S. Axler, F.W. Gehring, and K.A. Ribet

Aksoy/Khamsi: Nonstandard Methods in Fixed Point Theory
Andersson: Topics in Complex Analysis
Aupetit: A Primer on Spectral Theory
Bachman/Narici/Beckenstein: Fourier and Wavelet Analysis
Balakrishnan/Ranganathan: A Textbook of Graph Theory
Balser: Formal Power Series and Linear Systems of Meromorphic Ordinary Differential Equations
Bapat: Linear Algebra and Linear Models (2nd ed.)
Berberian: Fundamentals of Real Analysis
Booss/Bleecker: Topology and Analysis
Borkar: Probability Theory: An Advanced Course
Böttcher/Silbermann: Introduction to Large Truncated Toeplitz Matrices
Carleson/Gamelin: Complex Dynamics
Cecil: Lie Sphere Geometry: With Applications to Submanifolds
Chae: Lebesgue Integration (2nd ed.)
Charlap: Bieberbach Groups and Flat Manifolds
Chern: Complex Manifolds Without Potential Theory
Cohn: A Classical Invitation to Algebraic Numbers and Class Fields
Curtis: Abstract Linear Algebra
Curtis: Matrix Groups
DiBenedetto: Degenerate Parabolic Equations
Dimca: Singularities and Topology of Hypersurfaces
Edwards: A Formal Background to Mathematics I a/b
Edwards: A Formal Background to Mathematics II a/b
Foulds: Graph Theory Applications
Friedman: Algebraic Surfaces and Holomorphic Vector Bundles
Fuhrmann: A Polynomial Approach to Linear Algebra
Gardiner: A First Course in Group Theory
Gårding/Tambour: Algebra for Computer Science
Goldblatt: Orthogonality and Spacetime Geometry
Gustafson/Rao: Numerical Range: The Field of Values of Linear Operators and Matrices
Hahn: Quadratic Algebras, Clifford Algebras, and Arithmetic Witt Groups
Holmgren: A First Course in Discrete Dynamical Systems
Howe/Tan: Non-Abelian Harmonic Analysis: Applications of $SL(2, \mathrm{R})$
Howes: Modern Analysis and Topology
Hsieh/Sibuya: Basic Theory of Ordinary Differential Equations
Humi/Miller: Second Course in Ordinary Differential Equations
Hurwitz/Kritikos: Lectures on Number Theory
Jennings: Modern Geometry with Applications
Jones/Morris/Pearson: Abstract Algebra and Famous Impossibilities
Kannan/Krueger: Advanced Analysis
Kelly/Matthews: The Non-Euclidean Hyperbolic Plane
Kostrikin: Introduction to Algebra
Luecking/Rubel: Complex Analysis: A Functional Analysis Approach
MacLane/Moerdijk: Sheaves in Geometry and Logic
Marcus: Number Fields
McCarthy: Introduction to Arithmetical Functions

(continued after index)

R.B. Bapat

Linear Algebra and Linear Models

Second Edition

Springer

R.B. Bapat
Indian Statistical Institute
New Delhi 110016
India
rbb@isid.ac.in

Editorial Board
(North America):

S. Axler
Mathematics Department
San Francisco State University
San Francisco, CA 94132
USA

F.W. Gehring
Mathematics Department
East Hall
University of Michigan
Ann Arbor, MI 48109-1109
USA

K.A. Ribet
Mathematics Department
University of California at Berkeley
Berkeley, CA 94720-3840
USA

Mathematics Subject Classification (1991): 15-01, 62H12, 62H15, 62K10

Library of Congress Cataloging-in-Publication Data
Bapat, R.B.
Linear algebra and linear models/R.B. Bapat. — 2nd ed.
p. cm. — (Universitext)
Includes bibliographical references and index.
ISBN 0-387-98871-8 (hc.: alk. paper)
1. Algebras, Linear. 2. Linear models (Statistics) I. Title.
QA184.B37 1999
512′.5—dc21 99-30383

© 2000 Springer-Verlag New York, Inc., and Hindustan Book Agency.
First published in 1993 by Hindustan Book Agency, Delhi, India.
Published outside India by Springer-Verlag in cooperation with HBA, and this edition is not for sale in India.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

ISBN 0-387-98871-8 Springer-Verlag New York Berlin Heidelberg SPIN 10729850

Preface

The main purpose of the present monograph is to provide a rigorous introduction to the basic aspects of the theory of linear estimation and hypothesis testing. The necessary prerequisites in matrices, multivariate normal distribution, and distribution of quadratic forms are developed along the way. The monograph is primarily aimed at advanced undergraduate and first-year master's students taking courses in linear algebra, linear models, multivariate analysis, and design of experiments. It should also be of use to research workers as a source of several standard results and problems.

Some features in which we deviate from the standard textbooks on the subject are as follows.

We deal exclusively with real matrices, and this leads to some nonconventional proofs. One example is the proof of the fact that a symmetric matrix has real eigenvalues. We rely on ranks and determinants a bit more than is done usually. The development in the first two chapters is somewhat different from that in most texts.

It is not the intention to give an extensive introduction to matrix theory. Thus, several standard topics such as various canonical forms and similarity are not found here. We often derive only those results that are explicitly used later. The list of facts in matrix theory that are elementary, elegant, but not covered here is almost endless.

We put a great deal of emphasis on the generalized inverse and its applications. This amounts to avoiding the "geometric" or the "projections" approach that is favored by some authors and taking recourse to a more algebraic approach. Partly as a personal bias, I feel that the geometric approach works well in providing an understanding of why a result should be true but has limitations when it comes to proving the result rigorously.

The first three chapters are devoted to matrix theory, linear estimation, and tests of linear hypotheses, respectively. Chapter 4 collects several results on eigenvalues and singular values that are frequently required in statistics but usually are not proved in statistics texts. This chapter also includes sections on principal components and canonical correlations. Chapter 5 prepares the background for a course in designs, establishing the linear model as the underlying mathematical framework. The sections on optimality may be useful as motivation for further reading in this research area in which there is considerable activity at present. Similarly, the last chapter tries to provide a glimpse into the richness of a topic in generalized inverses (rank additivity) that has many interesting applications as well.

Several exercises are included, some of which are used in subsequent developments. Hints are provided for a few exercises, whereas reference to the original source is given in some other cases.

I am grateful to Professors Aloke Dey, H. Neudecker, K.P.S. Bhaskara Rao, and Dr. N. Eagambaram for their comments on various portions of the manuscript. Thanks are also due to B. Ganeshan for his help in getting the computer printouts at various stages.

About the Second Edition

This is a thoroughly revised and enlarged version of the first edition. Besides correcting the minor mathematical and typographical errors, the following additions have been made:

(1) A few problems have been added at the end of each section in the first four chapters. All the chapters now contain some new exercises.
(2) Complete solutions or hints are provided to several problems and exercises.
(3) Two new sections, one on the "volume of a matrix" and the other on the "star order," have been added.

New Delhi, India R.B. Bapat

Contents

Notation Index

1

Vector Spaces and Matrices

1.1 Preliminaries

In this section we review certain basic concepts. We consider only real matrices. Although our treatment is self-contained, the reader is assumed to be familiar with basic operations on matrices. We also assume knowledge of elementary properties of the determinant.

An $m \times n$ *matrix* consists of mn real numbers arranged in m rows and n columns. We denote matrices by bold letters. The entry in row i and column j of the matrix $\mathbf{A}$ is denoted by a_{ij}. An $m \times 1$ matrix is called a *column vector* of order m; similarly, a $1 \times n$ matrix is a *row vector* of order n. An $m \times n$ matrix is called a *square* matrix if $m = n$.

If $\mathbf{A}, \mathbf{B}$ are $m \times n$ matrices, then $\mathbf{A} + \mathbf{B}$ is defined as the $m \times n$ matrix with (i, j)-entry $a_{ij} + b_{ij}$. If $\mathbf{A}$ is a matrix and c is a real number, then $c\mathbf{A}$ is obtained by multiplying each element of $\mathbf{A}$ by c.

If $\mathbf{A}$ is $m \times p$ and $\mathbf{B}$ is $p \times n$, then their product $\mathbf{C} = \mathbf{AB}$ is an $m \times n$ matrix with (i, j)-entry given by

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

The following properties hold:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}),$$
$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC},$$
$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}.$$

The *transpose* of the $m \times n$ matrix $\mathbf{A}$, denoted by $\mathbf{A}'$, is the $n \times m$ matrix whose (i, j)-entry is a_{ji}. It can be verified that $(\mathbf{A}')' = \mathbf{A}$, $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$, $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

A good understanding of the definition of matrix multiplication is quite useful. We note some simple facts that are often required. We assume that all products occurring here are defined in the sense that the orders of the matrices make them compatible for multiplication.

(i) The jth column of $\mathbf{AB}$ is the same as $\mathbf{A}$ multiplied by the jth column of $\mathbf{B}$.
(ii) The ith row of $\mathbf{AB}$ is the same as the ith row of $\mathbf{A}$ multiplied by $\mathbf{B}$.
(iii) The (i, j)-entry of $\mathbf{ABC}$ is obtained as

$$(x_1, \dots, x_p)\mathbf{B}\begin{bmatrix} y_1 \\ \vdots \\ y_q \end{bmatrix},$$

where $(x_1, \dots, x_p)$ is the ith row of $\mathbf{A}$ and $(y_1, \dots, y_q)'$ is the jth column of $\mathbf{C}$.
(iv) If $\mathbf{A} = [\mathbf{a_1}, \dots, \mathbf{a_n}]$ and

$$\mathbf{B} = \begin{bmatrix} \mathbf{b_1}' \\ \vdots \\ \mathbf{b_n}' \end{bmatrix},$$

where $\mathbf{a_i}$ denote columns of $\mathbf{A}$ and $\mathbf{b_j}'$ denote rows of $\mathbf{B}$, then

$$\mathbf{AB} = \mathbf{a_1}\mathbf{b_1}' + \cdots + \mathbf{a_n}\mathbf{b_n}'.$$

A *diagonal matrix* is a square matrix $\mathbf{A}$ such that $a_{ij} = 0$, $i \neq j$. We denote the diagonal matrix

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

by $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$. When $\lambda_i = 1$ for all i, this matrix reduces to the *identity matrix* of order n, which we denote by $\mathbf{I_n}$, or often simply by $\mathbf{I}$ if the order is clear from the context. Observe that for any square matrix $\mathbf{A}$, we have $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$.

The entries $a_{11}, \dots, a_{nn}$ are said to constitute the (main) diagonal entries of $\mathbf{A}$. The *trace* of $\mathbf{A}$ is defined as

$$\operatorname{trace}\mathbf{A} = a_{11} + \cdots + a_{nn}.$$

It follows from this definition that if $\mathbf{A}, \mathbf{B}$ are matrices such that both $\mathbf{AB}$ and $\mathbf{BA}$ are defined, then

$$\operatorname{trace}\mathbf{AB} = \operatorname{trace}\mathbf{BA}.$$

The *determinant* of an $n \times n$ matrix $\mathbf{A}$, denoted by $|\mathbf{A}|$, is defined as

$$|\mathbf{A}| = \sum_{\sigma} \epsilon(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

where the summation is over all permutations $\{\sigma(1), \ldots, \sigma(n)\}$ of $\{1, \ldots, n\}$ and $\epsilon(\sigma)$ is 1 or -1 according as σ is even or odd.

We state some basic properties of the determinant without proof:

(i) The determinant can be evaluated by expansion along a row or a column. Thus, expanding along the first row,

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} |\mathbf{A_{1j}}|,$$

where $\mathbf{A_{1j}}$ is the submatrix obtained by deleting the first row and the jth column of $\mathbf{A}$. We also note that

$$\sum_{j=1}^{n} (-1)^{1+j} a_{ij} |\mathbf{A_{1j}}| = 0, \qquad i = 2, \ldots, n.$$

(ii) The determinant changes sign if two rows (or columns) are interchanged.

(iii) The determinant is unchanged if a constant multiple of one row is added to another row. A similar property is true for columns.

(iv) The determinant is a linear function of any column (row) when all the other columns (rows) are held fixed.

(v) $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$.

The matrix $\mathbf{A}$ is *upper triangular* if $a_{ij} = 0, i > j$. The transpose of an upper triangular matrix is *lower triangular.*

It will often be necessary to work with matrices in partitioned form. For example, let

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

be two matrices where each $\mathbf{A_{ij}}$, $\mathbf{B_{ij}}$ is itself a matrix. If compatibility for matrix multiplication is assumed throughout (in which case, we say that the matrices are partitioned *conformally*), then we can write

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}.$$

Problems

1. Construct a 3×3 matrix $\mathbf{A}$ such that both $\mathbf{A}$, $\mathbf{A^2}$ are nonzero but $\mathbf{A^3} = \mathbf{0}$.

2. Decide whether the determinant of the following matrix $\mathbf{A}$ is even or odd, without evaluating it explicitly:

$$\mathbf{A} = \begin{bmatrix} 387 & 456 & 589 & 238 \\ 488 & 455 & 677 & 382 \\ 440 & 982 & 654 & 651 \\ 892 & 564 & 786 & 442 \end{bmatrix}.$$

3. Let

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

Can you find 3×3 matrices $\mathbf{X}$, $\mathbf{Y}$ such that $\mathbf{XY} - \mathbf{YX} = \mathbf{A}$?

4. If $\mathbf{A}$, $\mathbf{B}$ are $n \times n$ matrices, show that

$$\begin{vmatrix} \mathbf{A}+\mathbf{B} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} \end{vmatrix} = |\mathbf{A}||\mathbf{B}|.$$

5. Evaluate the determinant of the $n \times n$ matrix $\mathbf{A}$, where $a_{ij} = ij$ if $i \neq j$ and $a_{ij} = 1 + ij$ if $i = j$.

6. Let $\mathbf{A}$ be an $n \times n$ matrix and suppose $\mathbf{A}$ has a zero submatrix of order $r \times s$ where $r + s = n + 1$. Show that $|\mathbf{A}| = 0$.

1.2 Vector Spaces and Subspaces

A nonempty set S is called a *vector space* if it satisfies the following conditions:

(i) For any $\mathbf{x}$, $\mathbf{y}$ in S, $\mathbf{x} + \mathbf{y}$ is defined and is in S. Furthermore,

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \qquad \text{(commutativity)}$$
$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}. \qquad \text{(associativity)}$$

(ii) There exists an element in S, denoted by $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x}$.

(iii) For any $\mathbf{x}$ in S, there exists an element $\mathbf{y}$ in S such that $\mathbf{x} + \mathbf{y} = \mathbf{0}$.

(iv) For any $\mathbf{x}$ in S and any real number c, $c\mathbf{x}$ is defined and is in S; moreover, $1\mathbf{x} = \mathbf{x}$ for any $\mathbf{x}$.

(v) For any $\mathbf{x_1}, \mathbf{x_2}$ in S and real numbers $c_1, c_2,\ c_1(\mathbf{x_1} + \mathbf{x_2}) = c_1\mathbf{x_1} + c_1\mathbf{x_2}$, $(c_1 + c_2)\mathbf{x_1} = c_1\mathbf{x_1} + c_2\mathbf{x_1}$ and $c_1(c_2\mathbf{x_1}) = (c_1c_2)\mathbf{x_1}$.

Elements in S are called *vectors.* If $\mathbf{x}$, $\mathbf{y}$ are vectors, then the operation of taking their sum $\mathbf{x} + \mathbf{y}$ is referred to as vector addition. The vector in (ii) is called the *zero vector.* The operation in (iv) is called *scalar multiplication.* A vector space may be defined with reference to any field. We have taken the field to be the field of real numbers as this will be sufficient for our purpose.

The set of column vectors of order n (or $n \times 1$ matrices) is a vector space. So is the set of row vectors of order n. These two vector spaces are the ones we consider most of the time.

Let R^n denote the set $R \times R \times \cdots \times R$, taken n times, where R is the set of real numbers. We will write elements of R^n either as column vectors or as row vectors depending upon whichever is convenient in a given situation.

If S, T are vector spaces and $S \subset T$, then S is called a *subspace* of T.

Let us describe all possible subspaces of R^3. Clearly, R^3 is a vector space, and so is the space consisting of only the zero vector, i.e., the vector of all zeros. Let c_1, c_2, c_3 be real numbers. The set of all vectors $\mathbf{x} \in R^3$ that satisfy

$$c_1x_1 + c_2x_2 + c_3x_3 = 0$$

is a subspace of R^3 (Here x_1, x_2, x_3 are the coordinates of $\mathbf{x}$). Geometrically, this set represents a plane passing through the origin. Intersection of two distinct planes through the origin is a straight line through the origin and is also a subspace. These are the only possible subspaces of R^3.

Problems

1. Which of the following sets are vector spaces (with the natural operations of addition and scalar multiplication)? (i) Vectors (a, b, c, d) such that $a + 2b = c - d$; (ii) $n \times n$ matrices $\mathbf{A}$ such that $\mathbf{A^2} = \mathbf{I}$; (iii) 3×3 matrices $\mathbf{A}$ such that $a_{11} + a_{13} = a_{22} + a_{31}$.

2. If S and T are vector spaces, then are $S \cup T$ and $S \cap T$ vector spaces as well?

1.3 Basis and Dimension

The *linear span* of (or the space spanned by) the vectors $\mathbf{x_1}, \dots, \mathbf{x_m}$ is defined to be the set of all linear combinations $c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m}$ where $c_1, \dots, c_m$ are real numbers. The linear span is a subspace; this follows from the definition.

A set of vectors $\mathbf{x_1}, \dots, \mathbf{x_m}$ is said to be *linearly dependent* if there exist real numbers $c_1, \dots, c_m$ such that at least one c_i is nonzero and $c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m} = \mathbf{0}$. A set is *linearly independent* if it is not linearly dependent. Strictly speaking, we should refer to a *collection* (or a *multiset*) of vectors rather than a set of vectors in the two preceding definitions. Thus when we talk of vectors $\mathbf{x_1}, \dots, \mathbf{x_m}$ being linearly dependent or independent, we allow for the possibility of the vectors not necessarily being distinct.

The following statements are easily proved:

(i) The set consisting of the zero vector alone is linearly dependent.
(ii) If $X \subset Y$ and if X is linearly dependent, then so is Y.
(iii) If $X \subset Y$ and if Y is linearly independent, then so is X.

A set of vectors is said to form a *basis* for the vector space S if it is linearly independent and its linear span equals S.

Let $\mathbf{e_i}$ be the ith column of the $n \times n$ identity matrix. The set $\mathbf{e_1}, \ldots, \mathbf{e_n}$ forms a basis for R^n, called the *standard basis.*

If $\mathbf{x_1}, \ldots, \mathbf{x_m}$ is a basis for S, then any vector $\mathbf{x}$ in S admits a unique representation as a linear combination $c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m}$. For if

$$\mathbf{x} = c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m} = d_1\mathbf{x_1} + \cdots + d_m\mathbf{x_m},$$

then

$$(c_1 - d_1)\mathbf{x_1} + \cdots + (c_m - d_m)\mathbf{x_m} = \mathbf{0},$$

and since $\mathbf{x_1}, \ldots, \mathbf{x_m}$ are linearly independent, $c_i = d_i$ for each i.

A vector space is said to be *finite-dimensional* if it has a basis consisting of finitely many vectors. The vector space containing only the zero vector is also finite-dimensional. We will consider only finite-dimensional vector spaces. Very often it will be implicitly assumed that the vector spaces under consideration are nontrivial, i.e., contain vectors other than the zero vector.

3.1. *Let S be a vector space. Then any two bases of S have the same cardinality.*

PROOF. Suppose $\mathbf{x_1}, \ldots, \mathbf{x_p}$ and $\mathbf{y_1}, \ldots, \mathbf{y_q}$ are bases for S and let, if possible, $p > q$. We can express every $\mathbf{x_i}$ as a linear combination of $\mathbf{y_1}, \ldots, \mathbf{y_q}$. Thus there exists a $p \times q$ matrix $\mathbf{A} = (a_{ij})$ such that

$$\mathbf{x_i} = \sum_{j=1}^{q} a_{ij}\mathbf{y_j}, \qquad i = 1, \ldots, p. \tag{1}$$

Similarly, there exists a $q \times p$ matrix $\mathbf{B} = (b_{ij})$ such that

$$\mathbf{y_j} = \sum_{k=1}^{p} b_{jk}\mathbf{x_k}, \qquad j = 1, \ldots, q. \tag{2}$$

From (1),(2) we see that

$$\mathbf{x_i} = \sum_{k=1}^{p} c_{ik}\mathbf{x_k}, \qquad i = 1, \ldots, p, \tag{3}$$

where $\mathbf{C} = \mathbf{AB}$. It follows from (3) and the observation made preceding **3.1** that $\mathbf{AB} = \mathbf{I}$, the identity matrix of order p. Add $p - q$ zero columns to $\mathbf{A}$ to get the $p \times p$ matrix $\mathbf{U}$. Similarly, add $p - q$ zero rows to $\mathbf{B}$ to get the $p \times p$ matrix $\mathbf{V}$. Then $\mathbf{UV} = \mathbf{AB} = \mathbf{I}$. Therefore, $|\mathbf{UV}| = 1$. However, $|\mathbf{U}| = |\mathbf{V}| = 0$, since $\mathbf{U}$ has a zero column and $\mathbf{V}$ has a zero row. Thus we have a contradiction, and hence $p \leq q$. We can similarly prove that $q \leq p$, it follows that $p = q$. □

In the process of proving **3.1** we have proved the following statement which will be useful. Let S be a vector space. Suppose $\mathbf{x_1}, \ldots, \mathbf{x_p}$ is a basis for S and suppose the set $\mathbf{y_1}, \ldots, \mathbf{y_q}$ spans S. Then $p \leq q$.

The *dimension* of the vector space S, denoted by $\dim(S)$, is defined to be the cardinality of a basis of S. By convention the dimension of the space containing only the zero vector is zero.

Let S, T be vector spaces. We say that S is *isomorphic* to T if there exists a one-to-one and onto map $f : S \longrightarrow T$ such that f is *linear*, i.e., $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x})+f(\mathbf{y})$ and $f(c\mathbf{x}) = cf(\mathbf{x})$ for all $\mathbf{x}$, $\mathbf{y}$ in S and real numbers c.

3.2. *Let S, T be vector spaces. Then S, T are isomorphic if and only if* $\dim(S) = \dim(T)$.

PROOF. We first prove the *only if* part. Suppose $f : S \longrightarrow T$ is an isomorphism. If $\mathbf{x_1}, \dots, \mathbf{x_k}$ is a basis for S, then we will show that $f(\mathbf{x_1}), \dots, f(\mathbf{x_k})$ is a basis for T. First suppose $c_1 f(\mathbf{x_1}) + \cdots + c_k f(\mathbf{x_k}) = \mathbf{0}$. It follows from the definition of isomorphism that $f(c_1\mathbf{x_1} + \cdots + c_k\mathbf{x_k}) = \mathbf{0}$ and hence $c_1\mathbf{x_1} + \cdots + c_k\mathbf{x_k} = \mathbf{0}$. Since $\mathbf{x_1}, \dots, \mathbf{x_k}$ are linearly independent, $c_1 = \cdots = c_k = 0$, and therefore $f(\mathbf{x_1}), \dots, f(\mathbf{x_k})$ are linearly independent. If $\mathbf{v} \in T$, then there exists $\mathbf{u} \in S$ such that $f(\mathbf{u}) = \mathbf{v}$. We can write $\mathbf{u} = d_1\mathbf{x_1} + \cdots + d_k\mathbf{x_k}$ for some $d_1, \dots, d_k$. Now, $\mathbf{v} = f(\mathbf{u}) = d_1 f(\mathbf{x_1}) + \cdots + d_k f(\mathbf{x_k})$. Thus $f(\mathbf{x_1}), \dots, f(\mathbf{x_k})$ span T and hence form a basis for T. It follows that $\dim(T) = k$.

To prove the converse, let $\mathbf{x_1}, \dots, \mathbf{x_k}; \mathbf{y_1}, \dots, \mathbf{y_k}$ be bases for S, T, respectively. (Since $\dim(S) = \dim(T)$, the bases have the same cardinality.) Any $\mathbf{x}$ in S admits a unique representation

$$\mathbf{x} = c_1\mathbf{x_1} + \cdots + c_k\mathbf{x_k}.$$

Define $f(\mathbf{x}) = \mathbf{y}$, where $\mathbf{y} = c_1\mathbf{y_1} + \cdots + c_k\mathbf{y_k}$. It can be verified that f satisfies the definition of isomorphism. □

3.3. *Let S be a vector space and suppose S is the linear span of the vectors $\mathbf{x_1}, \dots, \mathbf{x_m}$. If some $\mathbf{x_i}$ is a linear combination of $\mathbf{x_1}, \dots, \mathbf{x_{i-1}}, \mathbf{x_{i+1}} \dots, \mathbf{x_m}$, then these latter vectors also span S.*

The proof is easy.

3.4. *Let S be a vector space of dimension n and let $\mathbf{x_1}, \dots, \mathbf{x_m}$ be linearly independent vectors in S. Then there exists a basis for S containing $\mathbf{x_1}, \dots, \mathbf{x_m}$.*

PROOF. Let $\mathbf{y_1}, \dots, \mathbf{y_n}$ be a basis for S. The set $\mathbf{x_1}, \dots, \mathbf{x_m}, \mathbf{y_1}, \dots, \mathbf{y_n}$ is linearly dependent, and therefore there exists a linear combination

$$c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m} + d_1\mathbf{y_1} + \cdots + d_n\mathbf{y_n} = \mathbf{0}$$

where some c_i or d_i is nonzero. However, since $\mathbf{x_1}, \dots, \mathbf{x_m}$ are linearly independent, it must be true that some d_i is nonzero. Therefore, some $\mathbf{y_i}$ is a linear combination of the remaining vectors. By **3.3** the set

$$\mathbf{x_1}, \dots, \mathbf{x_m}, \mathbf{y_1}, \dots, \mathbf{y_{i-1}}, \mathbf{y_{i+1}}, \dots, \mathbf{y_n}$$

also spans S. If the set is linearly independent, then we have a basis as required. Otherwise, we continue the process until we get a basis containing $\mathbf{x_1}, \dots, \mathbf{x_m}$. □

3.5. *Any set of $n + 1$ vectors in R^n is linearly dependent.*

PROOF. If the set is linearly independent then by **3.4** we can find a basis for R^n containing the set. This is a contradiction since every basis for R^n must contain precisely n vectors. □

3.6. *Any subspace S of R^n admits a basis.*

PROOF. Choose vectors $\mathbf{x_1}, \ldots, \mathbf{x_m}$ in S successively so that at each stage they are linearly independent. At any stage if the vectors span S, then we have a basis. Otherwise, there exists a vector $\mathbf{x_{m+1}}$ in S that is not in the linear span of $\mathbf{x_1}, \ldots, \mathbf{x_m}$, and we arrive at the set $\mathbf{x_1}, \ldots, \mathbf{x_m}, \mathbf{x_{m+1}}$, which is linearly independent. The process must terminate, since by **3.5** any $n+1$ vectors in R^n are linearly dependent. □

3.7. *If S is a subspace of T, then* $\dim(S) \leq \dim(T)$. *Furthermore, equality holds if and only if $S = T$.*

PROOF. Recall that we consider only finite-dimensional vector spaces. Suppose $\dim S = p$, $\dim T = q$, and let $\mathbf{x_1}, \ldots, \mathbf{x_p}$ and $\mathbf{y_1}, \ldots, \mathbf{y_q}$, be bases for S, T, respectively. Using a similar argument as in the proof of **3.6** we can show that any set of r vectors in T is linearly dependent if $r > q$. Since $\mathbf{x_1}, \ldots, \mathbf{x_p}$ is a linearly independent set of vectors in $S \subset T$, we have $p \leq q$.

To prove the second part, supose $p = q$ and suppose $S \neq T$. Then there exists a vector $\mathbf{z} \in T$ that is not in the span of $\mathbf{x_1}, \ldots, \mathbf{x_p}$. Then the set $\mathbf{x_1}, \ldots, \mathbf{x_p}, \mathbf{z}$ is linearly independent. This is a contradiction, since by the remark made earlier, any $p+1$ vectors in T must be linearly dependent. Therefore, we have shown that if S is a subspace of T and if $\dim S = \dim T$, then $S = T$. Conversely, if $S = T$, then clearly $\dim S = \dim T$, and the proof is complete. □

Problems

1. Verify that each of the following sets is a vector space and find its dimension: (i) Vectors (a, b, c, d) such that $a + b = c + d$; (ii) $n \times n$ matrices with zero trace; (iii) The set of solutions (x, y, z) to the system $2x - y = 0, 2y + 3z = 0$.
2. If $\mathbf{x}, \mathbf{y}, \mathbf{z}$ is a basis for R^3, which of the following are also bases for R^3? (i) **x+2y, y + 3z, x+2z**; (ii) **x+y-2z, x - 2y+z, -2x+y+z**; (iii) **x, y, x+y+z**.
3. If $\{\mathbf{x_1}, \mathbf{x_2}\}$ and $\{\mathbf{y_1}, \mathbf{y_2}\}$ are both bases of R^2, show that at least one of the following statements is true: (i) $\{\mathbf{x_1}, \mathbf{y_2}\}, \{\mathbf{x_2}, \mathbf{y_1}\}$ are both bases of R^2; (ii) $\{\mathbf{x_1}, \mathbf{y_1}\}, \{\mathbf{x_2}, \mathbf{y_2}\}$ are both bases of R^2.

1.4 Rank

Let $\mathbf{A}$ be an $m \times n$ matrix. The subspace of R^m spanned by the column vectors of $\mathbf{A}$ is called the *column space* or the *column span* of $\mathbf{A}$ and is denoted by $\mathcal{C}(\mathbf{A})$. Similarly, the subspace of R^n spanned by the row vectors of $\mathbf{A}$ is called the *row space* of $\mathbf{A}$, denoted by $\mathcal{R}(\mathbf{A})$. Clearly, $\mathcal{R}(\mathbf{A})$ is isomorphic to $\mathcal{C}(\mathbf{A}')$. The dimension of the column space is called the *column rank*, whereas the dimension of the row space is called the *row rank* of the matrix. These two definitions turn out be very short-lived in any linear algebra book, since the two ranks are always equal, as we show in the next result.

4.1. *The column rank of a matrix equals its row rank.*

PROOF. Let $\mathbf{A}$ be an $m \times n$ matrix with column rank r. Then $\mathcal{C}(\mathbf{A})$ has a basis of r vectors, say $\mathbf{b_1}, \ldots, \mathbf{b_r}$. Let $\mathbf{B}$ be the $m \times r$ matrix $[\mathbf{b_1}, \ldots, \mathbf{b_r}]$. Since every column of $\mathbf{A}$ is a linear combination of $\mathbf{b_1}, \ldots, \mathbf{b_r}$, we can write $\mathbf{A} = \mathbf{BC}$ for some $r \times n$ matrix $\mathbf{C}$. Then every row of $\mathbf{A}$ is a linear combination of the rows of $\mathbf{C}$, and therefore $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{C})$. It follows by **3.7** that the dimension of $\mathcal{R}(\mathbf{A})$, which is the row rank of $\mathbf{A}$, is at most r. We can similarly show that the column rank does not exceed the row rank and, therefore, the two must be equal. □

The common value of the column rank and the row rank of $\mathbf{A}$ will henceforth be called the *rank* of $\mathbf{A}$, and we will denote it by $R(\mathbf{A})$. This notation should not be confused with the notation used to denote the row space of $\mathbf{A}$, namely, $\mathcal{R}(\mathbf{A})$.

It is obvious that $R(\mathbf{A}) = R(\mathbf{A}')$. The rank of $\mathbf{A}$ is zero if and only if $\mathbf{A}$ is the zero matrix.

4.2. *Let* $\mathbf{A}, \mathbf{B}$ *be matrices such that* $\mathbf{AB}$ *is defined. Then*

$$R(\mathbf{AB}) \le \min\{R(\mathbf{A}), R(\mathbf{B})\}.$$

PROOF. A vector in $\mathcal{C}(\mathbf{AB})$ is of the form $\mathbf{ABx}$ for some vector $\mathbf{x}$, and therefore it belongs to $\mathcal{C}(\mathbf{A})$. Thus $\mathcal{C}(\mathbf{AB}) \subset \mathcal{C}(\mathbf{A})$, and hence by **3.7**,

$$R(\mathbf{AB}) = \dim \mathcal{C}(\mathbf{AB}) \le \dim \mathcal{C}(\mathbf{A}) = R(\mathbf{A}).$$

Now using this fact we have

$$R(\mathbf{AB}) = R(\mathbf{B}'\mathbf{A}') \le R(\mathbf{B}') = R(\mathbf{B}).$$ □

4.3. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix of rank* $r, r \neq 0$. *Then there exist matrices* $\mathbf{B}, \mathbf{C}$ *of order* $m \times r$ *and* $r \times n$, *respectively, such that* $R(\mathbf{B}) = R(\mathbf{C}) = r$ *and* $\mathbf{A} = \mathbf{BC}$. *This decomposition is called a rank factorization of* $\mathbf{A}$.

PROOF. The proof proceeds along the same lines as that of **4.1**, so that we can write $\mathbf{A} = \mathbf{BC}$, where $\mathbf{B}$ is $m \times r$ and $\mathbf{C}$ is $r \times n$. Since the columns of $\mathbf{B}$ are linearly independent, $R(\mathbf{B}) = r$. Since $\mathbf{C}$ has r rows, $R(\mathbf{C}) \le r$. However, by **4.2**, $r = R(\mathbf{A}) \le R(\mathbf{C})$, and hence $R(\mathbf{C}) = r$. □

Throughout this book whenever we talk of rank factorization of a matrix it is implicitly assumed that the matrix is nonzero.

4.4. *Let* $\mathbf{A}, \mathbf{B}$ *be* $m \times n$ *matrices. Then* $R(\mathbf{A} + \mathbf{B}) \le R(\mathbf{A}) + R(\mathbf{B})$.

PROOF. Let $\mathbf{A} = \mathbf{XY}, \mathbf{B} = \mathbf{UV}$ be rank factorizations of $\mathbf{A}, \mathbf{B}$. Then

$$\mathbf{A} + \mathbf{B} = \mathbf{XY} + \mathbf{UV} = [\mathbf{X}, \mathbf{U}] \begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix}.$$

Therefore, by **4.2**,

$$R(\mathbf{A}+\mathbf{B}) \le R[\mathbf{X}, \mathbf{U}].$$

Let $\mathbf{x_1}, \dots, \mathbf{x_p}$ and $\mathbf{u_1}, \dots, \mathbf{u_q}$ be bases for $\mathcal{C}(\mathbf{X})$, $\mathcal{C}(\mathbf{U})$, respectively. Any vector in the column space of $[\mathbf{X}, \mathbf{U}]$ can be expressed as a linear combination of these $p+q$ vectors. Thus

$$R[\mathbf{X}, \mathbf{U}] \le R(\mathbf{X}) + R(\mathbf{U}) = R(\mathbf{A}) + R(\mathbf{B}),$$

and the proof is complete. □

The following operations performed on a matrix $\mathbf{A}$ are called *elementary column operations*:

(i) Interchange two columns of $\mathbf{A}$.
(ii) Multiply a column of $\mathbf{A}$ by a nonzero scalar.
(iii) Add a scalar multiple of one column to another column.

These operations clearly leave $\mathcal{C}(\mathbf{A})$ unaffected, and therefore they do not change the rank of the matrix. We can define elementary row operations similarly. The elementary row and column operations are particularly useful in computations. Thus to find the rank of a matrix we first reduce it to a matrix with several zeros by these operations and then compute the rank of the resulting matrix.

Problems

1. Find the rank of the following matrix for each real number α:

$$\begin{bmatrix} 1 & 4 & \alpha & 4 \\ 2 & -6 & 7 & 1 \\ 3 & 2 & -6 & 7 \\ 2 & 2 & -5 & 5 \end{bmatrix}.$$

2. Let $\{\mathbf{x_1}, \dots, \mathbf{x_p}\}$, $\{\mathbf{y_1}, \dots, \mathbf{y_q}\}$ be linearly independent sets in R^n, where $p < q \le n$. Show that there exists $i \in \{1, \dots, q\}$ such that $\{\mathbf{x_1}, \dots, \mathbf{x_p}, \mathbf{y_i}\}$ is linearly independent.

3. Let $\mathbf{A}$ be an $m \times n$ matrix and let $\mathbf{B}$ be obtained by changing any k entries of $\mathbf{A}$. Show that

$$R(\mathbf{A}) - k \le R(\mathbf{B}) \le R(\mathbf{A}) + k.$$

4. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be $n \times n$ matrices. Is it always true that $R(\mathbf{ABC}) \le R(\mathbf{AC})$?

1.5 Orthogonality

Let S be a vector space. A function that assigns a real number $\langle \mathbf{x}, \mathbf{y} \rangle$ to every pair of vectors $\mathbf{x}, \mathbf{y}$ in S is said to be an *inner product* if it satisfies the following conditions:

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.
(ii) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and equality holds if and only if $\mathbf{x} = \mathbf{0}$.
(iii) $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$.
(iv) $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.

In R^n, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y} = x_1 y_1 + \cdots + x_n y_n$ is easily seen to be an inner product. We will work with this inner product while dealing with R^n and its subspaces, unless indicated otherwise.

For a vector $\mathbf{x}$, the positive square root of the inner product $\langle \mathbf{x}, \mathbf{x} \rangle$ is called the *norm* of $\mathbf{x}$, denoted by $\|\mathbf{x}\|$. Vectors $\mathbf{x}$, $\mathbf{y}$ are said to be *orthogonal* or *perpendicular* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, in which case we write $\mathbf{x} \perp \mathbf{y}$.

5.1. *If $\mathbf{x_1}, \ldots, \mathbf{x_m}$ are pairwise orthogonal nonzero vectors, then they are linearly independent.*

PROOF. Suppose $c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m} = \mathbf{0}$. Then

$$\langle c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m}, \mathbf{x_1} \rangle = 0,$$

and hence,

$$\sum_{i=1}^{m} c_i \langle \mathbf{x_i}, \mathbf{x_1} \rangle = 0.$$

Since the vectors $\mathbf{x_1}, \ldots, \mathbf{x_m}$ are pairwise orthogonal, it follows that $c_1 \langle \mathbf{x_1}, \mathbf{x_1} \rangle = 0$, and since $\mathbf{x_1}$ is nonzero, $c_1 = 0$. Similarly, we can show that each c_i is zero. Therefore, the vectors are linearly independent. □

A set of vectors $\mathbf{x_1}, \ldots, \mathbf{x_m}$ is said to form an *orthonormal basis* for the vector space S if the set is a basis for S and furthermore, $\langle \mathbf{x_i}, \mathbf{x_j} \rangle$ is 0 if $i \neq j$ and 1 if $i = j$.

We now describe the *Gram–Schmidt procedure*, which produces an orthonormal basis starting with a given basis $\mathbf{x_1}, \ldots, \mathbf{x_n}$.

Set $\mathbf{y_1} = \mathbf{x_1}$. Having defined $\mathbf{y_1}, \ldots, \mathbf{y_{i-1}}$, we define

$$\mathbf{y_i} = \mathbf{x_i} - a_{i,i-1}\mathbf{y_{i-1}} - \cdots - a_{i1}\mathbf{y_1},$$

where $a_{i,i-1}, \ldots, a_{i1}$ are chosen so that $\mathbf{y_i}$ is orthogonal to $\mathbf{y_1}, \ldots, \mathbf{y_{i-1}}$. Thus we must solve $\langle \mathbf{y_i}, \mathbf{y_j} \rangle = 0$, $j = 1, \ldots, i-1$. This leads to

$$\langle \mathbf{x_i} - a_{i,i-1}\mathbf{y_{i-1}} - \cdots - a_{i1}\mathbf{y_1}, \mathbf{y_j} \rangle = 0, \qquad j = 1, \ldots, i-1,$$

which gives

$$\langle \mathbf{x_i}, \mathbf{y_j} \rangle - \sum_{k=1}^{i-1} a_{ik} \langle \mathbf{y_k}, \mathbf{y_j} \rangle = 0, \qquad j = 1, \ldots, i-1.$$

Now, since $\mathbf{y_1}, \ldots, \mathbf{y_{i-1}}$ is an orthogonal set, we get

$$\langle \mathbf{x_i}, \mathbf{y_j} \rangle - a_{ij} \langle \mathbf{y_j}, \mathbf{y_j} \rangle = 0,$$

and hence,

$$a_{ij} = \frac{\langle \mathbf{x_i}, \mathbf{y_j} \rangle}{\langle \mathbf{y_j}, \mathbf{y_j} \rangle}, \qquad j = 1, \ldots, i-1.$$

The process is continued to obtain the basis $\mathbf{y_1}, \ldots, \mathbf{y_n}$ of pairwise orthogonal vectors. Since $\mathbf{x_1}, \ldots, \mathbf{x_n}$ are linearly independent, each $\mathbf{y_i}$ is nonzero. Now if we set $\mathbf{z_i} = \frac{\mathbf{y_i}}{\|\mathbf{y_i}\|}$, then $\mathbf{z_1}, \ldots, \mathbf{z_n}$ is an orthonormal basis. Note that the linear span of $\mathbf{z_1}, \ldots, \mathbf{z_i}$ equals the linear span of $\mathbf{x_1}, \ldots, \mathbf{x_i}$ for each i.

We remark that given a set of linearly independent vectors $\mathbf{x_1}, \ldots, \mathbf{x_m}$, the Gram–Schmidt procedure described above can be used to produce a pairwise orthogonal set $\mathbf{y_1}, \ldots, \mathbf{y_m}$, such that $\mathbf{y_i}$ is a linear combination of $\mathbf{x_1}, \ldots, \mathbf{x_{i-1}}$, $i = 1, \ldots, m$. This fact is used in the proof of the next result.

Let W be a set (not necessarily a subspace) of vectors in a vector space S. We define

$$W^{\perp} = \{\mathbf{x} : \mathbf{x} \in S, \langle \mathbf{x}, \mathbf{y} \rangle = 0 \text{ for all } \mathbf{y} \in W\}.$$

It follows from the definitions that $W^{\perp}$ is a subspace of S.

5.2. *Let S be a subspace of the vector space T and let $\mathbf{x} \in T$. Then there exists a unique decomposition $\mathbf{x} = \mathbf{u} + \mathbf{v}$ such that $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$. The vector $\mathbf{u}$ is called the orthogonal projection of $\mathbf{x}$ on the vector space S.*

Proof. If $\mathbf{x} \in S$, then $\mathbf{x} = \mathbf{x} + \mathbf{0}$ is the required decomposition. Otherwise, let $\mathbf{x_1}, \ldots, \mathbf{x_m}$ be a basis for S. Use the Gram–Schmidt process on the set $\mathbf{x_1}, \ldots, \mathbf{x_m}, \mathbf{x}$ to obtain the sequence $\mathbf{y_1}, \ldots, \mathbf{y_m}, \mathbf{v}$ of pairwise orthogonal vectors. Since $\mathbf{v}$ is perpendicular to each $\mathbf{y_i}$ and since the linear span of $\mathbf{y_1}, \ldots, \mathbf{y_m}$ equals that of $\mathbf{x_1}, \ldots, \mathbf{x_m}$, then $\mathbf{v} \in S^{\perp}$. Also, according to the Gram–Schmidt process, $\mathbf{x} - \mathbf{v}$ is a linear combination of $\mathbf{y_1}, \ldots, \mathbf{y_m}$ and hence $\mathbf{x} - \mathbf{v} \in S$. Now $\mathbf{x} = (\mathbf{x} - \mathbf{v}) + \mathbf{v}$ is the required decomposition. It remains to show the uniqueness.

If $\mathbf{x} = \mathbf{u_1} + \mathbf{v_1} = \mathbf{u_2} + \mathbf{v_2}$ are two decompositions satisfying $\mathbf{u_1} \in S, \mathbf{u_2} \in S, \mathbf{v_1} \in S^{\perp}, \mathbf{v_2} \in S^{\perp}$, then

$$(\mathbf{u_1} - \mathbf{u_2}) + (\mathbf{v_1} - \mathbf{v_2}) = \mathbf{0}.$$

Since $\langle \mathbf{u_1} - \mathbf{u_2}, \mathbf{v_1} - \mathbf{v_2} \rangle = 0$, it follows from the preceding equation that $\langle \mathbf{u_1} - \mathbf{u_2}, \mathbf{u_1} - \mathbf{u_2} \rangle = 0$. Then $\mathbf{u_1} - \mathbf{u_2} = \mathbf{0}$, and hence $\mathbf{u_1} = \mathbf{u_2}$. It easily follows that $\mathbf{v_1} = \mathbf{v_2}$. Thus the decomposition is unique. □

5.3. *Let W be a subset of the vector space T and let S be the linear span of W. Then*

$$\dim(S) + \dim(W^{\perp}) = \dim(T).$$

Proof. Suppose $\dim(S) = m$, $\dim(W^{\perp}) = n$, and $\dim(T) = p$. Let $\mathbf{x_1}, \ldots, \mathbf{x_m}$ and $\mathbf{y_1}, \ldots, \mathbf{y_n}$ be bases for $S, W^{\perp}$, respectively. Suppose

$$c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m} + d_1\mathbf{y_1} + \cdots + d_n\mathbf{y_n} = \mathbf{0}.$$

Let $\mathbf{u} = c_1\mathbf{x_1} + \cdots + c_m\mathbf{x_m}$, $\mathbf{v} = d_1\mathbf{y_1} + \cdots + d_n\mathbf{y_n}$. Since $\mathbf{x_i}, \mathbf{y_j}$ are orthogonal for each i, j, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. However, $\mathbf{u} + \mathbf{v} = \mathbf{0}$ and hence $\mathbf{u} = \mathbf{v} = \mathbf{0}$.

It follows that $c_i = 0, d_j = 0$ for each i, j, and hence $\mathbf{x_1}, \ldots, \mathbf{x_m}, \mathbf{y_1}, \ldots, \mathbf{y_n}$ is a linearly independent set. Therefore, $m + n \leq p$. If $m + n < p$, then there exists a vector $\mathbf{z} \in T$ such that $\mathbf{x_1}, \ldots, \mathbf{x_m}, \mathbf{y_1}, \ldots, \mathbf{y_n}, \mathbf{z}$ is a linearly independent set. Let M be the linear span of $\mathbf{x_1}, \ldots, \mathbf{x_m}, \mathbf{y_1}, \ldots, \mathbf{y_n}$. By **5.2** there exists a decomposition $\mathbf{z} = \mathbf{u} + \mathbf{v}$ such that $\mathbf{u} \in M$, $\mathbf{v} \in M^{\perp}$. Then $\mathbf{v}$ is orthogonal to $\mathbf{x_i}$ for every i, and hence $\mathbf{v} \in W^{\perp}$. Also, $\mathbf{v}$ is orthogonal to $\mathbf{y_i}$ for every i, and hence $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ and therefore $\mathbf{v} = \mathbf{0}$. It follows that $\mathbf{z} = \mathbf{u}$. This contradicts the fact that $\mathbf{z}$ is linearly independent of $\mathbf{x_1}, \ldots, \mathbf{x_m}, \mathbf{y_1}, \ldots, \mathbf{y_n}$. Therefore, $m + n = p$. □

The proof of the next result is left as an exercise.

5.4. *If $S_1 \subset S_2 \subset T$ are vector spaces, then* (i) $(S_2)^{\perp} \subset (S_1)^{\perp}$; (ii) $(S_1^{\perp})^{\perp} = S_1$.

Let $\mathbf{A}$ be an $m \times n$ matrix. The set of all vectors $\mathbf{x} \in R^n$ such that $\mathbf{Ax} = \mathbf{0}$ is easily seen to be a subspace of R^n. This subspace is called the *null space* of $\mathbf{A}$, and we denote it by $\mathcal{N}(\mathbf{A})$.

5.5. *Let $\mathbf{A}$ be an $m \times n$ matrix. Then $\mathcal{N}(\mathbf{A}) = \mathcal{C}(\mathbf{A}')^{\perp}$.*

PROOF. If $\mathbf{x} \in \mathcal{N}(\mathbf{A})$, then $\mathbf{Ax} = \mathbf{0}$, and hence $\mathbf{y}'\mathbf{Ax} = \mathbf{0}$ for all $\mathbf{y} \in R^m$. Thus $\mathbf{x}$ is orthogonal to any vector in $\mathcal{C}(\mathbf{A}')$. Conversely, if $\mathbf{x} \in \mathcal{C}(\mathbf{A}')^{\perp}$, then $\mathbf{x}$ is orthogonal to every column of $\mathbf{A}'$, and therefore $\mathbf{Ax} = \mathbf{0}$. □

5.6. *Let $\mathbf{A}$ be an $m \times n$ matrix of rank r. Then* $\dim(\mathcal{N}(\mathbf{A})) = n - r$.

PROOF. We have

$$\begin{aligned} \dim(\mathcal{N}(\mathbf{A})) &= \dim \mathcal{C}(\mathbf{A}')^{\perp} && \text{by } \mathbf{5.5} \\ &= n - \dim \mathcal{C}(\mathbf{A}') && \text{by } \mathbf{5.3} \\ &= n - r. \end{aligned}$$

This completes the proof. □

The dimension of the null space of $\mathbf{A}$ is called the *nullity* of $\mathbf{A}$. Thus **5.6** says that *the rank plus the nullity equals the number of columns.*

Problems

1. Which of the following functions define an inner product on R^3? (i) $f(\mathbf{x}, \mathbf{y}) = x_1y_1 + x_2y_2 + x_3y_3 + 1$; (ii) $f(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + 3x_2y_2 + x_3y_3 - x_1y_2 - x_2y_1$; (iii) $f(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_2y_2 + x_3y_3 + 2x_1y_2 + 2x_2y_1$; (iv) $f(\mathbf{x}, \mathbf{y}) = x_1y_1 + x_2y_2$; (v) $f(\mathbf{x}, \mathbf{y}) = x_1^3y_1^3 + x_2^3y_2^3 + x_3^3y_3^3$.

2. Show that the following vectors form a basis for R^3. Use the Gram–Schmidt procedure to convert it into an orthonormal basis.

$$\mathbf{x} = \begin{bmatrix} 2 & 3 & -1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 3 & 1 & 0 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 4 & -1 & 2 \end{bmatrix}.$$

1.6 Nonsingularity

Suppose we have m linear equations in the n unknowns $x_1, \ldots, x_n$. The equations can conveniently be expressed as a single matrix equation $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is the $m \times n$ matrix of coefficients. The equation $\mathbf{Ax} = \mathbf{b}$ is said to be *consistent* if it has at least one solution; otherwise, it is *inconsistent.* The equation is *homogeneous* if $\mathbf{b} = \mathbf{0}$. The set of solutions of the homogeneous equation $\mathbf{Ax} = \mathbf{0}$ is clearly the null space of $\mathbf{A}$.

If the equation $\mathbf{Ax} = \mathbf{b}$ is consistent, then we can write

$$\mathbf{b} = x_1^0\mathbf{a_1} + \cdots + x_n^0\mathbf{a_n}$$

for some $x_1^0, \ldots, x_n^0$, where $\mathbf{a_1}, \ldots, \mathbf{a_n}$ are the columns of $\mathbf{A}$. Thus $\mathbf{b} \in \mathcal{C}(\mathbf{A})$. Conversely, if $\mathbf{b} \in \mathcal{C}(\mathbf{A})$, then $\mathbf{Ax} = \mathbf{b}$ must be consistent. If the equation is consistent and if $\mathbf{x^0}$ is a solution of the equation, then the set of all solutions of the equation is given by

$$\{\mathbf{x^0} + \mathbf{x} : \mathbf{x} \in \mathcal{N}(\mathbf{A})\}.$$

Clearly, the equation $\mathbf{Ax} = \mathbf{b}$ has either no solution, a unique solution, or infinitely many solutions.

A matrix $\mathbf{A}$ of order $n \times n$ is said to be *nonsingular* if $R(\mathbf{A}) = n$; otherwise, the matrix is *singular.*

6.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then the following conditions are equivalent:*

(i) $\mathbf{A}$ *is nonsingular, i.e.,* $R(\mathbf{A}) = n$.
(ii) *For any* $\mathbf{b} \in R^n$, $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*
(iii) *There exists a unique matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$.

PROOF. (i) $\Rightarrow$ (ii). Since $R(\mathbf{A}) = n$, we have $\mathcal{C}(\mathbf{A}) = R^n$, and therefore $\mathbf{Ax} = \mathbf{b}$ has a solution. If $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{Ay} = \mathbf{b}$, then $\mathbf{A}(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. By **5.6**, $\dim(\mathcal{N}(\mathbf{A})) = 0$ and therefore $\mathbf{x} = \mathbf{y}$. This proves the uniqueness.

(ii) $\Rightarrow$ (iii). By (ii), $\mathbf{Ax} = \mathbf{e_i}$ has a unique solution, say $\mathbf{b_i}$, where $\mathbf{e_i}$ is the ith column of the identity matrix. Then $\mathbf{B} = [\mathbf{b_1}, \ldots, \mathbf{b_n}]$ is a unique matrix satisfying $\mathbf{AB} = \mathbf{I}$. Applying the same argument to $\mathbf{A}'$, we conclude the existence of a unique matrix $\mathbf{C}$ such that $\mathbf{CA} = \mathbf{I}$. Now $\mathbf{B} = (\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{C}$.

(iii) $\Rightarrow$ (i). Suppose (iii) holds. Then any $\mathbf{x} \in R^n$ can be expressed as $\mathbf{x} = \mathbf{A}(\mathbf{Bx})$, and hence $\mathcal{C}(\mathbf{A}) = R^n$. Thus $R(\mathbf{A})$, which by definition is $\dim(\mathcal{C}(\mathbf{A}))$, must be n. □

The matrix $\mathbf{B}$ of (iii) of **6.1** is called the *inverse* of $\mathbf{A}$ and is denoted by $\mathbf{A^{-1}}$.

If $\mathbf{A}, \mathbf{B}$ are $n \times n$ matrices, then $(\mathbf{AB})(\mathbf{B^{-1}A^{-1}}) = \mathbf{I}$, and therefore $(\mathbf{AB})^{-1} = \mathbf{B^{-1}A^{-1}}$. In particular, the product of two nonsingular matrices is nonsingular.

Let $\mathbf{A}$ be an $n \times n$ matrix. We will denote by $\mathbf{A_{ij}}$ the submatrix of $\mathbf{A}$ obtained by deleting row i and column j. The *cofactor* of a_{ij} is defined to be $(-1)^{i+j}|\mathbf{A_{ij}}|$. The *adjoint* of $\mathbf{A}$, denoted by adj $\mathbf{A}$, is the $n \times n$ matrix whose (i, j)-entry is the cofactor of a_{ji}.

From the theory of determinants we have

$$\sum_{j=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{A_{ij}}| = |\mathbf{A}|,$$

and for $i \neq k$,

$$\sum_{j=1}^{n} a_{ij}(-1)^{j+k}|\mathbf{A_{kj}}| = 0.$$

These equations can be interpreted as

$$\mathbf{A}\mathrm{adj}\mathbf{A} = |\mathbf{A}|\mathbf{I}.$$

Thus if $|\mathbf{A}| \neq 0$, then $\mathbf{A^{-1}}$ exists and

$$\mathbf{A^{-1}} = \frac{1}{|\mathbf{A}|}\mathrm{adj}\mathbf{A}.$$

Conversely, if $\mathbf{A}$ is nonsingular, then from $\mathbf{AA^{-1}} = \mathbf{I}$ we conclude that $|\mathbf{AA^{-1}}| = |\mathbf{A}||\mathbf{A^{-1}}| = 1$ and therefore $|\mathbf{A}| \neq 0$. We have therefore proved the following result:

6.2. *A square matrix is nonsingular if and only if its determinant is nonzero.*

An $r \times r$ *minor* of a matrix is defined to be the determinant of an $r \times r$ submatrix of $\mathbf{A}$.

Let $\mathbf{A}$ be an $m \times n$ matrix of rank r, let $s > r$, and consider an $s \times s$ minor of $\mathbf{A}$, say the one formed by rows $i_1, \dots, i_s$ and columns $j_1, \dots, j_s$. Since the columns $j_1, \dots, j_s$ must be linearly dependent, then by **6.2** the minor must be zero.

Conversely, if $\mathbf{A}$ is of rank r, then $\mathbf{A}$ has r linearly independent rows, say the rows $i_1, \dots, i_r$. Let $\mathbf{B}$ be the submatrix formed by these r rows. Then $\mathbf{B}$ has rank r, and hence $\mathbf{B}$ has column rank r. Thus there is an $r \times r$ submatrix $\mathbf{C}$ of $\mathbf{B}$, and hence of $\mathbf{A}$, of rank r. By **6.2**, $\mathbf{C}$ has a nonzero determinant.

We therefore have the following definition of rank in terms of minors: The rank of the matrix $\mathbf{A}$ is r if (i) there is a nonzero $r \times r$ minor and (ii) every $s \times s$ minor, $s > r$, is zero. As remarked earlier, the rank is zero if and only if $\mathbf{A}$ is the zero matrix.

Problems

1. Let $\mathbf{A}$ be an $n \times n$ matrix. Show that $\mathbf{A}$ is nonsingular if and only if $\mathbf{Ax} = \mathbf{0}$ has no nonzero solution.

2. Let $\mathbf{A}$ be an $n \times n$ matrix and let $\mathbf{b} \in R^n$. Show that $\mathbf{A}$ is nonsingular if and only if $\mathbf{Ax} = \mathbf{b}$ has a unique solution.

3. Let $\mathbf{A}$ be an $n \times n$ matrix with only integer entries. Show that $\mathbf{A^{-1}}$ exists and has only integer entries if and only if $|\mathbf{A}| = \pm 1$.

4. Compute the inverses of the following matrices:

(i) $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where $ad - bc \neq 0$.

(ii) $\begin{bmatrix} 2 & -1 & 0 \\ 2 & 1 & -1 \\ 1 & 0 & 4 \end{bmatrix}$.

5. Let $\mathbf{A}$, $\mathbf{B}$ be matrices of order 9×7 and 4×3, respectively. Show that there exists a nonzero 7×4 matrix $\mathbf{X}$ such that $\mathbf{AXB} = \mathbf{0}$.

1.7 Frobenius Inequality

7.1. *Let* $\mathbf{B}$ *be an* $m \times r$ *matrix of rank* r. *Then there exists a matrix* $\mathbf{X}$ *(called a left inverse of* $\mathbf{B}$*), such that* $\mathbf{XB} = \mathbf{I}$.

PROOF. If $m = r$, then $\mathbf{B}$ is nonsingular and admits an inverse. So suppose $r < m$. The columns of $\mathbf{B}$ are linearly independent. Thus we can find a set of $m-r$ columns that together with the columns of $\mathbf{B}$ form a basis for R^m. In other words, we can find a matrix $\mathbf{U}$ of order $m \times (m-r)$ such that $[\mathbf{B}, \mathbf{U}]$ is nonsingular. Let the inverse of $[\mathbf{B}, \mathbf{U}]$ be partitioned as $\begin{bmatrix} \mathbf{X} \\ \mathbf{V} \end{bmatrix}$, where $\mathbf{X}$ is $r \times m$. Since

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{V} \end{bmatrix} [\mathbf{B}, \mathbf{U}] = \mathbf{I},$$

we have $\mathbf{XB} = \mathbf{I}$. □

We can similarly show that an $r \times n$ matrix $\mathbf{C}$ of rank r has a *right inverse,* i.e., a matrix $\mathbf{Y}$ such that $\mathbf{CY} = \mathbf{I}$. Note that a left inverse or a right inverse is not unique, unless the matrix is square and nonsingular.

7.2. *Let* $\mathbf{B}$ *be an* $m \times r$ *matrix of rank* r. *Then there exists a nonsingular matrix* $\mathbf{P}$ *such that*

$$\mathbf{PB} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}.$$

PROOF. The proof is the same as that of **7.1**. If we set $\mathbf{P} = \begin{bmatrix} \mathbf{X} \\ \mathbf{V} \end{bmatrix}$, then $\mathbf{P}$ satisfies the required condition. □

Similarly, if $\mathbf{C}$ is $r \times n$ of rank r, then there exists a nonsingular matrix $\mathbf{Q}$ such that $\mathbf{CQ} = [\mathbf{I}, \mathbf{0}]$. These two results and the rank factorization (see **4.3**) immediately lead to the following.

7.3. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix of rank* r. *Then there exist nonsingular matrices* $\mathbf{P}$, $\mathbf{Q}$ *such that*

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

7.4. *Let* $\mathbf{A}, \mathbf{B}$ *be matrices of order* $m \times n$ *and* $n \times p$*, respectively. If* $R(\mathbf{A}) = n$*, then* $R(\mathbf{AB}) = R(\mathbf{B})$*. If* $R(\mathbf{B}) = n$*, then* $R(\mathbf{AB}) = R(\mathbf{A})$*.*

PROOF. First suppose $R(\mathbf{A}) = n$. By **7.1**, there exists a matrix $\mathbf{X}$ such that $\mathbf{XA} = \mathbf{I}$. Then

$$R(\mathbf{B}) = R(\mathbf{XAB}) \leq R(\mathbf{AB}) \leq R(\mathbf{B}),$$

and hence $R(\mathbf{AB}) = R(\mathbf{B})$. The second part follows similarly. □

As an immediate corollary of **7.4** we see that the rank is not affected upon multiplying by a nonsingular matrix.

7.5. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix of rank* r*. Then there exists an* $m \times n$ *matrix* $\mathbf{Z}$ *of rank* $n - r$ *such that* $\mathbf{A} + \mathbf{Z}$ *has rank* n*.*

PROOF. By **7.3** there exist nonsingular matrices $\mathbf{P}, \mathbf{Q}$ such that

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Set

$$\mathbf{Z} = \mathbf{P}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix} \mathbf{Q}^{-1},$$

where $\mathbf{W}$ is any matrix of rank $n - r$. Then it is easily verified that $\mathbf{P}(\mathbf{A} + \mathbf{Z})\mathbf{Q}$ has rank n. Since $\mathbf{P}, \mathbf{Q}$ are nonsingular, it follows by the remark immediately preceding the result that $\mathbf{A} + \mathbf{Z}$ has rank n. □

Observe that **7.5** may also be proved using rank factorization; we leave this as an exercise.

7.6 (The Frobenius Inequality). *Let* $\mathbf{A}, \mathbf{B}$ *be matrices of order* $m \times n$ *and* $n \times p$ *respectively. Then*

$$R(\mathbf{AB}) \geq R(\mathbf{A}) + R(\mathbf{B}) - n.$$

PROOF. By **7.5** there exists a matrix $\mathbf{Z}$ of rank $n - R(\mathbf{A})$ such that $\mathbf{A} + \mathbf{Z}$ has rank n. We have

$$\begin{aligned} R(\mathbf{B}) &= R((\mathbf{A} + \mathbf{Z})\mathbf{B}) \qquad (\text{by } \mathbf{7.4}) \\ &= R(\mathbf{AB} + \mathbf{ZB}) \\ &\leq R(\mathbf{AB}) + R(\mathbf{ZB}) \qquad (\text{by } \mathbf{4.4}) \\ &\leq R(\mathbf{AB}) + R(\mathbf{Z}) \\ &= R(\mathbf{AB}) + n - R(\mathbf{A}). \end{aligned}$$

Hence $R(\mathbf{AB}) \geq R(\mathbf{A}) + R(\mathbf{B}) - n$. □

Problems

1. Let $\mathbf{A}, \mathbf{X}, \mathbf{B}$ be matrices such that the product $\mathbf{AXB}$ is defined. Prove the following generalization of the Frobenius inequality:

$$R(\mathbf{AXB}) \geq R(\mathbf{AX}) + R(\mathbf{XB}) - R(\mathbf{X}).$$

2. Let $\mathbf{A}$ be an $n \times n$ matrix such that $\mathbf{A^2} = \mathbf{I}$. Show that $R(\mathbf{I}+\mathbf{A})+R(\mathbf{I}-\mathbf{A}) = n$.

1.8 Eigenvalues and the Spectral Theorem

Let $\mathbf{A}$ be an $n \times n$ matrix. The determinant $|\mathbf{A}-\lambda\mathbf{I}|$ is a polynomial in the (complex) variable λ of degree n and is called the *characteristic polynomial* of $\mathbf{A}$. The equation

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

is called the *characteristic equation* of $\mathbf{A}$. By the fundamental theorem of algebra, the equation has n roots, and these roots are called the *eigenvalues* of $\mathbf{A}$.

The eigenvalues may not all be distinct. The number of times an eigenvalue occurs as a root of the characteristic equation is called the *algebraic multiplicity* of the eigenvalue.

We factor the characteristic polynomial as

$$|\mathbf{A} - \lambda\mathbf{I}| = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda). \tag{4}$$

Setting $\lambda = 0$ in (4) we see that $|\mathbf{A}|$ is just the product of the eigenvalues of $\mathbf{A}$. Similarly by equating the coefficient of λ^{n-1} on either side of (4) we see that the trace of $\mathbf{A}$ equals the sum of the eigenvalues.

A *principal submatrix* of a square matrix is a submatrix formed by a set of rows and the corresponding set of columns. A *principal minor* of $\mathbf{A}$ is the determinant of a principal submatrix.

A square matrix $\mathbf{A}$ is called *symmetric* if $\mathbf{A} = \mathbf{A}'$. An $n \times n$ matrix $\mathbf{A}$ is said to be *positive definite* if it is symmetric and if for any nonzero vector $\mathbf{x}$, $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$.

The identity matrix is clearly positive definite and so is a diagonal matrix with only positive entries along the diagonal.

8.1. *If* $\mathbf{A}$ *is positive definite, then it is nonsingular.*

PROOF. If $\mathbf{A}\mathbf{x} = \mathbf{0}$, then $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$, and since $\mathbf{A}$ is positive definite, $\mathbf{x} = \mathbf{0}$. Therefore, $\mathbf{A}$ must be nonsingular. □

The next result is obvious from the definition.

8.2. *If* $\mathbf{A}, \mathbf{B}$ *are positive definite and if* $\alpha \geq 0, \beta \geq 0$, *with* $\alpha + \beta > 0$, *then* $\alpha\mathbf{A} + \beta\mathbf{B}$ *is positive definite.*

8.3. *If* $\mathbf{A}$ *is positive definite then* $|\mathbf{A}| > 0$.

PROOF. For $0 \leq \alpha \leq 1$, define

$$f(\alpha) = |\alpha\mathbf{A} + (1 - \alpha)\mathbf{I}|.$$

By **8.2**, $\alpha\mathbf{A} + (1-\alpha)\mathbf{I}$ is positive definite, and therefore by **8.1**, $f(\alpha) \neq 0$, $0 \leq \alpha \leq 1$. Clearly, $f(0) = 1$, and since f is continuous, $f(1) = |\mathbf{A}| > 0$. □

8.4. *If* $\mathbf{A}$ *is positive definite, then any principal submatrix of* $\mathbf{A}$ *is positive definite.*

PROOF. Since $\mathbf{A}$ is positive definite, $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. Apply this condition to the set of vectors that have zeros in coordinates $j_1, \ldots, j_s$. For such a vector $\mathbf{x}$, $\mathbf{x}'\mathbf{A}\mathbf{x}$ reduces to an expression of the type $\mathbf{y}'\mathbf{B}\mathbf{y}$ where $\mathbf{B}$ is the principal submatrix of $\mathbf{A}$ formed by deleting rows and columns $j_1, \ldots, j_s$ from $\mathbf{A}$. It follows that $\mathbf{B}$, and similarly any principal submatrix of $\mathbf{A}$, is positive definite. □

A symmetric $n \times n$ matrix $\mathbf{A}$ is said to be *positive semidefinite* if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in R^n$.

8.5. *If* $\mathbf{A}$ *is a symmetric matrix, then the eigenvalues of* $\mathbf{A}$ *are all real.*

PROOF. Suppose μ is an eigenvalue of $\mathbf{A}$ and let $\mu = \alpha + i\beta$, where α, β are real and $i = \sqrt{-1}$. Since $|\mathbf{A} - \mu\mathbf{I}| = 0$, we have

$$|(\mathbf{A} - \alpha\mathbf{I}) - i\beta\mathbf{I}| = 0.$$

Taking the complex conjugate of the above determinant and multiplying the two, we get

$$|(\mathbf{A} - \alpha\mathbf{I}) - i\beta\mathbf{I}||(\mathbf{A} - \alpha\mathbf{I}) + i\beta\mathbf{I}| = 0.$$

Thus

$$|(\mathbf{A} - \alpha\mathbf{I})^2 + \beta^2\mathbf{I}| = 0. \tag{5}$$

Since $\mathbf{A}$ is symmetric, it is true that $\mathbf{A}^2$ is positive semidefinite (it follows from the definition that $\mathbf{B}\mathbf{B}'$ is positive semidefinite for any matrix $\mathbf{B}$). Thus if $\beta \neq 0$, then $|(\mathbf{A} - \alpha\mathbf{I})^2 + \beta^2\mathbf{I}|$ is positive definite, and then by **8.1**, (5) cannot hold. Thus $\beta = 0$, and μ must be real. □

If $\mathbf{A}$ is a symmetric $n \times n$ matrix, we will denote the eigenvalues of $\mathbf{A}$ by $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ and occasionally by $\lambda_1 \geq \cdots \geq \lambda_n$ if there is no possibility of confusion.

Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. Then for any i, $|\mathbf{A} - \lambda_i\mathbf{I}| = 0$ and therefore $\mathbf{A} - \lambda_i\mathbf{I}$ is singular. Thus the null space of $\mathbf{A} - \lambda_i\mathbf{I}$ has dimension at least one. This null space is called the *eigenspace* of $\mathbf{A}$ corresponding to λ_i, and any nonzero vector in the eigenspace is called an *eigenvector* of $\mathbf{A}$ corresponding to λ_i. The dimension of the null space is called the *geometric multiplicity* of λ_i.

8.6. *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix, and let* $\lambda \neq \mu$ *be eigenvalues of* $\mathbf{A}$ *with* $\mathbf{x}, \mathbf{y}$ *as corresponding eigenvectors respectively. Then* $\mathbf{x}'\mathbf{y} = 0$.

PROOF. We have $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{A}\mathbf{y} = \mu\mathbf{y}$. Therefore, $\mathbf{y}'\mathbf{A}\mathbf{x} = \mathbf{y}'(\mathbf{A}\mathbf{x}) = \lambda\mathbf{y}'\mathbf{x}$. Also, $\mathbf{y}'\mathbf{A}\mathbf{x} = (\mathbf{y}'\mathbf{A})\mathbf{x} = \mu\mathbf{y}'\mathbf{x}$. Thus $\lambda\mathbf{y}'\mathbf{x} = \mu\mathbf{y}'\mathbf{x}$. Since $\lambda \neq \mu$, it follows that $\mathbf{x}'\mathbf{y} = 0$. □

A square matrix $\mathbf{P}$ is said to be *orthogonal* if $\mathbf{P}^{-1} = \mathbf{P}'$, that is to say, if $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$. Thus an $n \times n$ matrix is orthogonal if its rows (as well as columns) form an orthonormal basis for R^n. The identity matrix is clearly orthogonal. A matrix obtained from the identity matrix by permuting its rows (and/or columns) is called a *permutation matrix* and is orthogonal as well. The product of orthogonal matrices is easily seen to be orthogonal.

8.7 (The Spectral Theorem). *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then there exists an orthogonal matrix* $\mathbf{P}$ *such that*

$$\mathbf{P}'\mathbf{A}\mathbf{P} = diag(\lambda_1, \dots, \lambda_n). \tag{6}$$

PROOF. The result is obvious for $n = 1$. Assume the result for matrices of order $n-1$ and proceed by induction. Let $\mathbf{x}$ be an eigenvector corresponding to λ_1 with $\|\mathbf{x}\| = 1$. Let $\mathbf{Q}$ be an orthogonal matrix with $\mathbf{x}$ as the first column (such a $\mathbf{Q}$ exists; first extend $\mathbf{x}$ to a basis for R^n and then apply the Gram–Schmidt process). Then

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{bmatrix}.$$

The eigenvalues of $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ are also $\lambda_1, \dots, \lambda_n$, and hence the eigenvalues of $\mathbf{B}$ are $\lambda_2, \dots, \lambda_n$. Clearly, $\mathbf{B}$ is symmetric since $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ is so. By the induction assumption there exists an orthogonal matrix $\mathbf{R}$ such that

$$\mathbf{R}'\mathbf{B}\mathbf{R} = \mathrm{diag}(\lambda_2, \dots, \lambda_n).$$

Now set

$$\mathbf{P} = \mathbf{Q}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{R} & \\ 0 & & & \end{bmatrix}.$$

Then $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. □

Suppose the matrix $\mathbf{P}$ in **8.7** has columns $\mathbf{x_1}, \dots, \mathbf{x_n}$. Then, since

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathrm{diag}(\lambda_1, \dots, \lambda_n),$$

we have $\mathbf{A}\mathbf{x_i} = \lambda_i \mathbf{x_i}$. In other words, $\mathbf{x_i}$ is an eigenvector of $\mathbf{A}$ corresponding to λ_i. Another way of writing (6) is

$$\mathbf{A} = \lambda_1 \mathbf{x_1}\mathbf{x_1'} + \cdots + \lambda_n \mathbf{x_n}\mathbf{x_n'}.$$

This is known as the *spectral decomposition* of $\mathbf{A}$.

8.8. *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then* $\mathbf{A}$ *is positive definite if and only if the eigenvalues of* $\mathbf{A}$ *are all positive.*

PROOF. By the Spectral Theorem, $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ for an orthogonal matrix $\mathbf{P}$. The result follows from the fact that $\mathbf{A}$ is positive definite if and only if $\mathbf{P}'\mathbf{A}\mathbf{P}$ is so. □

Similarly, a symmetric matrix is positive semidefinite if and only if its eigenvalues are all nonnegative.

8.9. *If* $\mathbf{A}$ *is positive semidefinite, then there exists a unique positive semidefinite matrix* $\mathbf{B}$ *such that* $\mathbf{B^2} = \mathbf{A}$. *The matrix* $\mathbf{B}$ *is called the square root of* $\mathbf{A}$ *and is denoted by* $\mathbf{A^{1/2}}$.

PROOF. There exists an orthogonal matrix $\mathbf{P}$ such that (6) holds. Since $\mathbf{A}$ is positive semidefinite, $\lambda_i \geq 0$, $i = 1, \ldots, n$. Set

$$\mathbf{B} = \mathbf{P}\mathrm{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_n^{\frac{1}{2}})\mathbf{P}'.$$

Then $\mathbf{B^2} = \mathbf{A}$.

To prove the uniqueness we must show that if $\mathbf{B}$, $\mathbf{C}$ are positive semidefinite matrices satisfying $\mathbf{A} = \mathbf{B^2} = \mathbf{C^2}$, then $\mathbf{B} = \mathbf{C}$. Let $\mathbf{D} = \mathbf{B} - \mathbf{C}$. By the spectral theorem, there exists an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Z} = \mathbf{Q}\mathbf{D}\mathbf{Q}'$ is a diagonal matrix. Let $\mathbf{E} = \mathbf{Q}\mathbf{B}\mathbf{Q}'$, $\mathbf{F} = \mathbf{Q}\mathbf{C}\mathbf{Q}'$, and it will be sufficient to show that $\mathbf{E} = \mathbf{F}$. Since $\mathbf{Z} = \mathbf{E} - \mathbf{F}$ is a diagonal matrix, $e_{ij} = f_{ij}$, $i \neq j$. Also,

$$\mathbf{EZ} + \mathbf{ZF} = \mathbf{E}(\mathbf{E} - \mathbf{F}) + (\mathbf{E} - \mathbf{F})\mathbf{F} = \mathbf{E^2} - \mathbf{F^2} = \mathbf{Q}(\mathbf{B^2} - \mathbf{C^2})\mathbf{Q}' = \mathbf{0},$$

and therefore,

$$(e_{ii} + f_{ii})z_{ii} = 0, \quad i = 1, \ldots, n.$$

If $z_{ii} = 0$, then $e_{ii} = f_{ii}$. If $z_{ii} \neq 0$, then $e_{ii} + f_{ii} = 0$. However, since $\mathbf{E}$, $\mathbf{F}$ are positive semidefinite, $e_{ii} \geq 0$, $f_{ii} \geq 0$, and it follows that $e_{ii} = f_{ii} = 0$. Thus $e_{ii} = f_{ii}$, $i = 1, \ldots, n$, and the proof is complete. □

A square matrix $\mathbf{A}$ is said to be *idempotent* if $\mathbf{A^2} = \mathbf{A}$.

8.10. *If* $\mathbf{A}$ *is idempotent, then each eigenvalue of* $\mathbf{A}$ *is either* 0 *or* 1.

PROOF. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$. Then $\lambda_1^2, \ldots, \lambda_n^2$ are the eigenvalues of $\mathbf{A^2}$ (see Exercise 12). Since $\mathbf{A} = \mathbf{A^2}$, $\{\lambda_1^2, \ldots, \lambda_n^2\} = \{\lambda_1, \ldots, \lambda_n\}$, and it follows that $\lambda_i = 0$ or 1 for each i. □

Conversely, if $\mathbf{A}$ is symmetric and if each eigenvalue of $\mathbf{A}$ is 0 or 1, then $\mathbf{A}$ is idempotent. This follows by an application of the spectral theorem.

We say that a matrix has full row (or column) rank if its rank equals the number of rows (or columns).

8.11. *If* $\mathbf{A}$ *is idempotent, then* $R(\mathbf{A}) =$ *trace* $\mathbf{A}$.

PROOF. Let $\mathbf{A} = \mathbf{BC}$ be a rank factorization. Since $\mathbf{B}$ has full column rank, it admits a left inverse by **7.1**. Similarly, $\mathbf{C}$ admits a right inverse. Let $\mathbf{B}_\ell^-$, $\mathbf{C}_\mathbf{r}^-$ be a left inverse and a right inverse of $\mathbf{B}$, $\mathbf{C}$, respectively. Then $\mathbf{A^2} = \mathbf{A}$ implies

$$\mathbf{B}_\ell^-\mathbf{BCBCC}_\mathbf{r}^- = \mathbf{B}_\ell^-\mathbf{BCC}_\mathbf{r}^- = \mathbf{I},$$

and hence $\mathbf{CB} = \mathbf{I}$, where the order of the identity matrix is the same as $R(\mathbf{A})$. Thus trace$\mathbf{A}$ = trace$\mathbf{BC}$ = trace$\mathbf{CB}$ = $R(\mathbf{A})$. □

More results on positive definite matrices and idempotent matrices are given in the Exercises.

Problems

1. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix such that the sum of the entries in any row of $\mathbf{A}$ is α. Show that α is an eigenvalue of $\mathbf{A}$. Let $\alpha_2, \ldots, \alpha_n$ be the remaining eigenvalues. What are the eigenvalues of $\mathbf{A} + \beta\mathbf{J}$, where β is a real number and $\mathbf{J}$ is a matrix of all ones?
2. Find the eigenvalues of the $n \times n$ matrix with all diagonal entries equal to a and all the remaining entries equal to b.
3. If $\mathbf{A}$ is a symmetric matrix, then show that the algebraic multiplicity of any eigenvalue of $\mathbf{A}$ equals its geometric multiplicity.
4. If $\mathbf{A}$ is a symmetric matrix, what would be a natural way to define matrices sin$\mathbf{A}$ and cos$\mathbf{A}$? Does your definition respect the identity $(\sin\mathbf{A})^2 + (\cos\mathbf{A})^2 = \mathbf{I}$?
5. Let $\mathbf{A}$ be a symmetric, nonsingular matrix. Show that $\mathbf{A}$ is positive definite if and only if $\mathbf{A}^{-1}$ is positive definite.
6. Let $\mathbf{A}$ be an $n \times n$ positive definite matrix and let $\mathbf{x} \in R^n$ with $\|\mathbf{x}\| = 1$. Show that
$$(\mathbf{x}'\mathbf{A}\mathbf{x})(\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}) \geq 1.$$
7. Let $\theta_1, \ldots, \theta_n \in [-\pi, \pi]$ and let $\mathbf{A}$ be the $n \times n$ matrix with its (i, j)-entry given by $\cos(\theta_i - \theta_j)$ for all i, j. Show that $\mathbf{A}$ is positive semidefinite. What can you say about the rank of $\mathbf{A}$?

1.9 Exercises

1. Consider the set of all vectors $\mathbf{x}$ in R^n such that $\sum_{i=1}^n x_i = 0$. Show that the set is a vector space and find a basis for the space.
2. Consider the set of all $n \times n$ matrices $\mathbf{A}$ such that trace$\mathbf{A} = 0$. Show that the set is a vector space and find its dimension.
3. Let $\mathbf{A}$ be an $n \times n$ matrix such that trace$\mathbf{AB} = 0$ for every $n \times n$ matrix $\mathbf{B}$. Can we conclude that $\mathbf{A}$ must be the zero matrix?
4. Let $\mathbf{A}$ be an $n \times n$ matrix of rank r. If rows $i_1, \ldots, i_r$ are linearly independent and if columns $j_1, \ldots, j_r$ are linearly independent, then show that the $r \times r$ submatrix formed by these rows and columns has rank r.
5. For any matrix $\mathbf{A}$, show that $\mathbf{A} = \mathbf{0}$ if and only if trace$\mathbf{A}'\mathbf{A} = 0$.
6. Let $\mathbf{A}$ be a square matrix. Prove that the following conditions are equivalent: (i) $\mathbf{A}$ is symmetric; (ii) $\mathbf{A}^2 = \mathbf{AA}'$; (iii) trace$\mathbf{A}^2$ = trace$\mathbf{AA}'$; (iv) $\mathbf{A}^2 = \mathbf{A}'\mathbf{A}$; (v) trace$\mathbf{A}^2$ = trace$\mathbf{A}'\mathbf{A}$.

7. Let $\mathbf{A}$ be a square matrix with all row sums equal to 1. If $\mathbf{AA}' = \mathbf{A}'\mathbf{A}$, then show that the column sums of $\mathbf{A}$ are also equal to 1.

8. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ be $n \times n$ matrices such that the matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

has rank n. Show that $|\mathbf{AD}| = |\mathbf{BC}|$.

9. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices such that $R(\mathbf{AB}) = R(\mathbf{B})$. Show that the following type of cancellation is valid: Whenever $\mathbf{ABX} = \mathbf{ABY}$, then $\mathbf{BX} = \mathbf{BY}$.

10. Let $\mathbf{A}$ be an $n \times n$ matrix such that $R(\mathbf{A}) = R(\mathbf{A}^2)$. Show that $R(\mathbf{A}) = R(\mathbf{A}^k)$ for any positive integer k.

11. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices such that $\mathbf{AB} = \mathbf{0}$. Show that $R(\mathbf{A}) + R(\mathbf{B}) \le n$.

12. If $\mathbf{A}$ has eigenvalues $\lambda_1, \dots, \lambda_n$, then show that $\mathbf{A}^2$ has eigenvalues $\lambda_1^2, \dots, \lambda_n^2$.

13. If $\mathbf{A}, \mathbf{B}$ are $n \times n$ matrices, then show that $\mathbf{AB}$ and $\mathbf{BA}$ have the same eigenvalues.

14. Let $\mathbf{A}, \mathbf{B}$ be matrices of order $m \times n$, $n \times m$, respectively. Consider the identity

$$\begin{bmatrix} \mathbf{I_m} - \mathbf{AB} & \mathbf{A} \\ \mathbf{0} & \mathbf{I_n} \end{bmatrix} \begin{bmatrix} \mathbf{I_m} & \mathbf{0} \\ \mathbf{B} & \mathbf{I_m} \end{bmatrix} = \begin{bmatrix} \mathbf{I_m} & \mathbf{0} \\ \mathbf{B} & \mathbf{I_n} \end{bmatrix} \begin{bmatrix} \mathbf{I_m} & \mathbf{A} \\ \mathbf{0} & \mathbf{I_n} - \mathbf{BA} \end{bmatrix}$$

and show that

$$|\mathbf{I_m} - \mathbf{AB}| = |\mathbf{I_n} - \mathbf{BA}|.$$

Now obtain a relationship between the characteristic polynomoials of $\mathbf{AB}$ and $\mathbf{BA}$. Conclude that the nonzero eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ are the same.

15. If $\mathbf{S}$ is a nonsingular matrix, then show that $\mathbf{A}$ and $\mathbf{S}^{-1}\mathbf{AS}$ have the same eigenvalues.

16. Suppose $\mathbf{A}$ is an $n \times n$ matrix, and let

$$|\mathbf{A} - \lambda\mathbf{I}| = c_0 - c_1\lambda + c_2\lambda^2 - \cdots + c_n(-1)^n\lambda^n$$

be the characteristic polynomial of $\mathbf{A}$. The Cayley–Hamilton theorem asserts that $\mathbf{A}$ satisfies its characteristic equation, i.e.,

$$c_0\mathbf{I} - c_1\mathbf{A} + c_2\mathbf{A^2} - \cdots + c_n(-1)^n\mathbf{A^n} = \mathbf{0}.$$

Prove the theorem for a diagonal matrix. Then prove the theorem for any symmetric matrix.

17. Prove the following: If $\mathbf{A} = \mathbf{B}'\mathbf{B}$ for some matrix $\mathbf{B}$, then $\mathbf{A}$ is positive semidefinite. Further, $\mathbf{A}$ is positive definite if $\mathbf{B}$ has full column rank.

18. Prove the following: (i) If $\mathbf{A}$ is positive semidefinite, then $|\mathbf{A}| \ge 0$. (ii) If $\mathbf{A}$ is positive semidefinite, then all principal minors of $\mathbf{A}$ are nonnegative. (iii) Suppose $\mathbf{A}$ is positive semidefinite. Then $\mathbf{A}$ is positive definite if and only if it is nonsingular.

19. For any matrix $\mathbf{X}$, show that $R(\mathbf{X}'\mathbf{X}) = R(\mathbf{X})$. If $\mathbf{A}$ is positive definite, then show that $R(\mathbf{X}'\mathbf{AX}) = R(\mathbf{X})$ for any $\mathbf{X}$.

20. Let $\mathbf{A}$ be a square matrix such that $\mathbf{A}+\mathbf{A}'$ is positive definite. Then prove that $\mathbf{A}$ is nonsingular.

21. If $\mathbf{A}$ is symmetric, then show that $R(\mathbf{A})$ equals the number of nonzero eigenvalues of $\mathbf{A}$, counting multiplicity.

22. Let $\mathbf{A}$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $1 \le k \le n$. Show that

$$\sum_{i_1 < \cdots < i_k} \lambda_{i_1} \cdots \lambda_{i_k}$$

equals the sum of the $k \times k$ principal minors of $\mathbf{A}$.

23. If $\mathbf{A}$ is a symmetric matrix of rank r, then prove that $\mathbf{A}$ has a principal submatrix of order r that is nonsingular.

24. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices such that $\mathbf{A}$ is positive definite and $\mathbf{B}$ is symmetric. Show that there exists a nonsingular matrix $\mathbf{E}$ such that $\mathbf{E}'\mathbf{A}\mathbf{E} = \mathbf{I}$ and $\mathbf{E}'\mathbf{B}\mathbf{E}$ is diagonal. Conclude that if $\mathbf{A}$ is positive definite, then there exists a nonsingular matrix $\mathbf{E}$ such that $\mathbf{E}'\mathbf{A}\mathbf{E} = \mathbf{I}$.

25. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices where $\mathbf{A}$ is symmetric and $\mathbf{B}$ is positive semidefinite. Show that $\mathbf{AB}$ has only real eigenvalues. If $\mathbf{A}$ is positive semidefinite, then show that $\mathbf{AB}$ has only nonnegative eigenvalues.

26. Suppose $\mathbf{A}$ is a symmetric $n \times n$ matrix. Show that $\mathbf{A}$ is positive semidefinite if and only if trace $(\mathbf{AB}) \ge 0$ for any positive semidefinite matrix $\mathbf{B}$.

27. Let S, T be subspaces of R^n. Define $S + T$ as the set of vectors of the form $\mathbf{x} + \mathbf{y}$ where $\mathbf{x} \in S$, $\mathbf{y} \in T$. Prove that $S + T$ is a subspace and that

$$\dim(S + T) = \dim(S) + \dim(T) - \dim(S \cap T).$$

28. Let S, T be subspaces of R^n such that $\dim(S) + \dim(T) > n$. Show that $\dim(S \cap T) \ge 1$.

29. Let $X = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$, $Y = \{\mathbf{y_1}, \ldots, \mathbf{y_n}\}$ be bases for R^n and let $S \subset X$ be a set of cardinality r, $1 \le r \le n$. Show that there exists $T \subset Y$ of cardinality r such that $(X \setminus S) \cup T$ is a basis for R^n.

30. Let $\mathbf{A}, \mathbf{B}$ be matrices of order $m \times n$, $n \times q$, respectively. Show that

$$R(\mathbf{AB}) = R(\mathbf{B}) - \dim(\mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B})).$$

31. Let $\mathbf{A}$ be an $m \times n$ matrix and let $\mathbf{B}$ be an $r \times s$ submatrix of $\mathbf{A}$. Show that $r + s - R(\mathbf{B}) \le m + n - R(\mathbf{A})$.

32. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ positive semidefinite matrices and let $\mathbf{C}$ be the matrix with its (i, j)-entry given by $c_{ij} = a_{ij} b_{ij}$, $i, j, = 1, \ldots, n$. Show that $\mathbf{C}$ is positive semidefinite.

33. Let $x_1, \ldots, x_n$ be positive numbers. Show that the $n \times n$ matrix with its (i, j)-entry $\frac{1}{x_i + x_j}$, $i, j = 1, \ldots, n$, is positive semidefinite.

34. Let $\mathbf{X}, \mathbf{Y}$ be $n \times n$ symmetric matrices such that $\mathbf{X}$ is positive definite and $\mathbf{XY} + \mathbf{YX}$ is positive semidefinite. Show that $\mathbf{Y}$ must be positive semidefinite.

35. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices such that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A} - \mathbf{B}$ are positive semidefinite. Show that $\mathbf{A}^{\frac{1}{2}} - \mathbf{B}^{\frac{1}{2}}$ is positive semidefinite.

1.10 Hints and Solutions

Section 4

1. Answer: The rank is 2 if $\alpha = -7$ and is 3 otherwise.

2. Let $\mathbf{A}$, $\mathbf{B}$ be matrices with columns

$$\{\mathbf{x_1}, \ldots, \mathbf{x_p}\}, \quad \{\mathbf{y_1}, \ldots, \mathbf{y_q}\},$$

respectively. If $\{\mathbf{x_1}, \ldots, \mathbf{x_p}, \mathbf{y_i}\}$ is linearly dependent for each i, then each $\mathbf{y_i}$ is a linear combination of $\{\mathbf{x_1}, \ldots, \mathbf{x_p}\}$ and hence $\mathbf{B} = \mathbf{AC}$ for some matrix $\mathbf{C}$. Now

$$q = R(\mathbf{B}) \leq R(\mathbf{A}) = p$$

gives a contradiction.

3. We may assume $k \leq \min\{m, n\}$, for otherwise the result is trivial. Let $\mathbf{C} = \mathbf{B} - \mathbf{A}$. Then $\mathbf{C}$ has at most k nonzero entries, and thus $R(\mathbf{C}) \leq k$. (To see this, note that a set of s rows of $\mathbf{C}$, where $s > k$, must have a zero row and hence is linearly dependent.) Since $R(\mathbf{B}) = R(\mathbf{A} + \mathbf{C}) \leq R(\mathbf{A}) + R(\mathbf{C}) \leq R(\mathbf{A}) + k$, we get the second inequality in the exercise. The first one follows similarly.

Section 5

1. Answer: Only (ii) defines an inner product.

2. Answer: We get the following orthonormal basis (rounded to three decimal places):

$$\begin{bmatrix} 0.535 & 0.802 & -0.267 \end{bmatrix}, \quad \begin{bmatrix} 0.835 & -0.452 & 0.313 \end{bmatrix},$$
$$\begin{bmatrix} -0.131 & 0.391 & 0.911 \end{bmatrix}.$$

Section 7

1. Hint: Let $\mathbf{X} = \mathbf{UV}$ be a rank factorization. Then $\mathbf{AXB} = (\mathbf{AU})(\mathbf{VB})$. Now deduce the result from the Frobenius inequality.

Section 8

1. Hint: Note that α is an eigenvalue, since the vector of all ones is an eigenvector for it. The eigenvalues of $\mathbf{A} + \beta\mathbf{J}$ are given by $\alpha + n\beta, \alpha_2, \ldots, \alpha_n$. This can be seen using the spectral theorem and the fact that the eigenvectors of $\mathbf{A}$ for $\alpha_2, \ldots, \alpha_n$ can be taken to be orthogonal to the vector of all ones.

2. Answer: $a + (n-1)b$ and $a - b$ with multiplicities $1, n-1$, respectively.

Section 9

4. Let $\mathbf{B}$ be the submatrix of $\mathbf{A}$ formed by rows $i_1, \ldots, i_r$ and let $\mathbf{C}$ be the submatrix of $\mathbf{A}$ formed by columns $j_1, \ldots, j_r$. Also, let $\mathbf{D}$ be the submatrix formed by rows $i_1, \ldots, i_r$ and columns $j_1, \ldots, j_r$. If $R(\mathbf{D}) < r$, then, since $R(\mathbf{B}) = r$, there exists a column of $\mathbf{B}$ that is not a linear combination of columns of $\mathbf{D}$. Then the corresponding column of $\mathbf{A}$ is not a linear combination of columns of $\mathbf{C}$. This is a contradiction, since coulmns of $\mathbf{C}$ form a basis for the column space of $\mathbf{A}$.

5. Hint: trace$\mathbf{A}'\mathbf{A}$ equals the sum of squares of the entries of $\mathbf{A}$.

6. Hint: Let $\mathbf{B} = \mathbf{A} - \mathbf{A}'$. Then $\mathbf{A}$ is symmetric if and only if $\mathbf{B} = \mathbf{0}$, which, by the preceding exercise, happens if and only if trace$\mathbf{B}\mathbf{B}' = 0$. Expand trace$\mathbf{B}\mathbf{B}'$.

7. Let $\mathbf{1}$ be the column vector of all ones and let $\mathbf{x} = \mathbf{A}'\mathbf{1} - \mathbf{1}$. We must show $\mathbf{x} = \mathbf{0}$ and for this it is sufficient to show $\mathbf{x}'\mathbf{x} = \mathbf{0}$. Now $\mathbf{x}'\mathbf{x}$ equals, by expansion, $\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} - \mathbf{1}'\mathbf{A}\mathbf{1} - \mathbf{1}'\mathbf{A}'\mathbf{1} + \mathbf{1}'\mathbf{1}$. This is seen to be zero, since $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A}$ and each row sum of $\mathbf{A}$ is 1.

8. If $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are all singular, the result is trivial. So assume, without loss of generality, that $\mathbf{A}$ is nonsingular. Then the rank of $\mathbf{A}$ as well as that of the partitioned matrix being n, the last n columns are linear combinations of the first n. Thus there exists an $n \times n$ matrix $\mathbf{X}$ such that $\mathbf{B} = \mathbf{A}\mathbf{X}$, $\mathbf{D} = \mathbf{C}\mathbf{X}$. Then $|\mathbf{A}\mathbf{D}| = |\mathbf{A}||\mathbf{C}||\mathbf{X}| = |\mathbf{B}\mathbf{C}|$.

10. Clearly, $\mathcal{C}(\mathbf{A^2}) \subset \mathcal{C}(\mathbf{A})$, and since $R(\mathbf{A}) = R(\mathbf{A^2})$, the spaces are equal. Thus $\mathbf{A} = \mathbf{A^2}\mathbf{X}$ for some matrix $\mathbf{X}$. Now, $\mathbf{A^2} = \mathbf{A^3}\mathbf{X}$, and thus $R(\mathbf{A^2}) \leq R(\mathbf{A^3})$. Since $R(\mathbf{A^3}) \leq R(\mathbf{A^2})$, the ranks must be equal. The general case is proved by induction.

12. The following "proof," which is often given, is incomplete: If λ is an eigenvalue of $\mathbf{A}$, then $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some nonzero (complex) vector $\mathbf{x}$. Then $\mathbf{A^2}\mathbf{x} = \lambda^2\mathbf{x}$. Hence λ^2 is an eigenvalue of $\mathbf{A^2}$. Thus we have proved that if λ is an eigenvalue of $\mathbf{A}$, then λ^2 is an eigenvalue of $\mathbf{A^2}$. However, this does not rule out the possibility of, for example, $\mathbf{A}$ of order 3×3 having eigenvalues $1, 1, -1$ and $\mathbf{A^2}$ having eigenvalues $1, 2, 4$. We now give a proof. We have $|\mathbf{A} - \lambda\mathbf{I}| = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda)$ and hence, replacing λ by $-\lambda$, $|\mathbf{A} + \lambda\mathbf{I}| = (\lambda_1 + \lambda)\cdots(\lambda_n + \lambda)$. Multiplying these equations we get, setting $\lambda^2 = \mu$,

$$|\mathbf{A^2} - \mu\mathbf{I}| = (\lambda_1^2 - \mu)\cdots(\lambda_n^2 - \mu),$$

and the result follows.

13. If $\mathbf{A}$ is nonsingular, then

$$|\mathbf{A}\mathbf{B} - \lambda\mathbf{I}| = |\mathbf{A^{-1}}(\mathbf{A}\mathbf{B} - \lambda\mathbf{I})\mathbf{A}| = |\mathbf{B}\mathbf{A} - \lambda\mathbf{I}|.$$

Thus $\mathbf{A}\mathbf{B}$, $\mathbf{B}\mathbf{A}$ have the same characteristic polynomial and hence the same eigenvalues. To settle the general case first observe that if a sequence of matrices $\mathbf{X_k}$ converges (entrywise) to the matrix $\mathbf{X}$, then the eigenvalues of $\mathbf{X_k}$ can be labeled, say $\lambda_1^k, \ldots, \lambda_n^k$, so that λ_i^k approaches λ_i, $i = 1, \ldots, n$, where λ_i, $i = 1, \ldots, n$, are the eigenvalues of $\mathbf{X}$. This follows from the more general fact that the roots of a polynomial are continuous functions of its coefficients. Now, if $\mathbf{A}$ is singular, we may construct a sequence of nonsingular matrices with limit $\mathbf{A}$ and use a continuity argument. See the next exercise for a different proof.

16. Hint: Use the spectral theorem to deduce the general case from the diagonal case.

19. If $\mathbf{Xz} = \mathbf{0}$ for some $\mathbf{z}$, then clearly $\mathbf{X'Xz} = \mathbf{0}$. Conversely, if $\mathbf{X'Xz} = \mathbf{0}$, then $\mathbf{z'X'Xz} = 0$, and it follows that $\mathbf{Xz} = \mathbf{0}$. Thus $\mathbf{X}$ and $\mathbf{X'X}$ have the same null space, and by **5.6**, Chapter 1, they have the same rank. (We may similarly prove $R(\mathbf{XX'}) = R(\mathbf{X})$.) Now, if $\mathbf{A}$ is positive definite, then $R(\mathbf{X'AX}) = R(\mathbf{X'A}^{\frac{1}{2}})$ by the first part, which equals $R(\mathbf{X})$, since $\mathbf{A}^{\frac{1}{2}}$ is nonsingular.

20. If $\mathbf{Ax} = \mathbf{0}$, then $\mathbf{x'A'} = \mathbf{0}$, and hence $\mathbf{x'(A + A')x} = 0$. Since $\mathbf{A}+\mathbf{A'}$ is positive definite, $\mathbf{x} = \mathbf{0}$. It follows that $\mathbf{A}$ is nonsingular.

22. Hint: Equate the coefficients of λ^{n-k} on either side of the equation

$$|\mathbf{A} - \lambda\mathbf{I}| = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda).$$

23. Hint: Use the two preceding exercises.

24. By the spectral theorem, there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P'AP} = \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \mathbf{D}$, say, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\mathbf{A}$. Since $\mathbf{A}$ is positive definite, $\lambda_i > 0, i = 1, \dots, n$, and $\mathbf{D}^{-\frac{1}{2}}$ is well-defined. Then $\mathbf{D}^{-\frac{1}{2}}\mathbf{P'APD}^{-\frac{1}{2}} = \mathbf{I}$. The matrix $\mathbf{D}^{-\frac{1}{2}}\mathbf{P'BPD}^{-\frac{1}{2}}$ is symmetric, and again by the spectral theorem there exists an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Q'D}^{-\frac{1}{2}}\mathbf{P'BPD}^{-\frac{1}{2}}\mathbf{Q}$ is a diagonal matrix. Set $\mathbf{E} = \mathbf{PD}^{-\frac{1}{2}}\mathbf{Q}$. Then $\mathbf{E}$ is nonsingular, $\mathbf{E'AE} = \mathbf{I}$, and $\mathbf{E'BE}$ is diagonal.

25. By Exercise 13, $\mathbf{AB} = \mathbf{AB}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}$ has the same eigenvalues as $\mathbf{B}^{\frac{1}{2}}\mathbf{AB}^{\frac{1}{2}}$, but the latter matrix is symmetric and hence has only real eigenvalues. If $\mathbf{A}$ is positive semidefinite, then so is $\mathbf{B}^{\frac{1}{2}}\mathbf{AB}^{\frac{1}{2}}$, and has only nonnegative eigenvalues.

27. Clearly, $S + T$ is a subspace, since it is closed under addition and scalar multiplication. Let $\mathbf{x_1}, \dots, \mathbf{x_p}$ be a basis for $S \cap T$. Then there exists a basis

$$\mathbf{x_1}, \dots, \mathbf{x_p}, \quad \mathbf{y_1}, \dots, \mathbf{y_q}$$

for S and a basis

$$\mathbf{x_1}, \dots, \mathbf{x_p}, \quad \mathbf{z_1}, \dots, \mathbf{z_r}$$

for T. We show that

$$\mathbf{x_1}, \dots, \mathbf{x_p}, \quad \mathbf{y_1}, \dots, \mathbf{y_q}, \quad \mathbf{z_1}, \dots, \mathbf{z_r}$$

is a basis for $S + T$. Obviously, the set spans $S + T$. We now show that the set is linearly independent. Suppose

$$u_1\mathbf{x_1} + \cdots + u_p\mathbf{x_p} + v_1\mathbf{y_1} + \cdots + v_q\mathbf{y_q} + w_1\mathbf{z_1} + \cdots + w_r\mathbf{z_r} = \mathbf{0}. \quad (7)$$

Thus $w_1\mathbf{z_1} + \cdots + w_r\mathbf{z_r}$, which belongs to T, can be expressed as a linear combination of $\mathbf{x_1}, \dots, \mathbf{x_p}, \mathbf{y_1}, \dots, \mathbf{y_q}$, and hence it belongs to $S \cap T$. Thus there exist $\alpha_1, \cdots, \alpha_p$ such that $w_1\mathbf{z_1} + \cdots + w_r\mathbf{z_r} + \alpha_1\mathbf{x_1} + \cdots + \alpha_p\mathbf{x_p} = \mathbf{0}$. Since $\mathbf{x_1}, \dots, \mathbf{x_p}, \mathbf{z_1}, \dots, \mathbf{z_r}$ are linearly independent, it follows that $w_1 = \cdots = w_r = 0$. We can similarly show that $v_1 = \cdots = v_q = 0$, and then it follows from (7) that $u_1 = \cdots = u_p = 0$. Hence the set is linearly independent and the proof is complete.

29. Without loss of generality, we take $S = \{\mathbf{x_1}, \dots, \mathbf{x_r}\}$. Let $\mathbf{A}$ be the $n \times n$ matrix such that $\mathbf{y_i} = \sum_{j=1}^{n} a_{ij}\mathbf{x_j}$, $i = 1, \dots, n$. Then $\mathbf{A}$ is nonsingular and hence $|\mathbf{A}| \neq 0$. Expanding $|\mathbf{A}|$ along the first r columns, we see that there exists an $r \times r$ nonsingular submatrix of $\mathbf{A}$, formed by rows, say, $i_1, \dots, i_r$ and columns $1, \dots, r$. Let $T = \{\mathbf{y_{i_1}}, \dots, \mathbf{y_{i_r}}\}$. Let $\mathbf{B}$ be the $n \times n$ matrix defined as follows. The first r rows of $\mathbf{B}$ are identical to rows $i_1, \dots, i_r$ of $\mathbf{A}$, while the last $n - r$ rows of $\mathbf{B}$ are identical to the last $n - r$ rows of $\mathbf{I_n}$. Then $|\mathbf{B}| \neq 0$ and hence $\mathbf{B}$ is nonsingular. Let $(X \setminus S) \cup T = \{\mathbf{y_{i_1}}, \dots, \mathbf{y_{i_r}}, \mathbf{x_{r+1}}, \dots, \mathbf{x_n}\} = \{\mathbf{u_1}, \dots, \mathbf{u_n}\}$. Then $\mathbf{u_i} = \sum_{j=1}^{n} b_{ij}\mathbf{x_j}$, $i = 1, \dots, n$, and hence $\{\mathbf{u_1}, \dots, \mathbf{u_n}\}$ is a basis for R^n.

30. By **7.3** we may assume, without loss of generality, that $\mathbf{B} = \begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $R(\mathbf{B}) = r$. Partition $\mathbf{A} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix}$ conformally. Then $\mathbf{AB} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{0} \\ \mathbf{A_{21}} & \mathbf{0} \end{bmatrix}$. By **5.6**, $R(\mathbf{AB}) = R\begin{bmatrix} \mathbf{A_{11}} \\ \mathbf{A_{21}} \end{bmatrix} = r - \dim \mathcal{N}\begin{bmatrix} \mathbf{A_{11}} \\ \mathbf{A_{21}} \end{bmatrix}$. Now observe that $\mathcal{N}\begin{bmatrix} \mathbf{A_{11}} \\ \mathbf{A_{21}} \end{bmatrix} = \mathcal{N}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B})$.

31. Assume, without loss of generality, that $\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix}$. Then $R(\mathbf{A}) \leq R[\mathbf{B}, \mathbf{C}] + R[\mathbf{D}, \mathbf{E}] \leq R(\mathbf{B}) + R(\mathbf{C}) + m - r \leq R(\mathbf{B}) + n - s + m - r$, and the result follows.

32. Hint: First suppose $\mathbf{B}$ has rank one. Then there exist $u_1, \dots, u_n$ such that $b_{ij} = u_i u_j$, $i, j = 1, \dots, n$. Then $\mathbf{C} = \mathbf{UAU}$ where $\mathbf{U} = \operatorname{diag}(u_1, \dots, u_n)$, and hence $\mathbf{C}$ is positive semidefinite. The general case is obtained using the spectral decomposition of $\mathbf{B}$.

33. Hint: For $t > 0$, the $n \times n$ matrix $(t^{x_i + x_j})$ is positive semidefinite. Now use the fact that $\frac{1}{x_i + x_j} = \int_0^1 t^{x_i + x_j - 1} dt$.

34. Using the spectral theorem we may assume, without loss of generality, that $\mathbf{X} = \operatorname{diag}(x_1, \dots, x_n)$. Let $\mathbf{XY} + \mathbf{YX} = \mathbf{Z}$. Then $y_{ij}(x_i + x_j) = z_{ij}$, and hence $y_{ij} = \frac{z_{ij}}{x_i + x_j}$ for all i, j. Now use the preceding two exercises.

35. Hint: Let $\mathbf{X} = (\mathbf{A}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}})$, $\mathbf{Y} = (\mathbf{A}^{\frac{1}{2}} - \mathbf{B}^{\frac{1}{2}})$. Then $\mathbf{XY} + \mathbf{YX} = 2(\mathbf{A} - \mathbf{B})$, which is positive semidefinite. Now use the preceding exercise.

2

Linear Estimation

2.1 Generalized Inverses

Let $\mathbf{A}$ be an $m \times n$ matrix. A matrix $\mathbf{G}$ of order $n \times m$ is said to be a *generalized inverse* (or a *g-inverse*) of $\mathbf{A}$ if $\mathbf{AGA} = \mathbf{A}$.

If $\mathbf{A}$ is square and nonsingular, then $\mathbf{A}^{-1}$ is the unique g-inverse of $\mathbf{A}$. Otherwise, $\mathbf{A}$ has infinitely many g-inverses, as we will see shortly.

1.1. *Let* $\mathbf{A}, \mathbf{G}$ *be matrices of order* $m \times n$ *and* $n \times m$ *respectively. Then the following conditions are equivalent:*

(i) $\mathbf{G}$ *is a g-inverse of* $\mathbf{A}$.
(ii) *For any* $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, $\mathbf{x} = \mathbf{Gy}$ *is a solution of* $\mathbf{Ax} = \mathbf{y}$.

PROOF. (i) $\Rightarrow$ (ii). Any $\mathbf{y} \in \mathcal{C}(\mathbf{A})$ is of the form $\mathbf{y} = \mathbf{Az}$ for some $\mathbf{z}$. Then $\mathbf{A}(\mathbf{Gy}) = \mathbf{AGAz} = \mathbf{Az} = \mathbf{y}$.

(ii) $\Rightarrow$ (i). Since $\mathbf{AGy} = \mathbf{y}$ for any $\mathbf{y} \in \mathcal{C}(\mathbf{A})$ we have $\mathbf{AGAz} = \mathbf{Az}$ for all $\mathbf{z}$. In particular, if we let $\mathbf{z}$ be the ith column of the identity matrix, then we see that the ith columns of $\mathbf{AGA}$ and $\mathbf{A}$ are identical. Therefore, $\mathbf{AGA} = \mathbf{A}$. □

Let $\mathbf{A} = \mathbf{BC}$ be a rank factorization. We have seen that $\mathbf{B}$ admits a left inverse $\mathbf{B}_\ell^-$, and $\mathbf{C}$ admits a right inverse $\mathbf{C}_\mathbf{r}^-$. Then $\mathbf{G} = \mathbf{C}_\mathbf{r}^-\mathbf{B}_\ell^-$ is a g-inverse of $\mathbf{A}$, since

$$\mathbf{AGA} = \mathbf{BC}(\mathbf{C}_\mathbf{r}^-\mathbf{B}_\ell^-)\mathbf{BC} = \mathbf{BC} = \mathbf{A}.$$

Alternatively, if $\mathbf{A}$ has rank r, then by **7.3** of Chapter 1 there exist nonsingular matrices $\mathbf{P}$, $\mathbf{Q}$ such that

$$\mathbf{A} = \mathbf{P}\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}.$$

It can be verified that for *any* $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ of appropriate dimensions,

$$\begin{bmatrix} \mathbf{I}_r & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$$

is a g-inverse of

$$\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Then

$$\mathbf{G} = \mathbf{Q}^{-1}\begin{bmatrix} \mathbf{I}_r & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}\mathbf{P}^{-1}$$

is a g-inverse of $\mathbf{A}$. This also shows that any matrix that is not a square nonsingular matrix admits infinitely many g-inverses.

Another method that is particularly suitable for computing a g-inverse is as follows. Let $\mathbf{A}$ be of rank r. Choose any $r \times r$ nonsingular submatrix of $\mathbf{A}$. For convenience let us assume

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

where $\mathbf{A}_{11}$ is $r \times r$ and nonsingular. Since $\mathbf{A}$ has rank r, there exists a matrix $\mathbf{X}$ such that $\mathbf{A}_{12} = \mathbf{A}_{11}\mathbf{X}$, $\mathbf{A}_{22} = \mathbf{A}_{21}\mathbf{X}$. Now it can be verified that the $n \times m$ matrix $\mathbf{G}$ defined as

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

is a g-inverse of $\mathbf{A}$. (Just multiply $\mathbf{AGA}$ out.) We will often use the notation $\mathbf{A}^-$ to denote a g-inverse of $\mathbf{A}$.

1.2. *If* $\mathbf{G}$ *is a g-inverse of* $\mathbf{A}$*, then* $R(\mathbf{A}) = R(\mathbf{AG}) = R(\mathbf{GA})$.

PROOF. $R(\mathbf{A}) = R(\mathbf{AGA}) \leq R(\mathbf{AG}) \leq R(\mathbf{A})$. The second part follows similarly. □

A g-inverse of $\mathbf{A}$ is called a *reflexive g-inverse* if it also satisfies $\mathbf{GAG} = \mathbf{G}$. Observe that if $\mathbf{G}$ is any g-inverse of $\mathbf{A}$, then $\mathbf{GAG}$ is a reflexive g-inverse of $\mathbf{A}$.

1.3. *Let* $\mathbf{G}$ *be a g-inverse of* $\mathbf{A}$. *Then* $R(\mathbf{A}) \leq R(\mathbf{G})$. *Furthermore, equality holds if and only if* $\mathbf{G}$ *is reflexive.*

PROOF. For any g-inverse $\mathbf{G}$ we have $R(\mathbf{A}) = R(\mathbf{AGA}) \leq R(\mathbf{G})$. If $\mathbf{G}$ is reflexive, then $R(\mathbf{G}) = R(\mathbf{GAG}) \leq R(\mathbf{A})$ and hence $R(\mathbf{A}) = R(\mathbf{G})$.

Conversely, suppose $R(\mathbf{A}) = R(\mathbf{G})$. First observe that $\mathcal{C}(\mathbf{GA}) \subset \mathcal{C}(\mathbf{G})$. By **1.2**, $R(\mathbf{G}) = R(\mathbf{GA})$, and hence $\mathcal{C}(\mathbf{G}) = \mathcal{C}(\mathbf{GA})$. Therefore, $\mathbf{G} = \mathbf{GAX}$ for some $\mathbf{X}$. Now,

$$\mathbf{GAG} = \mathbf{GAGAX} = \mathbf{GAX} = \mathbf{G},$$

and $\mathbf{G}$ is reflexive. □

1.4. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix, let* $\mathbf{G}$ *be a g-inverse of* $\mathbf{A}$ *and let* $\mathbf{y} \in \mathcal{C}(\mathbf{A})$. *Then the class of solutions of* $\mathbf{Ax} = \mathbf{y}$ *is given by* $\mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$, *where* $\mathbf{z}$ *is arbitrary.*

PROOF. For any $\mathbf{z}$,

$$\mathbf{A}\{\mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}\} = \mathbf{AGy} = \mathbf{y},$$

since $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, and hence $\mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$ is a solution. Conversely, if $\mathbf{u}$ is a solution, then set $\mathbf{z} = \mathbf{u} - \mathbf{Gy}$ and verify that

$$\mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z} = \mathbf{u}.$$

That completes the proof. □

A g-inverse $\mathbf{G}$ of $\mathbf{A}$ is said to be a *minimum norm g-inverse* of $\mathbf{A}$ if in addition to $\mathbf{AGA} = \mathbf{A}$, it satisfies $(\mathbf{GA})' = \mathbf{GA}$. The reason for this terminology will be clear from the next result.

1.5. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix. Then the following conditions are equivalent:*

(i) $\mathbf{G}$ *is a minimum norm g-inverse of* $\mathbf{A}$.
(ii) *For any* $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, $\mathbf{x} = \mathbf{Gy}$ *is a solution of* $\mathbf{Ax} = \mathbf{y}$ *with minimum norm.*

PROOF. (i) ⇒ (ii). In view of **1.4** we must show that

$$\|\mathbf{Gy}\| \leq \|\mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}\| \tag{1}$$

for any $\mathbf{y} \in \mathcal{C}(\mathbf{A})$ and for any $\mathbf{z}$.

We have

$$\|\mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}\|^2 = \|\mathbf{Gy}\|^2 + \|(\mathbf{I} - \mathbf{GA})\mathbf{z}\|^2 + 2\mathbf{y}'\mathbf{G}'(\mathbf{I} - \mathbf{GA})\mathbf{z}. \tag{2}$$

Since $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, then $\mathbf{y} = \mathbf{Au}$ for some $\mathbf{u}$. Hence

$$\begin{aligned} \mathbf{y}'\mathbf{G}'(\mathbf{I} - \mathbf{GA})\mathbf{z} &= \mathbf{u}'\mathbf{A}'\mathbf{G}'(\mathbf{I} - \mathbf{GA})\mathbf{z} \\ &= \mathbf{u}'\mathbf{GA}(\mathbf{I} - \mathbf{GA})\mathbf{z} \\ &= 0. \end{aligned}$$

Inserting this in (2) we get (1).

(ii) ⇒ (i). Since for any $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, $\mathbf{x} = \mathbf{Gy}$ is a solution of $\mathbf{Ax} = \mathbf{y}$, by **1.1**, $\mathbf{G}$ is a g-inverse of $\mathbf{A}$. Now we have (1) for all $\mathbf{z}$, and therefore for all $\mathbf{u}$, $\mathbf{z}$,

$$0 \leq \|(\mathbf{I} - \mathbf{GA})\mathbf{z}\|^2 + 2\mathbf{u}'\mathbf{A}'\mathbf{G}'(\mathbf{I} - \mathbf{GA})\mathbf{z}. \tag{3}$$

Replace $\mathbf{u}$ by $\alpha\mathbf{u}$ in (3). If $\mathbf{u}'\mathbf{A}'\mathbf{G}'(\mathbf{I}-\mathbf{GA})\mathbf{z} < 0$, then choosing α large and positive we get a contradiction to (3). Similarly, if $\mathbf{u}'\mathbf{A}'\mathbf{G}'(\mathbf{I}-\mathbf{GA})\mathbf{z} > 0$, then choosing α large and negative we get a contradiction. We therefore conclude that

$$\mathbf{u}'\mathbf{A}'\mathbf{G}'(\mathbf{I}-\mathbf{GA})\mathbf{z} = 0$$

for all $\mathbf{u}, \mathbf{z}$ and hence $\mathbf{A}'\mathbf{G}'(\mathbf{I}-\mathbf{GA}) = \mathbf{0}$. Thus $\mathbf{A}'\mathbf{G}'$ equals $(\mathbf{GA})'\mathbf{GA}$, which is symmetric. $\square$

A g-inverse $\mathbf{G}$ of $\mathbf{A}$ is said to be a *least squares g-inverse* of $\mathbf{A}$ if in addition to $\mathbf{AGA} = \mathbf{A}$, it satisfies $(\mathbf{AG})' = \mathbf{AG}$.

1.6. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix. Then the following conditions are equivalent:*

(i) $\mathbf{G}$ *is a least squares g-inverse of* $\mathbf{A}$.
(ii) *For any* $\mathbf{x}, \mathbf{y}$, $\|\mathbf{AGy}-\mathbf{y}\| \le \|\mathbf{Ax}-\mathbf{y}\|$.

PROOF. (i) $\Rightarrow$ (ii). Let $\mathbf{x}-\mathbf{Gy} = \mathbf{w}$. Then we must show

$$\|\mathbf{AGy}-\mathbf{y}\| \le \|\mathbf{AGy}-\mathbf{y}+\mathbf{Aw}\|. \tag{4}$$

We have

$$\|\mathbf{AGy}-\mathbf{y}+\mathbf{Aw}\|^2 = \|(\mathbf{AG}-\mathbf{I})\mathbf{y}\|^2 + \|\mathbf{Aw}\|^2 + 2\mathbf{w}'\mathbf{A}'(\mathbf{AG}-\mathbf{I})\mathbf{y}. \tag{5}$$

But

$$\mathbf{w}'\mathbf{A}'(\mathbf{AG}-\mathbf{I})\mathbf{y} = \mathbf{w}'(\mathbf{A}'\mathbf{G}'\mathbf{A}'-\mathbf{A}')\mathbf{y} = 0,$$

since $(\mathbf{AG})' = \mathbf{AG}$. Inserting this in (5) we get (4).

(ii) $\Rightarrow$ (i). For any vector $\mathbf{x}$, set $\mathbf{y} = \mathbf{Ax}$ in (ii). Then we see that

$$\|\mathbf{AGAx}-\mathbf{Ax}\| \le \|\mathbf{Ax}-\mathbf{Ax}\| = 0$$

and hence $\mathbf{AGAx} = \mathbf{Ax}$. Since $\mathbf{x}$ is arbitrary, $\mathbf{AGA} = \mathbf{A}$, and therefore $\mathbf{G}$ is a g-inverse of $\mathbf{A}$. The remaining part of the proof parallels that of (ii) $\Rightarrow$ (i) in **1.5** and is left as an exercise. $\square$

Suppose we have the equation $\mathbf{Ax} = \mathbf{y}$ that is not necessarily consistent and suppose we wish to find a solution $\mathbf{x}$ such that $\|\mathbf{Ax}-\mathbf{y}\|$ is minimized. Then according to **1.6** this is achieved by taking $\mathbf{x} = \mathbf{Gy}$ for any least squares g-inverse $\mathbf{G}$ of $\mathbf{A}$.

If $\mathbf{G}$ is a reflexive g-inverse of $\mathbf{A}$ that is both minimum norm and least squares then it is called a *Moore–Penrose inverse* of $\mathbf{A}$. In other words, $\mathbf{G}$ is a Moore–Penrose inverse of $\mathbf{A}$ if it satisfies

$$\mathbf{AGA} = \mathbf{A}, \quad \mathbf{GAG} = \mathbf{G}, \quad (\mathbf{AG})' = \mathbf{AG}, \quad (\mathbf{GA})' = \mathbf{GA}. \tag{6}$$

We will show that such a $\mathbf{G}$ exists and is, in fact, unique. We first show uniqueness. Suppose $\mathbf{G_1}, \mathbf{G_2}$ both satisfy (6). Then we must show $\mathbf{G_1} = \mathbf{G_2}$. Each of the following steps follows by applying (6). The terms that are underlined are to be reinterpreted to get the next step each time.

$$\mathbf{G_1} = \mathbf{G_1}\underline{\mathbf{AG_1}}$$

$$
\begin{aligned}
&= \mathbf{G}_1\mathbf{G}_1'\underline{\mathbf{A}'}\\
&= \mathbf{G}_1\mathbf{G}_1'\mathbf{A}'\underline{\mathbf{G}_2'\mathbf{A}'}\\
&= \mathbf{G}_1\underline{\mathbf{G}_1'\mathbf{A}'}\mathbf{A}\mathbf{G}_2\\
&= \mathbf{G}_1\underline{\mathbf{A}\mathbf{G}_1\mathbf{A}}\mathbf{G}_2\\
&= \mathbf{G}_1\underline{\mathbf{A}}\mathbf{G}_2\\
&= \mathbf{G}_1\mathbf{A}\underline{\mathbf{G}_2\mathbf{A}}\mathbf{G}_2\\
&= \underline{\mathbf{G}_1\mathbf{A}}\mathbf{A}'\mathbf{G}_2'\mathbf{G}_2\\
&= \underline{\mathbf{A}'\mathbf{G}_1'\mathbf{A}'}\mathbf{G}_2'\mathbf{G}_2\\
&= \underline{\mathbf{A}'\mathbf{G}_2'}\mathbf{G}_2\\
&= \mathbf{G}_2\mathbf{A}\mathbf{G}_2\\
&= \mathbf{G}_2.
\end{aligned}
$$

We will denote the Moore–Penrose inverse of $\mathbf{A}$ by $\mathbf{A}^+$. We now show the existence. Let $\mathbf{A} = \mathbf{BC}$ be a rank factorization. Then it can be easily verified that

$$\mathbf{B}^+ = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}', \quad \mathbf{C}^+ = \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1},$$

and then

$$\mathbf{A}^+ = \mathbf{C}^+\mathbf{B}^+.$$

Problems

1. Find two different g-inverses of

$$\begin{bmatrix} 1 & 0 & -1 & 2 \\ 2 & 0 & -2 & 4 \\ -1 & 1 & 1 & 3 \\ -2 & 2 & 2 & 6 \end{bmatrix}.$$

2. Find the minimum norm solution of the system of equations

$$
\begin{aligned}
2x + y - z &= 1,\\
x - 2y + z &= -2,\\
x + 3y - 2z &= 3.
\end{aligned}
$$

3. Find the Moore–Penrose inverse of $\begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix}$.

2.2 Linear Model

Let $\mathbf{y}$ be a column vector with components $y_1, \ldots, y_n$. We call $\mathbf{y}$ a *random vector* if each y_i is a random variable. The expectation of $\mathbf{y}$, denoted by $E(\mathbf{y})$, is the column vector whose ith component is $E(y_i)$. Clearly,

$$E(\mathbf{Bx} + \mathbf{Cy}) = \mathbf{B}E(\mathbf{x}) + \mathbf{C}E(\mathbf{y}),$$

where $\mathbf{x}$, $\mathbf{y}$ are random vectors and $\mathbf{B}$, $\mathbf{C}$ are constant nonrandom matrices.

If $\mathbf{x}$, $\mathbf{y}$ are random vectors of order m, n, respectively, then the *covariance matrix* between $\mathbf{x}$, $\mathbf{y}$, denoted by cov($\mathbf{x}$, $\mathbf{y}$), is an $m \times n$ matrix whose (i, j)-entry is $\mathrm{cov}(x_i, y_j)$.

The *dispersion matrix,* or the variance-covariance matrix of $\mathbf{y}$, denoted by $D(\mathbf{y})$, is defined to be cov($\mathbf{y}$, $\mathbf{y}$). The dispersion matrix is obviously symmetric.

If $\mathbf{b}$, $\mathbf{c}$ are constant vectors, then

$$\begin{aligned}\mathrm{cov}(\mathbf{b}'\mathbf{x}, \mathbf{c}'\mathbf{y}) &= \mathrm{cov}(b_1x_1 + \cdots + b_mx_m, c_1y_1, + \cdots + c_ny_n)\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} b_ic_j\mathrm{cov}(x_i, y_j)\\ &= \mathbf{b}'\mathrm{cov}(\mathbf{x}, \mathbf{y})\mathbf{c}.\end{aligned}$$

It follows that if $\mathbf{B}$, $\mathbf{C}$ are constant matrices, then

$$\mathrm{cov}(\mathbf{Bx}, \mathbf{Cy}) = \mathbf{B}\mathrm{cov}(\mathbf{x}, \mathbf{y})\mathbf{C}'.$$

Setting $\mathbf{x} = \mathbf{y}$ and $\mathbf{b} = \mathbf{c}$ gives

$$\mathrm{var}(\mathbf{b}'\mathbf{x}) = \mathbf{b}'D(\mathbf{x})\mathbf{b}.$$

Since variance is nonnegative, we conclude that $D(\mathbf{x})$ is positive semidefinite. Note that $D(\mathbf{x})$ is positive definite unless there exists a linear combination $\mathbf{b}'\mathbf{x}$ that is constant with probability one.

We now introduce the concept of a linear model. Suppose we conduct an experiment that gives rise to the random variables $y_1, \ldots, y_n$. We make the assumption that the distribution of the random variables is controlled by some (usually a small number of) unknown parameters. In a linear model, the basic assumption is that $E(y_i)$ is a linear function of the parameters $\beta_1, \ldots, \beta_p$ with known coefficients. In matrix notation this can be expressed as

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{y}$ is the $n \times 1$ vector with components $y_1, \ldots, y_n$; $\mathbf{X}$ is a known nonrandom matrix of order $n \times p$; and $\boldsymbol{\beta}$ is the $p \times 1$ vector of parameters $\beta_1, \ldots, \beta_p$. We also assume that $y_1, \ldots, y_n$ are uncorrelated and that $\mathrm{var}(y_i) = \sigma^2$ for all i; this property is called *homoscedasticity.* Thus

$$D(\mathbf{y}) = \sigma^2\mathbf{I}.$$

Another way to write the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon,$$

where the vector ϵ satisfies $E(\epsilon) = \mathbf{0}$, $D(\epsilon) = \sigma^2\mathbf{I}$.

We do not make any further assumptions about the distribution of $\mathbf{y}$ at present. Our first objective is to find estimates of $\beta_1 \ldots, \beta_p$ and their linear combinations. We also seek an estimate of σ^2.

Problems

Answer the following questions with reference to the linear model $E(y_1) = \beta_1 + \beta_2$, $E(y_2) = 2\beta_1 - \beta_2$, $E(y_3) = \beta_1 - \beta_2$, where y_1, y_2, y_3 are uncorrelated with a common variance σ^2:

1. Find two different linear functions of y_1, y_2, y_3 that are unbiased for β_1. Determine their variances and the covariance between the two.
2. Find two linear functions that are both unbiased for β_2 and are uncorrelated.
3. Write the model in terms of the new parameters $\theta_1 = \beta_1 + 2\beta_2, \theta_2 = \beta_1 - 2\beta_2$.

2.3 Estimability

Consider the linear model

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \quad D(\mathbf{y}) = \sigma^2\mathbf{I}, \tag{7}$$

where $\mathbf{y}$ is $n \times 1$, $\mathbf{X}$ is $n \times p$, and $\boldsymbol{\beta}$ is $p \times 1$.

The linear parametric function $\boldsymbol{\ell}'\boldsymbol{\beta}$ is said to be *estimable* if there exists a linear function $\mathbf{c}'\mathbf{y}$ of the observations such that $E(\mathbf{c}'\mathbf{y}) = \boldsymbol{\ell}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in R^p$.

The condition $E(\mathbf{c}'\mathbf{y}) = \boldsymbol{\ell}'\boldsymbol{\beta}$ is equivalent to $\mathbf{c}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\ell}'\boldsymbol{\beta}$, and since this must hold for all $\boldsymbol{\beta}$ in R^p, we must have $\mathbf{c}'\mathbf{X} = \boldsymbol{\ell}'$. Thus $\boldsymbol{\ell}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{\ell}' \in \mathcal{R}(\mathbf{X})$.

The following facts concerning generalized inverse are frequently used in this as well as the next chapter:

(i) For any matrix $\mathbf{X}$, $\mathcal{R}(\mathbf{X}) = \mathcal{R}(\mathbf{X}'\mathbf{X})$. This is seen as follows. Clearly, $\mathcal{R}(\mathbf{X}'\mathbf{X}) \subset \mathcal{R}(\mathbf{X})$. However, $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}$ have the same rank (see Exercise 19, Chapter 1), and therefore their row spaces have the same dimension. This implies that the spaces must be equal. As a consequence we can write $\mathbf{X} = \mathbf{M}\mathbf{X}'\mathbf{X}$ for some matrix $\mathbf{M}$.

(ii) The matrix $\mathbf{A}\mathbf{C}^-\mathbf{B}$ is invariant under the choice of the g-inverse $\mathbf{C}^-$ of $\mathbf{C}$ if $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{C})$ and $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{C})$. This is seen as follows. We can write $\mathbf{B} = \mathbf{C}\mathbf{U}$ and $\mathbf{A} = \mathbf{V}\mathbf{C}$ for some matrices $\mathbf{U}, \mathbf{V}$. Then

$$\mathbf{A}\mathbf{C}^-\mathbf{B} = \mathbf{V}\mathbf{C}\mathbf{C}^-\mathbf{C}\mathbf{U} = \mathbf{V}\mathbf{C}\mathbf{U},$$

which does not depend on the choice of the g-inverse. (Note that the matrices $\mathbf{U}, \mathbf{V}$ are not necessarily unique. However, if $\mathbf{B} = \mathbf{C}\mathbf{U}_1$, $\mathbf{A} = \mathbf{V}_1\mathbf{C}$ is another representation, then

$$\mathbf{V}_1\mathbf{C}\mathbf{U}_1 = \mathbf{V}_1\mathbf{C}\mathbf{C}^-\mathbf{C}\mathbf{U}_1 = \mathbf{A}\mathbf{C}^-\mathbf{B} = \mathbf{V}\mathbf{C}\mathbf{C}^-\mathbf{C}\mathbf{U} = \mathbf{V}\mathbf{C}\mathbf{U}.)$$

The statement has a converse, which we will establish in Chapter 6.

(iii) The matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ is invariant under the choice of the g-inverse. This is immediate from (ii), since $\mathcal{R}(\mathbf{X}) = \mathcal{R}(\mathbf{X}'\mathbf{X})$.

(iv) $\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \mathbf{X}$, $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{X}'$. This is easily proved by writing $\mathbf{X} = \mathbf{M}\mathbf{X}'\mathbf{X}$.

3.1. *Let $\boldsymbol{\ell}'\boldsymbol{\beta}$ be an estimable function and let $\mathbf{G}$ be a least squares g-inverse of $\mathbf{X}$. Then $\boldsymbol{\ell}'\mathbf{Gy}$ is an unbiased linear estimate of $\boldsymbol{\ell}'\boldsymbol{\beta}$ with minimum variance among all unbiased linear estimates of $\boldsymbol{\ell}'\boldsymbol{\beta}$. We say that $\boldsymbol{\ell}'\mathbf{Gy}$ is BLUE (best linear unbiased estimate) of $\boldsymbol{\ell}'\boldsymbol{\beta}$. The variance of $\boldsymbol{\ell}'\mathbf{Gy}$ is $\sigma^2\boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\ell}$.*

PROOF. Since $\boldsymbol{\ell}'\boldsymbol{\beta}$ is estimable, $\boldsymbol{\ell}' = \mathbf{u}'\mathbf{X}$ for some $\mathbf{u}$. Then

$$E(\boldsymbol{\ell}'\mathbf{Gy}) = \mathbf{u}'\mathbf{XGX}\boldsymbol{\beta} = \mathbf{u}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\ell}'\boldsymbol{\beta},$$

and hence $\boldsymbol{\ell}'\mathbf{Gy}$ is unbiased for $\boldsymbol{\ell}'\boldsymbol{\beta}$. Any other linear unbiased estimate is of the form $(\boldsymbol{\ell}'\mathbf{G} + \mathbf{w}')\mathbf{y}$, where $\mathbf{w}'\mathbf{X} = 0$. Now

$$\begin{aligned} \operatorname{var}\{(\boldsymbol{\ell}'\mathbf{G} + \mathbf{w}')\mathbf{y}\} &= \sigma^2(\boldsymbol{\ell}'\mathbf{G} + \mathbf{w}')(\mathbf{G}'\boldsymbol{\ell} + \mathbf{w}) \\ &= \sigma^2(\mathbf{u}'\mathbf{XG} + \mathbf{w}')(\mathbf{G}'\mathbf{X}'\mathbf{u} + \mathbf{w}). \end{aligned}$$

Since $\mathbf{G}$ is a least squares g-inverse of $\mathbf{X}$,

$$\mathbf{u}'\mathbf{XGw} = \mathbf{u}'\mathbf{G}'\mathbf{X}'\mathbf{w} = 0,$$

and therefore

$$\begin{aligned} \operatorname{var}\{(\boldsymbol{\ell}'\mathbf{G} + \mathbf{w}')\mathbf{y}\} &= \sigma^2(\mathbf{u}'(\mathbf{XG})(\mathbf{XG})'\mathbf{u} + \mathbf{w}'\mathbf{w}) \\ &\geq \sigma^2\mathbf{u}'(\mathbf{XG})(\mathbf{XG})'\mathbf{u} \\ &= \operatorname{var}(\boldsymbol{\ell}'\mathbf{Gy}). \end{aligned}$$

Therefore, $\boldsymbol{\ell}'\mathbf{Gy}$ is BLUE of $\boldsymbol{\ell}'\boldsymbol{\beta}$. The variance of $\boldsymbol{\ell}'\mathbf{Gy}$ is $\sigma^2\boldsymbol{\ell}'\mathbf{GG}'\boldsymbol{\ell}$. It is easily seen that for any choice of g-inverse, $(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ is a least squares g-inverse of $\mathbf{X}$. In particular, using the Moore–Penrose inverse,

$$\begin{aligned} \boldsymbol{\ell}'\mathbf{GG}'\boldsymbol{\ell} &= \boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^+\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^+\boldsymbol{\ell} \\ &= \boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^+\boldsymbol{\ell} \\ &= \boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\ell}, \end{aligned}$$

since $\boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\ell} = \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{u}$ is invariant with respect to the choice of g-inverse. □

Example
Consider the model

$$E(y_{ij}) = \alpha_i + \beta_j, \qquad i = 1, 2; \quad j = 1, 2.$$

We can express the model in standard form as

$$E\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix},$$

so that $\mathbf{X}$ is the 4×4 matrix on the right-hand side. Let S be the set of all vectors (l_1, l_2, m_1, m_2) such that $l_1 + l_2 = m_1 + m_2$. Note that if $\mathbf{x} \in \mathcal{R}(\mathbf{X})$, then $\mathbf{x} \in S$.

Thus $\mathcal{R}(\mathbf{X}) \subset S$. Clearly, $\dim(S) = 3$, and the rank of $\mathbf{X}$ is 3 as well. Therefore, $\mathcal{R}(\mathbf{X}) = S$, and we conclude that $l_1\alpha_1 + l_2\alpha_2 + m_1\beta_1 + m_2\beta_2$ is estimable if and only if $l_1 + l_2 = m_1 + m_2$.

We compute

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix},$$

and

$$(\mathbf{X}'\mathbf{X})^- = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & -2 & -2 \\ 0 & -2 & 3 & 1 \\ 0 & -2 & 1 & 3 \end{bmatrix}$$

is one possible g-inverse. Thus

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \frac{1}{4}\begin{bmatrix} 3 & 1 & 1 & -1 \\ 1 & 3 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{bmatrix}.$$

Now we can compute the BLUE of any estimable function $\mathbf{u}'\mathbf{X}\boldsymbol{\beta}$ as $\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$. For example, if $\mathbf{u}' = (1, 0, 0, 0)'$, then we get the BLUE of $\alpha_1 + \beta_1$ as

$$\frac{1}{4}(3y_{11} + y_{12} + y_{21} - y_{22}).$$

The model (7) is said to be a *full-rank model* (or a *regression model*) if $\mathbf{X}$ has full column rank, i.e., $R(\mathbf{X}) = p$. For such models the following results can easily be verified.

(i) $\mathcal{R}(\mathbf{X}) = R^p$, and therefore every function $\boldsymbol{\ell}'\boldsymbol{\beta}$ is estimable.
(ii) $\mathbf{X}'\mathbf{X}$ is nonsingular.
(iii) Let $\widehat{\beta}_i$ be the BLUE of β_i and let $\widehat{\boldsymbol{\beta}}$ be the column vector with components $\widehat{\beta}_1, \dots, \widehat{\beta}_p$. Then $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The dispersion matrix of $\widehat{\boldsymbol{\beta}}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
(iv) The BLUE of $\boldsymbol{\ell}'\boldsymbol{\beta}$ is $\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}$ with variance $\sigma^2\boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\ell}$.

Parts (iii) and (iv) constitute the *Gauss–Markov theorem.*

Problems

1. Consider the model $E(y_1) = 2\beta_1 - \beta_2 - \beta_3$, $E(y_2) = \beta_2 - \beta_4$, $E(y_3) = \beta_2 + \beta_3 - 2\beta_4$ with the usual assumptions. Describe the estimable functions.

2. Consider the model $E(y_1) = \beta_1 + \beta_2$, $E(y_2) = \beta_1 - \beta_2$, $E(y_3) = \beta_1 + 2\beta_2$ with the usual assumptions. Obtain the BLUE of $2\beta_1 + \beta_2$ and find its variance.

3. Consider the model $E(y_1) = 2\beta_1 + \beta_2$, $E(y_2) = \beta_1 - \beta_2$, $E(y_3) = \beta_1 + \alpha\beta_2$ with the usual assumptions. Determine α such that the BLUEs of β_1, β_2 are uncorrelated.

2.4 Weighing Designs

The next result is the Hadamard inequality for positive semidefinite matrices.

4.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *positive semidefinite matrix. Then*

$$|\mathbf{A}| \le a_{11} \cdots a_{nn}.$$

Furthermore, if $\mathbf{A}$ *is positive definite, then equality holds in the above inequality if and only if* $\mathbf{A}$ *is a diagonal matrix.*

PROOF. If $\mathbf{A}$ is singular, then $|\mathbf{A}| = 0$, whereas $a_{ii} \ge 0$ for all i, and the result is trivial. So suppose $\mathbf{A}$ is nonsingular. Then each $a_{ii} > 0$. Let $\mathbf{D} = \operatorname{diag}(\sqrt{a_{11}}, \ldots, \sqrt{a_{nn}})$ and let $\mathbf{B} = \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$. Then $\mathbf{B}$ is positive semidefinite and $b_{ii} = 1$ for each i. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{B}$. By the arithmetic mean-geometric mean inequality,

$$\frac{1}{n}\sum_{i=1}^{n} \lambda_i \ge \prod_{i=1}^{n} \lambda_i^{\frac{1}{n}}.$$

Since

$$\sum_{i=1}^{n} \lambda_i = \operatorname{trace}\mathbf{B} = n$$

and $\prod_{i=1}^{n} \lambda_i = |\mathbf{B}|$, we get $|\mathbf{B}| \le 1$. Therefore,

$$|\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}| \le 1,$$

and the inequality follows. If $\mathbf{A}$ is positive definite and if equality holds in the inequality, then it must hold in the arithmetic mean–geometric mean inequality in the proof above. But then $\lambda_1, \ldots, \lambda_n$ are all equal, and it follows by the spectral theorem that $\mathbf{B}$ is a scalar multiple of the identity matrix. Then $\mathbf{A}$ must be diagonal. □

4.2. *Let* $\mathbf{X}$ *be an* $n \times n$ *matrix and suppose* $|x_{ij}| \le 1$ *for all* i, j. *Then*

$$|\mathbf{X}'\mathbf{X}| \le n^n.$$

PROOF. Let $\mathbf{A} = \mathbf{X}'\mathbf{X}$. Then

$$a_{ii} = \sum_{j=1}^{n} x_{ji}^2 \le n$$

and $|\mathbf{A}| = |\mathbf{X}'\mathbf{X}|$. The result follows by **4.1**. □

An application of the inequality in **4.2** is illustrated by the following example. Suppose four objects are to be weighed using an ordinary chemical balance (without bias) with two pans. We are allowed four weighings. In each weighing we may put some of the objects in the right pan and some in the left pan. Any procedure that specifies this allocation is called a *weighing design*. Let $\beta_1, \beta_2, \beta_3, \beta_4$ be the true weights of the objects. Define $x_{ij} = 1$ or -1 depending upon whether we put the jth object in the right pan or in the left pan in the ith weighing. We set $x_{ij} = 0$ if the jth object is not used at all in the ith weighing. Let y_i denote the weight needed to achieve balance in the ith weighing. If the sign of y_i is positive, then the weight is required in the left pan, otherwise in the right pan. Then we have the model $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X} = (x_{ij})$, $\mathbf{y}$ is the 4×1 vector with components y_i, and $\boldsymbol{\beta}$ is the 4×1 vector with components β_i. As usual, we make the assumption that the y_i's are uncorrelated with common variance σ^2. The dispersion matrix of $\widehat{\boldsymbol{\beta}}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, assuming $\mathbf{X}'\mathbf{X}$ to be nonsingular. Thus to get more precision we must make the $\mathbf{X}'\mathbf{X}$ matrix "large." One measure of largeness of a positive semidefinite matrix is the determinant. (This is the D-optimality criterion, which we will encounter again in Chapter 5 in the context of block designs.) The matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{8}$$

satisfies $|\mathbf{X}'\mathbf{X}| = 4^4$, and by **4.2** this is the maximum determinant possible.

A square matrix is called a *Hadamard matrix* if each entry is 1 or -1 and the rows are orthogonal. The matrix (8) is a Hadamard matrix.

Problems

1. Suppose the matrix

$$A = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix}$$

is positive definite, where a, b, c are real numbers, not all zero. Show that

$$a^2 + b^2 + c^2 - 2abc > 0.$$

2. Let

$$A = \begin{bmatrix} 2 & a & 1 \\ -1 & 1 & a \\ 1 & -1 & a \end{bmatrix}.$$

Show that $|\mathbf{A}| \le (a^2 + 2)\sqrt{a^2 + 5}$.

3. Show that there exists a Hadamard matrix of order 2^k for any positive integer $k \geq 1$.

2.5 Residual Sum of Squares

We continue to consider the model (7) of Section 3. The equations $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ are called the *normal equations.* The equations are consistent, since $\mathcal{C}(\mathbf{X}') = \mathcal{C}(\mathbf{X}'\mathbf{X})$. Let $\widehat{\boldsymbol{\beta}}$ be a solution of the normal equations. Then $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ for some choice of the g-inverse. The *residual sum of squares* (RSS) is defined to be

$$(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}).$$

The RSS is invariant under the choice of the g-inverse $(\mathbf{X}'\mathbf{X})^{-}$, although $\widehat{\boldsymbol{\beta}}$ depends on the choice. Thus $\widehat{\boldsymbol{\beta}}$ is not unique and does not admit any statistical interpretation. By "fitting the model" we generally mean calculating the BLUEs of parametric functions of interest and computing RSS.

5.1. *The minimum of* $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ *is attained at* $\widehat{\boldsymbol{\beta}}$.

PROOF. We have

$$\begin{aligned}
&(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&\quad = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}) \\
&\quad = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}),
\end{aligned}$$

since

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{X}'(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}) = \mathbf{0}.$$

It follows that

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \geq (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}).$$

That completes the proof. □

5.2. *Let* $R(\mathbf{X}) = r$. *Then* $E(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = (n - r)\sigma^2$.

PROOF. We have

$$E(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' = D(\mathbf{y}) = \sigma^2\mathbf{I}.$$

Thus

$$\begin{aligned}
E(\mathbf{y}\mathbf{y}') &= E(\mathbf{y})\boldsymbol{\beta}'\mathbf{X}' + \mathbf{X}\boldsymbol{\beta}E(\mathbf{y}') - \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I} \\
&= \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I}. \qquad (9)
\end{aligned}$$

We will use the notation

$$\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$$

throughout this and the next chapter. Observe that $\mathbf{P}$ is a symmetric, idempotent matrix and $\mathbf{PX} = \mathbf{0}$. These properties will be useful. Now,

$$\begin{aligned}
E(\mathbf{y} - \mathbf{X}\widehat{\beta})'(\mathbf{y} - \mathbf{X}\widehat{\beta}) &= E(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y})'(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}) \\
&= E\mathbf{y}'\mathbf{P}\mathbf{y} \\
&= E \text{ trace}(\mathbf{y}'\mathbf{P}\mathbf{y}) \\
&= E \text{ trace}(\mathbf{P}\mathbf{y}\mathbf{y}') \\
&= \text{trace}\mathbf{P}E(\mathbf{y}\mathbf{y}') \\
&= \sigma^2 \text{trace}\mathbf{P},
\end{aligned}$$

by (9) and the fact that $\mathbf{PX} = \mathbf{0}$. Finally,

$$\begin{aligned}
\text{trace}\mathbf{P} &= n - \text{trace}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' \\
&= n - \text{trace}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} \\
&= n - R((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}),
\end{aligned}$$

since $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}$ is idempotent. However,

$$R((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}) = R(\mathbf{X}'\mathbf{X}) = R(\mathbf{X}) = r,$$

and the proof is complete. □

We conclude from **5.2** that $\text{RSS}/(n - r)$ is an unbiased estimator of σ^2. For computations it is more convenient to use the expressions

$$\text{RSS} = \mathbf{y}'\mathbf{y} - \widehat{\beta}'\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\widehat{\beta}.$$

5.3 Example (One-Way Classification)
Consider the model

$$y_{ij} = \alpha_i + \epsilon_{ij}, \qquad i = 1, \ldots, k, \quad j = 1, \ldots, n_i,$$

where ϵ_{ij} are independent with mean 0 and variance σ^2. The model can be written as

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & & 0 \\ \cdots & & \cdots & \\ 0 & & 0 & 1 \\ \vdots & & \vdots & \vdots \\ 0 & & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} + \boldsymbol{\epsilon}.$$

Thus

$$\mathbf{X}'\mathbf{X} = \operatorname{diag}(n_1, \dots, n_k)$$

and

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} y_{1.} \\ \vdots \\ y_{k.} \end{bmatrix},$$

where

$$y_{i.} = \sum_{j=1}^{n_i} y_{ij}, \qquad i = 1, \dots, k.$$

Thus the model is of full rank and the BLUEs of α_i are given by the components of

$$\widehat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \overline{y}_{1.} \\ \vdots \\ \overline{y}_{k.} \end{bmatrix},$$

where $\overline{y}_{i.} = y_{i.}/n_i, \; i = 1, \dots, k$. Now

$$\text{RSS} = \mathbf{y}'\mathbf{y} - \widehat{\alpha}'\mathbf{X}'\mathbf{y} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{y_{i.}^2}{n_i}.$$

Since the rank of $\mathbf{X}$ is k, by **5.2**, $E(\text{RSS}) = (n-k)\sigma^2$, where $n = \sum_{i=1}^{n} n_i$ and $\text{RSS}/(n-k)$ is an unbiased estimator of σ^2.

Problems

1. Consider the model $E(y_1) = \beta_1 + \beta_2$, $E(y_2) = 2\beta_1$, $E(y_3) = \beta_1 - \beta_2$ with the usual assumptions. Find the RSS.
2. Suppose the one-way classification model is written as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad i = 1, \dots, k, \quad j = 1, \dots, n_i,$$

where ϵ_{ij} are independent with mean 0 and variance σ^2. The parameter μ is normally referred to as the "general effect." What are the estimable functions? Is it correct to say that the grand mean $\overline{y}_{..}$ is an unbiased estimator of μ?

2.6 Estimation Subject to Restrictions

Consider the usual model $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \; D(\mathbf{y}) = \sigma^2\mathbf{I}$, where $\mathbf{y}$ is $n \times 1$, $\mathbf{X}$ is $n \times p$. Suppose we have a priori linear restrictions $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$ on the parameters. We assume that $\mathcal{R}(\mathbf{L}) \subset \mathcal{R}(\mathbf{X})$ and that the equation $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$ is consistent.

Let $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ for a fixed g-inverse $(\mathbf{X}'\mathbf{X})^{-}$ and let

$$\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{z}).$$

6.1. *The minimum of* $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ *subject to* $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$ *is attained at* $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$.

PROOF. Since $\mathcal{R}(\mathbf{L}) \subset \mathcal{R}(\mathbf{X})$ and since $R(\mathbf{X}) = R(\mathbf{X}'\mathbf{X})$, then $\mathbf{L} = \mathbf{W}\mathbf{X}'\mathbf{X}$ for some $\mathbf{W}$. Let $\mathbf{T} = \mathbf{W}\mathbf{X}'$. Now,

$$\begin{aligned}\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}' &= \mathbf{W}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\mathbf{W}' \\ &= \mathbf{W}\mathbf{X}'\mathbf{X}\mathbf{W}' \\ &= \mathbf{T}\mathbf{T}'.\end{aligned}$$

Since $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$ is consistent, $\mathbf{L}\mathbf{v} = \mathbf{z}$ for some $\mathbf{v}$. Thus

$$\begin{aligned}\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}\mathbf{z} &= \mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}\mathbf{L}\mathbf{v} \\ &= \mathbf{T}\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-}\mathbf{T}\mathbf{X}\mathbf{v} \\ &= \mathbf{T}\mathbf{X}\mathbf{v} \\ &= \mathbf{L}\mathbf{v} \\ &= \mathbf{z}. \qquad (10)\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}\mathbf{L}\widehat{\boldsymbol{\beta}} &= \mathbf{T}\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-}\mathbf{W}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \\ &= \mathbf{T}\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-}\mathbf{W}\mathbf{X}'\mathbf{y} \\ &= \mathbf{T}\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-}\mathbf{T}\mathbf{y} \\ &= \mathbf{T}\mathbf{y} \qquad (11)\end{aligned}$$

and

$$\begin{aligned}\mathbf{L}\widehat{\boldsymbol{\beta}} &= \mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \\ &= \mathbf{W}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \\ &= \mathbf{W}\mathbf{X}'\mathbf{y} \\ &= \mathbf{T}\mathbf{y}. \qquad (12)\end{aligned}$$

Using (10), (11), (12) we see that $\mathbf{L}\tilde{\boldsymbol{\beta}} = \mathbf{z}$, and therefore $\tilde{\boldsymbol{\beta}}$ satisfies the restriction $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$.

Now, for any $\boldsymbol{\beta}$ satisfying $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$,

$$\begin{aligned}&(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &\quad= (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}))'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\ &\quad= (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}), \qquad (13)\end{aligned}$$

since we can show $(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = 0$ as follows: We have

$$\begin{aligned}\mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{z}) \\ &= \mathbf{X}'\mathbf{y} - \mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{z}),\end{aligned}$$

since $\mathbf{L}' = \mathbf{X}'\mathbf{X}\mathbf{W}'$. Hence

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{z}),$$

and since $\mathbf{L}\tilde{\boldsymbol{\beta}} = \mathbf{L}\boldsymbol{\beta} = \mathbf{z}$, it follows that

$$(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{z}) = 0.$$

From (13) it is clear that

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \geq (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$$

if $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$, and the proof is complete. □

6.2. $R(\mathbf{L}) = R(\mathbf{T}) = R(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')$.

PROOF. Since $\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}' = \mathbf{T}\mathbf{T}'$, then $R(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}') = R(\mathbf{T}\mathbf{T}') = R(\mathbf{T})$. Clearly, $R(\mathbf{L}) = R(\mathbf{T}\mathbf{X}) \leq R(\mathbf{T})$. Since $R(\mathbf{X}) = R(\mathbf{X}'\mathbf{X})$, then $\mathbf{X} = \mathbf{M}\mathbf{X}'\mathbf{X}$ for some $\mathbf{M}$. Thus $\mathbf{T} = \mathbf{W}\mathbf{X}' = \mathbf{W}\mathbf{X}'\mathbf{X}\mathbf{M}' = \mathbf{L}\mathbf{M}'$. Therefore, $R(\mathbf{T}) \leq R(\mathbf{L})$, and hence $R(\mathbf{T}) = R(\mathbf{L})$. □

We note some simplifications that occur if additional assumptions are made. Thus suppose that $R(\mathbf{X}) = p$, so that we have a full-rank model. We also assume that $\mathbf{L}$ is $m \times p$ of rank m. Then by **6.2**,

$$R(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}') = R(\mathbf{L}) = m,$$

and hence $\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'$ is nonsingular. It reduces to a scalar if $m = 1$.

6.3 Example

Consider the model $E(y_i) = \theta_i$, $i = 1, 2, 3, 4$, where y_i are uncorrelated with variance σ^2. Suppose we have the restriction $\theta_1+\theta_2+\theta_3+\theta_4 = 0$ on the parameters. We find the RSS. The model in standard form has $\mathbf{X} = \mathbf{I_4}$. The restriction on the parameters can be written as $\mathbf{L}\boldsymbol{\theta} = 0$, where $\mathbf{L} = (1, 1, 1, 1)$. Thus

$$\widehat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$$

and

$$\tilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{L}'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-}\mathbf{L}\widehat{\boldsymbol{\theta}} = \mathbf{y} - \begin{bmatrix} \overline{\mathbf{y}} \\ \overline{\mathbf{y}} \\ \overline{\mathbf{y}} \\ \overline{\mathbf{y}} \end{bmatrix}.$$

Thus

$$\text{RSS} = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\theta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\theta}}) = 4\overline{\mathbf{y}}^{\mathbf{2}}.$$

6.4 Example

Consider an alternative formulation of the one-way classification model considered earlier in Section 5:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad i = 1, \ldots, k; \quad j = 1, \ldots, n_i;$$

where ϵ_{ij} are independent with mean 0 and variance σ^2. This model arises when we want to compare k treatments. We have n_i observations on the ith treatment. The parameter μ is interpreted as the "general effect," and α_i is the "effect due to the ith treatment." We wish to find the RSS. Instead of writing the model in standard form we follow a different approach. The RSS is the minimum value of

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i)^2. \tag{14}$$

We use the fact that if $u_1,\dots,u_m$ are real numbers, then

$$\sum_{i=1}^{m}(u_i-\theta)^2$$

is minimized when $\theta=\overline{u}$, the mean of $u_1,\dots,u_n$. This is easily proved using calculus. Thus (14) is minimized when $\mu+\alpha_i=\overline{y}_{i.}$, $i=1,\dots,k$; and therefore,

$$\text{RSS}=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{i.})^2.$$

Now suppose we wish to find the RSS subject to the constraints $\alpha_i-\alpha_j=0$ for all i,j. Since $\alpha_i-\alpha_j$ is estimable, we may proceed to apply **6.1**. Thus we must calculate $\tilde{\alpha}$ using the formula immediately preceding **6.1**. However, again there is a more elementary way. Let α denote the common value of $\alpha_1,\dots,\alpha_k$. Then we must minimize

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha)^2,$$

and this is achieved by setting

$$\mu+\alpha=\overline{y}_{..}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}y_{ij},$$

where $n=\sum_{i=1}^{k}n_i$. Thus the RSS now is

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{..})^2.$$

The computation of RSS subject to linear restrictions will be useful in deriving a test of the hypothesis that the restrictions are indeed valid. This will be achieved in the next chapter.

Problems

1. Consider the model $E(y_1)=\beta_1+2\beta_2$, $E(y_2)=2\beta_1$, $E(y_3)=\beta_1+\beta_2$ with the usual assumptions. Find the RSS subject to the restriction $\beta_1=\beta_2$.

2. Consider the one-way classification model (with $k \geq 2$)

$$y_{ij} = \alpha_i + \epsilon_{ij}, \qquad i = 1, \dots, k, \quad j = 1, \dots, n_i,$$

where ϵ_{ij} are independent with mean 0 and variance σ^2. Find the RSS subject to the restriction $\alpha_1 = \alpha_2$.

2.7 Exercises

1. Let $\mathbf{A}$ be a matrix and let $\mathbf{G}$ be a g-inverse of $\mathbf{A}$. Show that the class of all g-inverses of $\mathbf{A}$ is given by $\mathbf{G} + (\mathbf{I} - \mathbf{GA})\mathbf{U} + \mathbf{V}(\mathbf{I} - \mathbf{AG})$, where $\mathbf{U}$, $\mathbf{V}$ are arbitrary.

2. Find a g-inverse of the following matrix such that it does not contain any zero entry:

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

3. Show that the class of g-inverses of $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ is given by

$$\begin{bmatrix} 1+a+c & a+d \\ b+c & b+d \end{bmatrix},$$

where a, b, c, d are arbitrary.

4. Let $\mathbf{A}$ be an $m \times n$ matrix of rank r and let k be an integer, $r \leq k \leq \min(m, n)$. Show that $\mathbf{A}$ has a g-inverse of rank k. Conclude that a square matrix has a nonsingular g-inverse.

5. Let $\mathbf{x}$ be an $n \times 1$ vector. Find the g-inverse of $\mathbf{x}$ that is closest to the origin.

6. Let $\mathbf{X}$ be an $n \times m$ matrix and let $\mathbf{y} \in R^n$. Show that the orthogonal projection of $\mathbf{y}$ onto $\mathcal{C}(\mathbf{X})$ is given by $\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$ for any choice of the g-inverse.

7. For any matrix $\mathbf{X}$, show that $\mathbf{X}^+ = (\mathbf{X}'\mathbf{X})^+\mathbf{X}'$ and $\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{X}\mathbf{X}^+$.

8. Let $\mathbf{A}$ be an $m \times n$ matrix and let $\mathbf{P}$, $\mathbf{Q}$ be matrices of order $r \times m$. Then prove that $\mathbf{PA} = \mathbf{QA}$ if and only if $\mathbf{PAA}' = \mathbf{QAA}'$.

9. Let $\mathbf{A}$, $\mathbf{G}$ be matrices of order $m \times n$, $n \times m$, respectively. Then show that $\mathbf{G}$ is a minimum norm g-inverse of $\mathbf{A}$ if and only if $\mathbf{GAA}' = \mathbf{A}'$.

10. Is it true that any positive semidefinite matrix is the dispersion matrix of a random vector?

11. Let $x_1, \dots, x_n$ be real numbers with mean $\overline{x}$. Consider the linear model $Y_i = \alpha + \beta(x_i - \overline{x}) + \epsilon_i$, $i = 1, 2, \dots, n$, with the usual assumptions. Show that the BLUEs of α and β are uncorrelated.

12. Consider the linear model (7) and let $\mathbf{x_i}$ be the ith column of $\mathbf{X}$, $i = 1, \dots, p$. Show that the function $\ell_1\beta_1 + \ell_2\beta_2$ is estimable if and only if $\mathbf{x_1}, \mathbf{x_2}$ do not belong to the linear span of $\ell_2\mathbf{x_1} - \ell_1\mathbf{x_2}, \mathbf{x_3}, \dots, \mathbf{x_p}$.

TABLE 2.1.

distance d	time t
9	1
15	2
19	3
20	4
45	10
55	12
78	18

13. For any vector $\boldsymbol{\ell}$ show that the following conditions are equivalent: (i) $\boldsymbol{\ell}'\boldsymbol{\beta}$ is estimable. (ii) $\boldsymbol{\ell}' = \boldsymbol{\ell}'\mathbf{X}^-\mathbf{X}$ for some g-inverse $\mathbf{X}^-$. (iii) $\boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \boldsymbol{\ell}'$ for some g-inverse $(\mathbf{X}'\mathbf{X})^-$.

14. Prove that the BLUE of an estimable function is unique. (It is to be shown that if $\boldsymbol{\ell}'\boldsymbol{\beta}$ is an estimable function and if $\mathbf{c}'\mathbf{y}$, $\mathbf{d}'\mathbf{y}$ are both BLUE of the function, then $\mathbf{c} = \mathbf{d}$.)

15. Consider the data in Table 2.1, which gives the distance d (in meters) traveled by an object in time t (in seconds). Fit the model $d_i = d_0 + vt_i + e_i, i = 1, 2, \ldots, 7$, where e_i denote uncorrelated errors with zero mean and variance σ^2. Find $\widehat{d_0}, \widehat{v}, \widehat{\sigma^2}$.

16. Suppose $\mathbf{x_i}, \mathbf{y_i}, \mathbf{z_i}$, $i = 1, \ldots, n$, are $3n$ independent observations with common variance σ^2 and expectations given by $E(\mathbf{x_i}) = \theta_1$, $E(\mathbf{y_i}) = \theta_2$, $E(\mathbf{z_i}) = \theta_1 - \theta_2$, $i = 1, \ldots, n$. Find BLUEs of θ_1, θ_2 and compute the RSS.

17. In Example **6.3** suppose a further restriction $\theta_1 = \theta_2$ is imposed. Find the RSS.

18. In the standard linear model set up suppose the error space (the space of linear functions of $\mathbf{y}$ with expectation zero) is one-dimensional, and let z, a linear function of the observations, span the error space. Let $\mathbf{u}'\mathbf{y}$ be unbiased for the function $\mathbf{p}'\boldsymbol{\beta}$. Show that the BLUE of $\mathbf{p}'\boldsymbol{\beta}$ is

$$\mathbf{u}'\mathbf{y} - \frac{\operatorname{cov}(\mathbf{u}'\mathbf{y}, z)}{\operatorname{var}(z)} z.$$

19. Let $\mathbf{A}$ be an $m \times n$ matrix of rank r and suppose $\mathbf{A}$ is partitioned as

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

where $\mathbf{A_{11}}$ is $r \times r$ nonsingular. Show that

$$\mathbf{A} \begin{bmatrix} \operatorname{adj}\mathbf{A_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{A} = |\mathbf{A_{11}}|\mathbf{A}.$$

20. Let $\mathbf{A}$ be an $m \times n$ matrix of rank r with only integer entries. If there exists an integer linear combination of the $r \times r$ minors of $\mathbf{A}$ that equals 1, then show that $\mathbf{A}$ admits a g-inverse with only integer entries.

21. Let $\mathbf{A}$ be a positive semidefinite matrix that is partitioned as

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

where $\mathbf{A}_{11}$ is a square matrix. Then show that

$$|\mathbf{A}| \le |\mathbf{A}_{11}||\mathbf{A}_{22}|.$$

22. Let $X_1, \dots, X_n$ be n random variables and let $\mathbf{A}$ be their correlation matrix, which is defined as an $n \times n$ matrix with its (i, j)-entry equal to the correlation between X_i, X_j. Suppose $|\mathbf{A}| = 1$. What can you conclude about $\operatorname{cov}(X_i, X_j)$ for any i, j?

23. If there exists a Hadamard matrix of order $n, n > 2$, then show that n is divisible by 4.

24. Let $X_1, \dots, X_n$ be random variables with equal mean μ and suppose $\operatorname{var}(X_i) = \lambda^i \sigma^2, i = 1, 2, \dots, n$, where $\lambda > 0$ is known. Find the BLUE of μ.

2.8 Hints and Solutions

Section 3

1. $c_1\beta_1 + c_2\beta_2 + c_3\beta_3 + c_4\beta_4$ is estimable if and only if $c_1 + c_2 + c_3 + c_4 = 0$.

2. The BLUE of $2\beta_1 + \beta_2$ is $\frac{1}{14}(9y_1 + 11y_2 + 8y_3)$ and has variance $\frac{19}{14}\sigma^2$, where σ^2 is the variance of each y_i.

3. $\alpha = -1$.

Section 5

1. $\text{RSS} = \frac{1}{3}(y_1 - y_2 + y_3)^2$.

2. Answer: $c\mu + d_1\alpha_1 + \cdots + d_k\alpha_k$ is estimable if and only if $c = d_1 + \cdots + d_k$. Since μ is not estimable, it is incorrect to say that $\overline{y}_{..}$ (or any linear function of y_{ij}) is an unbiased estimator of μ.

Section 6

1. RSS subject to $\beta_1 = \beta_2$ is

$$\frac{1}{17}(8y_1^2 + 13y_2^2 + 13y_3^2 - 12y_1y_2 - 12y_1y_3 - 8y_2y_3).$$

2.

$$\sum_{j=1}^{n_1}(y_{1j} - \overline{y}_{12.})^2 + \sum_{j=1}^{n_2}(y_{2j} - \overline{y}_{12.})^2 + \sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_{i.})^2$$

where $\overline{y}_{12.} = \frac{n_1\overline{y}_{1.} + n_2\overline{y}_{2.}}{n_1 + n_2}$.

Section 7

1. It is easily verified that $\mathbf{G} + (\mathbf{I} - \mathbf{GA})\mathbf{U} + \mathbf{V}(\mathbf{I} - \mathbf{AG})$ is a g-inverse of $\mathbf{A}$ for any $\mathbf{U}, \mathbf{V}$. Conversely, if $\mathbf{H}$ is a g-inverse of $\mathbf{A}$, then set $\mathbf{U} = \mathbf{HAG}$, $\mathbf{V} = \mathbf{H} - \mathbf{G}$, and verify that $\mathbf{G} + (\mathbf{I} - \mathbf{GA})\mathbf{U} + \mathbf{V}(\mathbf{I} - \mathbf{AG}) = \mathbf{H}$.

4. Using **7.3** of Chapter 1, there exist nonsingular matrices $\mathbf{P}$, $\mathbf{Q}$ such that

$$\mathbf{A} = \mathbf{P}\begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}.$$

Let

$$\mathbf{G} = \mathbf{Q}^{-1}\begin{bmatrix} \mathbf{I_r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I_{k-r}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{P}^{-1}$$

be an $n \times m$ matrix. Then $\mathbf{G}$ is a g-inverse of $\mathbf{A}$ of rank k.

6. Let $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$. Then it can be verified that $(\mathbf{I} - \mathbf{P})\mathbf{y} \in \mathcal{C}(\mathbf{X})$, and $(\mathbf{I} - \mathbf{P})\mathbf{y}$ is orthogonal to $\mathbf{Py}$. Since $\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y} + \mathbf{Py}$, it follows that $(\mathbf{I} - \mathbf{P})\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto $\mathcal{C}(\mathbf{X})$.

12. First suppose $\ell_1\beta_1 + \ell_2\beta_2$ is estimable. Then there exists $\mathbf{u}$ such that $(\ell_1, \ell_2, 0, \dots, 0) = \mathbf{u}'\mathbf{X}$. If $\mathbf{x_1}, \mathbf{x_2}$ belong to the span of

$$\ell_2\mathbf{x_1} - \ell_1\mathbf{x_2}, \mathbf{x_3}, \dots, \mathbf{x_p},$$

then there exists a matrix $\mathbf{C}$ such that

$$[\mathbf{x_1}, \mathbf{x_2}] = [\ell_2\mathbf{x_1} - \ell_1\mathbf{x_2}, \mathbf{x_3}, \dots, \mathbf{x_p}]\mathbf{C}.$$

Premultiplying both sides of this equation by $\mathbf{u}'$, we get $\ell_1 = \ell_2 = 0$, which is a contradiction.

Conversely, suppose $\ell_1\beta_1 + \ell_2\beta_2$ is not estimable. Assume, without loss of generality, that $\ell_2 \neq 0$.

Let $\mathbf{A} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \mathbf{x_3} & \cdots & \mathbf{x_p} \\ \ell_1 & \ell_2 & 0 & \cdots & 0 \end{bmatrix}$. We must show $R(\mathbf{A}) = R(\mathbf{X})$. By elementary column operations, $\mathbf{A}$ has the same rank as

$$\begin{bmatrix} \ell_2\mathbf{x_1} - \ell_1\mathbf{x_2} & \mathbf{x_2} & \mathbf{x_3} & \cdots & \mathbf{x_p} \\ 0 & \ell_2 & 0 & 0 & 0 \end{bmatrix},$$

which in turn equals one plus the rank of $\begin{bmatrix} \ell_2\mathbf{x_1} - \ell_1\mathbf{x_2} & \mathbf{x_3} & \cdots & \mathbf{x_p} \end{bmatrix}$, since $\mathbf{x_2}$ does not belong to the span of $\ell_2\mathbf{x_1} - \ell_1\mathbf{x_2}, \mathbf{x_3}, \dots, \mathbf{x_p}$. It follows that $\mathbf{A}$ has the same rank as

$$\begin{bmatrix} \ell_2\mathbf{x_1} - \ell_1\mathbf{x_2} & \mathbf{x_2} & \mathbf{x_3} & \cdots & \mathbf{x_p} \end{bmatrix}.$$

Finally, the rank of this latter matrix equals that of $\mathbf{X}$, since $\mathbf{x_1}$ can be written as a linear combination of $\ell_2\mathbf{x_1} - \ell_1\mathbf{x_2}$.

14. Suppose $\mathbf{c}'\mathbf{y}$, $\mathbf{d}'\mathbf{y}$ are BLUEs of $\boldsymbol{\ell}'\boldsymbol{\beta}$. Then for any α, $(\alpha\mathbf{c}' + (1-\alpha)\mathbf{d}')\mathbf{y}$ is unbiased for $\boldsymbol{\ell}'\boldsymbol{\beta}$. Now,

$$\operatorname{var}[(\alpha\mathbf{c}' + (1-\alpha)\mathbf{d}')\mathbf{y}] = \sigma^2(\alpha^2\mathbf{c}'\mathbf{c} + 2\alpha(1-\alpha)\mathbf{c}'\mathbf{d} + (1-\alpha)^2\mathbf{d}'\mathbf{d}).$$

This is a quadratic in α and has minima at $\alpha = 0$ and $\alpha = 1$. This is possible only if the quadratic is a constant; in particular, the coefficient of α^2, which is $(\mathbf{c}-\mathbf{d})(\mathbf{c}-\mathbf{d})'$ must be zero. It follows that $\mathbf{c} = \mathbf{d}$.

15. Answer: $\widehat{d_0} = 5.71$, $\widehat{v} = 4.02$, $\widehat{\sigma^2} = 2.22$.

16. Let m_i, v_i, $i = 1, 2, 3$, denote the mean and the variance of the numbers $\mathbf{x_i}$, $\mathbf{y_i}$, $\mathbf{z_i}$, $i = 1, \dots, n$, respectively. The BLUEs of θ_1, θ_2 are given by $\frac{1}{3}(2m_1 + m_2 + m_3)$ and $\frac{1}{3}(m_1 + 2m_2 - m_3)$, respectively. The RSS is given by

$$\frac{n}{3}(m_1 - m_2 - m_3)^2 + n(v_1 + v_2 + v_3).$$

17. Answer: $4\overline{y}^2 + \frac{1}{2}(y_1 - y_2)^2$.

21. Hint: By a suitable transformation reduce the problem to the case where $\mathbf{A_{11}}, \mathbf{A_{22}}$ are both diagonal matrices. Then use the Hadamard inequality.

23. Suppose $\mathbf{A}$ is a Hadamard matrix of order n, $n > 2$. If we multilply any column by -1, then the matrix remains Hadamard. So assume that the first row of $\mathbf{A}$ consists entirely of 1's. Then the second row, being orthogonal to the first row, must contain an equal number of 1's and -1's. Therefore, n is even. A similar analysis using the third row, which must be orthogonal to the first as well as the second row, shows that n must be divisible by 4. (It is strongly believed that conversely, when n is divisible by 4, there exists a Hadamard matrix of order n; however, no proof has been found.)

3

Tests of Linear Hypotheses

3.1 Schur Complements

If $\mathbf{A}$ is positive definite then all principal submatrices of $\mathbf{A}$ are positive definite, and therefore all principal minors of $\mathbf{A}$ are positive. We now prove the converse. The following result will be used, whose proof follows by expanding the determinant along a column several times.

1.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then for any* μ

$$|\mathbf{A} + \mu\mathbf{I}| = \sum_{i=0}^{n} \mu^{n-i} s_i, \tag{1}$$

where s_i *is the sum of all* $i \times i$ *principal minors of* $\mathbf{A}$. *We set* $s_0 = 1$. *Note that* $s_n = |\mathbf{A}|$.

If $\mathbf{A}$ is a symmetric $n \times n$ matrix and if all principal minors of $\mathbf{A}$ are positive, then by **1.1**, $|\mathbf{A} + \mu\mathbf{I}| > 0$ for any $\mu \geq 0$ (when $\mu = 0$ use the fact that $|\mathbf{A}| > 0$). Thus $\mathbf{A}$ cannot have a nonpositive eigenvalue, and therefore $\mathbf{A}$ is positive definite. Combining this observation with **8.3**, **8.4** of Chapter 1 we get the following:

1.2. *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then* $\mathbf{A}$ *is positive definite if and only if all principal minors of* $\mathbf{A}$ *are positive.*

Similarly, a symmetric matrix is positive semidefinite if and only if all its principal minors are nonnegative.

Let $\mathbf{A}$ be a symmetric matrix that is partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D} \end{bmatrix}, \tag{2}$$

where $\mathbf{B}, \mathbf{D}$ are square matrices. If $\mathbf{B}$ is nonsingular, then the *Schur complement* of $\mathbf{B}$ in $\mathbf{A}$ is defined to be the matrix $\mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$. Similarly, if $\mathbf{D}$ is nonsingular, then the Schur complement of $\mathbf{D}$ in $\mathbf{A}$ is $\mathbf{B} - \mathbf{C}\mathbf{D}^{-1}\mathbf{C}'$.

Let $\mathbf{B}$ be nonsingular and let $\mathbf{X} = -\mathbf{C}'\mathbf{B}^{-1}$. The following identity can be verified by simple matrix multiplication:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{X} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{X}' \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C} \end{bmatrix}. \tag{3}$$

Several useful facts can be proved using (3):

1.3. *The following assertions are true:*

(i) *If* $\mathbf{A}$ *is positive definite, then* $\mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$ *is positive definite.*
(ii) *Let* $\mathbf{A}$ *be symmetric. If a principal submatrix of* $\mathbf{A}$ *and its Schur complement in* $\mathbf{A}$ *are positive definite, then* $\mathbf{A}$ *is positive definite.*
(iii) $|\mathbf{A}| = |\mathbf{B}||\mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}|$.

PROOF. (i) Clearly, if $\mathbf{A}$ is positive definite, then $\mathbf{S}\mathbf{A}\mathbf{S}'$ is positive definite for any nonsingular $\mathbf{S}$. If

$$\mathbf{S} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{X} & \mathbf{I} \end{bmatrix},$$

where, as before, $\mathbf{X} = -\mathbf{C}'\mathbf{B}^{-1}$, then $|\mathbf{S}| = 1$, and hence $\mathbf{S}$ is nonsingular. Thus $\mathbf{S}\mathbf{A}\mathbf{S}'$ is positive definite (see (3)), and since $\mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$ is a principal submatrix of $\mathbf{S}\mathbf{A}\mathbf{S}'$, it is positive definite.

(ii) Suppose $\mathbf{A}$ is partitioned as in (2) and suppose $\mathbf{B}$ and $\mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$ are positive definite. Then the right-hand side of (3) is positive definite, and it follows that $\mathbf{A}$ is positive definite, since $\mathbf{S}$ defined in (i) is nonsingular.

(iii) This is immediate by taking the determinant of both sides in (3). □

In (2), suppose $\mathbf{A}$ is $n \times n$ and $\mathbf{B}$ is $(n-1) \times (n-1)$. Then $\mathbf{C}$ is a column vector and $\mathbf{D}$ is 1×1. Let us rewrite (2) as

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{c} \\ \mathbf{c}' & \mathbf{d} \end{bmatrix}. \tag{4}$$

The Schur complement of $\mathbf{B}$ in $\mathbf{A}$ is $\mathbf{d} - \mathbf{c}'\mathbf{B}^{-1}\mathbf{c}$, which is a scalar. By **1.3** (iii),

$$\mathbf{d} - \mathbf{c}'\mathbf{B}^{-1}\mathbf{c} = \frac{|\mathbf{A}|}{|\mathbf{B}|}. \tag{5}$$

A principal submatrix formed by rows $1, \ldots, k$ and columns $1, \ldots, k$ for any k is called a *leading principal submatrix*, and its determinant is a *leading principal* *minor*. We are ready to obtain yet another characterization of positive definite matrices.

1.4. *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then* $\mathbf{A}$ *is positive definite if and only if all leading principal minors of* $\mathbf{A}$ *are positive.*

PROOF. Clearly, if $\mathbf{A}$ is positive definite, then all its leading principal minors are positive. We prove the converse by induction. The result is obvious for $n = 1$. Assume the result for $(n-1) \times (n-1)$ matrices. Let $\mathbf{A}$ be partitioned as in (4). Since any leading principal minor of $\mathbf{B}$ must be positive, by the induction assumption $\mathbf{B}$ is positive definite. Also, $|\mathbf{A}| > 0$, and therefore by (5), the Schur complement of $\mathbf{B}$ in $\mathbf{A}$, $\mathbf{d} - \mathbf{c}'\mathbf{B}^{-1}\mathbf{c}$, is positive. Thus by **1.3** (iii), $\mathbf{A}$ is positive definite, and the proof is complete. □

Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices. We write $\mathbf{A} \geq \mathbf{B}$ to denote the fact that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A} - \mathbf{B}$ are all positive semidefinite matrices. We write $\mathbf{A} \leq \mathbf{B}$ if it is true that $\mathbf{B} \geq \mathbf{A}$.

1.5. *Let* $\mathbf{A}, \mathbf{B}$ *be positive definite matrices such that* $\mathbf{A} \geq \mathbf{B}$. *Then* $\mathbf{A}^{-1} \leq \mathbf{B}^{-1}$.

PROOF. First suppose $\mathbf{B} = \mathbf{I}$. Then $\mathbf{A} \geq \mathbf{I}$ implies that $\mathbf{A} - \mathbf{I}$ is positive semidefinite. Thus each eigenvalue of $\mathbf{A}$ is greater than or equal to 1. Therefore, each eigenvalue of $\mathbf{A}^{-1}$ is less than or equal to 1, and $\mathbf{A}^{-1} \leq \mathbf{I}$. In general, $\mathbf{A} \geq \mathbf{B}$ implies that

$$\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} \geq \mathbf{I},$$

and now the first part can be used to complete the proof. □

Problems

1. Let $\mathbf{A}$ be an $n \times n$ positive definite matrix, $n > 1$, and suppose $a_{ij} \leq 0$ for all $i \neq j$. Let $\mathbf{B}$ be the Schur complement of a_{11} in $\mathbf{A}$. Show that $b_{ij} \leq 0$ for all $i \neq j$.
2. Let $\mathbf{A}$ be an $n \times n$ matrix, not necessarily symmetric, and suppose all principal minors of $\mathbf{A}$ are positive. Show that any real eigenvalue of $\mathbf{A}$ must be positive.
3. Let $\mathbf{A}$ be a symmetric matrix. If every leading principal minor of $\mathbf{A}$ is nonnegative, can we conclude that $\mathbf{A}$ is positive semidefinite?
4. Let $\mathbf{A}$ be an $n \times n$ positive definite matrix partitioned as in (4). Give a proof of the Hadamard inequality (see **4.1** in Chapter 2) using (5).

3.2 Multivariate Normal Distribution

Let $\mathbf{u}$ be a random vector of order n whose components $u_1, \ldots, u_n$ are independent standard normal variables. Let $\mathbf{X}$ be an $r \times n$ matrix, and let $\boldsymbol{\mu}$ be a constant $r \times 1$ vector. The vector $\mathbf{y} = \mathbf{X}\mathbf{u} + \boldsymbol{\mu}$ is said to have (an r-dimensional) *multivariate normal distribution.*

Clearly, $E(\mathbf{y}) = \mathbf{X}E(\mathbf{u}) + \boldsymbol{\mu} = \boldsymbol{\mu}$ and $D(\mathbf{y}) = \mathbf{X}D(\mathbf{u})\mathbf{X}' = \mathbf{X}\mathbf{X}'$. Let $\Sigma = \mathbf{X}\mathbf{X}'$. We now obtain the characteristic function $\phi_{\mathbf{y}}(\mathbf{t})$ of $\mathbf{y}$, defined as

$$\phi_{\mathbf{y}}(\mathbf{t}) = E(\exp(i\mathbf{t}'\mathbf{y})).$$

First, we have

$$\begin{aligned}\phi_{\mathbf{u}}(\mathbf{t}) &= E(\exp(i\mathbf{t}'\mathbf{u})) = \prod_{j=1}^{n} E(\exp(it_j u_j))\\ &= \prod_{j=1}^{n} \exp\Big(-\frac{t_j^2}{2}\Big) = \exp\Big(-\frac{\mathbf{t}'\mathbf{t}}{2}\Big).\end{aligned}$$

Now,

$$\begin{aligned}\phi_{\mathbf{y}}(\mathbf{t}) &= E(\exp(i\mathbf{t}'\mathbf{y})) = E(\exp(i\mathbf{t}'(\mathbf{X}\mathbf{u} + \boldsymbol{\mu}))\\ &= \exp(i\mathbf{t}'\boldsymbol{\mu})E(\exp(i\mathbf{t}'\mathbf{X}\mathbf{u})) = \exp(i\mathbf{t}'\boldsymbol{\mu})\phi_{\mathbf{u}}(\mathbf{t}'\mathbf{X})\\ &= \exp(i\mathbf{t}'\boldsymbol{\mu})\exp(-\frac{1}{2}\mathbf{t}'\mathbf{X}\mathbf{X}'\mathbf{t}) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}).\end{aligned} \tag{6}$$

Thus the distribution of $\mathbf{y}$ depends only on $\boldsymbol{\mu}$ and Σ. Therefore, we will use the notation $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$.

We now show that when Σ is nonsingular, $\mathbf{y}$ has the density function given by

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\Sigma^{-\mathbf{1}}(\mathbf{y} - \boldsymbol{\mu})\right). \tag{7}$$

We will show that if a random vector $\mathbf{y}$ has the density function given by (7), then the characteristic function of $\mathbf{y}$ is (6). Then by the uniqueness of the distribution corresponding to a characteristic function it will follow that if $\mathbf{y}$ is $N(\boldsymbol{\mu}, \Sigma)$, where Σ is nonsingular, then the density function of $\mathbf{y}$ is (7).

We first verify that the function in (7) integrates to 1 and hence is a density function. Make the transformation $\mathbf{z} = \Sigma^{-\mathbf{1/2}}(\mathbf{y} - \boldsymbol{\mu})$. The Jacobian of the transformation is the absolute value of $\left|\left(\frac{\partial z_i}{\partial y_j}\right)\right|$ and is easily seen to be $|\Sigma^{-\mathbf{1/2}}|$. Thus

$$\begin{aligned}&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\mathbf{y})dy_1\cdots dy_n\\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\\ &\qquad\times\exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right)|\Sigma|^{\frac{1}{2}}dz_1\cdots dz_n\\ &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\sum_{j=1}^{n} z_j^2\right)dz_1\cdots dz_n\\ &= \prod_{j=1}^{n}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_j^2\right)dz_j\\ &= 1,\end{aligned}$$

since each term in the product is the total integral of a standard normal density.

The characteristic function of $\mathbf{y}$ is given by

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right) \exp(i\mathbf{t}'\mathbf{y}) dy_1 \cdots dy_n. \tag{8}$$

Make the transformation $\mathbf{z} = \mathbf{y} - \boldsymbol{\mu}$ in (8). The Jacobian is clearly 1. Thus the integral in (8) equals

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{z}'\Sigma^{-1}\mathbf{z}\right) \exp(i\mathbf{t}'(\mathbf{z}+\mu)) dz_1 \cdots dz_n, \tag{9}$$

which is the same as

$$\exp\left(i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right) \times \Delta,$$

where

$$\Delta = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\mathbf{z}-\Sigma\mathbf{t})'\Sigma^{-\mathbf{1}}(\mathbf{z}-\Sigma\mathbf{t})\right) dz_1 \cdots dz_n.$$

Make the transformation $\mathbf{u} = \mathbf{z} - \Sigma\mathbf{t}$ in Δ. Then it reduces to the integral of a standard normal density and therefore equals 1. We conclude that the characteristic function of $\mathbf{y}$ is given by

$$\exp\left(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right).$$

2.1 Exercise
If $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$, then for any matrix $\mathbf{B}$,

$$\mathbf{By} \sim N(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}').$$

(Hint: Find the characteristic function of $\mathbf{By}$.)

Let $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$, and suppose $\mathbf{y}$, $\boldsymbol{\mu}$, and Σ are conformally partitioned as

$$\mathbf{y} = \begin{pmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{\mathbf{11}} & \Sigma_{\mathbf{12}} \\ \Sigma_{\mathbf{21}} & \Sigma_{\mathbf{22}} \end{bmatrix}. \tag{10}$$

The characteristic function of $\mathbf{y_1}$ is obtained by setting $\mathbf{t_2} = \mathbf{0}$, where

$$\mathbf{t} = \begin{pmatrix} \mathbf{t_1} \\ \mathbf{t_2} \end{pmatrix}$$

is the corresponding partitioning of $\mathbf{t}$. Thus

$$\phi_{\mathbf{y_1}}(\mathbf{t_1}) = \exp\left(i\mathbf{t}_1'\boldsymbol{\mu_1} - \frac{1}{2}\mathbf{t}_1'\Sigma_{\mathbf{11}}\mathbf{t_1}\right),$$

and therefore $\mathbf{y_1} \sim N(\boldsymbol{\mu_1}, \Sigma_{\mathbf{11}})$. Similarly, $\mathbf{y_2} \sim N(\boldsymbol{\mu_2}, \Sigma_{\mathbf{22}})$.

2.2. *Let* $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$ *and suppose that* $\mathbf{y}$, $\boldsymbol{\mu}$, *and* Σ *are conformally partitioned as in (10). Then* $\mathbf{y_1}, \mathbf{y_2}$ *are independent, if and only if* $\Sigma_{12} = \mathbf{0}$.

PROOF. If $\mathbf{y_1}, \mathbf{y_2}$ are independent, then $\operatorname{cov}(\mathbf{y_1}, \mathbf{y_2}) = \Sigma_{12} = \mathbf{0}$. We now show that the converse is also true. Thus suppose that $\Sigma_{12} = \mathbf{0}$. Then

$$\mathbf{t}'\Sigma\mathbf{t} = \mathbf{t}_1'\Sigma_{11}\mathbf{t}_1 + \mathbf{t}_2'\Sigma_{22}\mathbf{t}_2.$$

Therefore,

$$\phi_{\mathbf{y}}(\mathbf{t}) = \phi_{\mathbf{y}_1}(\mathbf{t}_1)\phi_{\mathbf{y}_2}(\mathbf{t}_2),$$

and hence $\mathbf{y_1}, \mathbf{y_2}$ are independent. □

2.3. *Let* $\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ *and let* $\mathbf{A}, \mathbf{B}$ *be matrices such that* $\mathbf{AB}' = \mathbf{0}$. *Then* $\mathbf{Ay}, \mathbf{By}$ *are independent.*

PROOF. Observe that by **2.1**,

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{Ay} \\ \mathbf{By} \end{bmatrix}$$

has multivariate normal distribution. So by **2.2**, $\mathbf{Ay}, \mathbf{By}$ are independent if $\operatorname{cov}(\mathbf{Ay}, \mathbf{By}) = \mathbf{AB}' = \mathbf{0}$. □

Now suppose Σ is nonsingular, and we will obtain the conditional distribution of $\mathbf{y_2}$ given $\mathbf{y_1}$. Consider the identity (3) applied to Σ. Then $\mathbf{X} = -\Sigma_{21}\Sigma_{11}^{-1}$. Let

$$\mathbf{S} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{X} & \mathbf{I} \end{bmatrix}.$$

Then

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{X} & \mathbf{I} \end{bmatrix},$$

and we conclude that

$$\Sigma = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{X} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{X}' \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Therefore,

$$\Sigma^{-1} = \mathbf{S}' \begin{bmatrix} \Sigma_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\Sigma}_{22}^{-1} \end{bmatrix} \mathbf{S}, \tag{11}$$

where

$$\tilde{\Sigma}_{22} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

the Schur complement of Σ_{11} in Σ.

Now,

$$
\begin{aligned}
&(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\
&\quad = ((\mathbf{y_1}-\boldsymbol{\mu_1})', (\mathbf{y_2}-\boldsymbol{\mu_2})')\Sigma^{-1}\begin{bmatrix} \mathbf{y_1}-\boldsymbol{\mu_1} \\ \mathbf{y_2}-\boldsymbol{\mu_2} \end{bmatrix} \\
&\quad = (\mathbf{y_1}-\boldsymbol{\mu_1})'\Sigma_{\mathbf{11}}^{-1}(\mathbf{y_1}-\boldsymbol{\mu_1}) \\
&\qquad + ((\mathbf{y_2}-\boldsymbol{\mu_2})' + (\mathbf{y_1}-\boldsymbol{\mu_1})'\mathbf{X}')\tilde{\Sigma}_{22}^{-1}((\mathbf{y_2}-\boldsymbol{\mu_2}) + \mathbf{X}(\mathbf{y_1}-\boldsymbol{\mu_1}))
\end{aligned}
$$

using (11). Also,

$$|\Sigma| = |\Sigma_{\mathbf{11}}||\tilde{\Sigma}_{22}|.$$

Substitute these expressions in the density function of $\mathbf{y}$ given in (7) and then divide by the marginal density of $\mathbf{y_1}$, i.e., an $N(\boldsymbol{\mu_1}, \Sigma_{\mathbf{11}})$ density, to get the conditional density of $\mathbf{y_2}$ given $\mathbf{y_1}$. It turns out that the conditional distribution of $\mathbf{y_2}$ given $\mathbf{y_1}$ is multivariate normal with mean vector

$$\boldsymbol{\mu_2} - \mathbf{X}(\mathbf{y_1}-\boldsymbol{\mu_1}) = \boldsymbol{\mu_2} + \Sigma_{21}\Sigma_{\mathbf{11}}^{-1}(\mathbf{y_1}-\boldsymbol{\mu_1})$$

and dispersion matrix $\tilde{\Sigma}_{22}$.

Problems

1. Let $\mathbf{X} = (X_1, X_2)$ follow a bivariate normal distribution with mean vector $(1, 2)$ and dispersion matrix $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. (i) Find the joint distribution of $X_1 + X_2$ and $X_1 - X_2$. (ii) Find the conditional distribution of X_1 given $X_2 = -2$.

2. Let X_1, X_2 be a random sample from a standard normal distribution. Determine Y_1, Y_2, both linear functions of X_1, X_2, such that $\mathbf{Y} = (Y_1, Y_2)$ has bivariate normal distribution with mean vector $(-2, 3)$ and dispersion matrix $\begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$.

3. Let X_1, X_2 be a random sample from a standard normal distribution. Determine the linear functions of X_1, X_2 that are distributed independently of $(X_1 - X_2)^2$.

3.3 Quadratic Forms and Cochran's Theorem

3.1. *Let* $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I_n})$ *and let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *has the chi-square distribution with* r *degrees of freedom* (χ_r^2) *if and only if* $\mathbf{A}$ *is idempotent and* $R(\mathbf{A}) = r$.

PROOF. If $\mathbf{A}$ is idempotent with rank r, then there exists an orthogonal matrix $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}' \begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{P}.$$

Let $\mathbf{z} = \mathbf{P}\mathbf{y}$. Then $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I_n})$. We have

$$\begin{aligned} \mathbf{y}'\mathbf{A}\mathbf{y} &= \mathbf{y}'\mathbf{P}' \begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{P}\mathbf{y} \\ &= \mathbf{z}' \begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{z} \\ &= z_1^2 + \cdots + z_r^2 \\ &\sim \chi_r^2. \end{aligned}$$

Conversely, suppose $\mathbf{y}'\mathbf{A}\mathbf{y} \sim \chi_r^2$. There exists an orthogonal matrix $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}' \mathrm{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{P},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$. Again, let $\mathbf{z} = \mathbf{P}\mathbf{y}$, so that $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I_n})$. The characteristic function of $\mathbf{y}$ is given by

$$\begin{aligned} \phi(t) &= E(\exp(it\mathbf{y}'\mathbf{A}\mathbf{y})) \\ &= E(\exp(it\sum_{j=1}^{n} \lambda_j z_j^2)) \\ &= \prod_{j=1}^{n} E(\exp(it\lambda_j z_j^2)) \\ &= \prod_{j=1}^{n} (1 - 2it\lambda_j)^{-\frac{1}{2}}, \end{aligned} \tag{12}$$

since $z_j^2 \sim \chi_1^2$. However, since $\mathbf{y}'\mathbf{A}\mathbf{y} \sim \chi_r^2$, its characteristic function is

$$\phi(t) = (1 - 2it)^{-r/2}. \tag{13}$$

Equating (12) and (13), we get

$$(1 - 2it)^r = \prod_{j=1}^{n} (1 - 2it\lambda_j) \tag{14}$$

for all t. The left-hand side of (14) is a polynomial in t with r roots, all equal to $1/2i$. Therefore, the right-hand side also must have the same roots. This is possible precisely when r of the λ_i's are equal to 1, the rest being zero. Therefore, $\mathbf{A}$ is idempotent with rank r. □

3.2. *Let* $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I_n})$ *and let* $\mathbf{A_1}, \mathbf{A_2}$ *be symmetric idempotent matrices. Then* $\mathbf{y}'\mathbf{A_1}\mathbf{y}, \mathbf{y}'\mathbf{A_2}\mathbf{y}$ *are independent if and only if* $\mathbf{A_1}\mathbf{A_2} = \mathbf{0}$.

PROOF. Suppose $\mathbf{A_1A_2} = \mathbf{0}$. Then by **2.3**, $\mathbf{A_1y}$, $\mathbf{A_2y}$ are independent. Hence

$$\mathbf{y'A_1y} = (\mathbf{A_1y})'(\mathbf{A_1y}), \quad \mathbf{y'A_2y} = (\mathbf{A_2y})'(\mathbf{A_2y})$$

are independent, since they are (measurable) functions of independent random variables.

Conversely, let $\mathbf{y'A_1y}, \mathbf{y'A_2y}$ be independent. By **3.1**, $\mathbf{y'A_1y}, \mathbf{y'A_2y}$ are chi-square, and therefore

$$\mathbf{y'A_1y} + \mathbf{y'A_2y} = \mathbf{y'(A_1 + A_2)y}$$

must be chi-square. Again, by **3.1**, $\mathbf{A_1} + \mathbf{A_2}$ is idempotent. Therefore,

$$\begin{aligned}\mathbf{A_1} + \mathbf{A_2} &= (\mathbf{A_1} + \mathbf{A_2})^2 \\ &= \mathbf{A_1^2} + \mathbf{A_2^2} + \mathbf{A_1A_2} + \mathbf{A_2A_1} \\ &= \mathbf{A_1} + \mathbf{A_2} + \mathbf{A_1A_2} + \mathbf{A_2A_1}.\end{aligned}$$

Hence

$$\mathbf{A_1A_2} + \mathbf{A_2A_1} = \mathbf{0}.$$

This gives, upon postmultiplying by $\mathbf{A_2}$,

$$\mathbf{A_1A_2} + \mathbf{A_2A_1A_2} = \mathbf{0}. \tag{15}$$

Premultiply (15) by $\mathbf{A_2}$ to get

$$\mathbf{A_2A_1A_2} + \mathbf{A_2^2A_1A_2} = 2\mathbf{A_2A_1A_2} = \mathbf{0}.$$

Hence $\mathbf{A_2A_1A_2} = \mathbf{0}$. Substituting in (15), we get $\mathbf{A_1A_2} = \mathbf{0}$. □

Remark

We have not used the assumption that $\mathbf{A_1}, \mathbf{A_2}$ are idempotent while proving the sufficiency in **3.2**. The necessity can also be shown to be true without this assumption. The proof employs characteristic functions and is more complicated. Since we will not need the more general result, we omit the proof.

3.3. *Let* $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I_n})$. *Let* $\mathbf{A}$ *be a symmetric idempotent matrix and let* $\boldsymbol{\ell} \in R^n$ *be a nonzero vector. Then* $\mathbf{y'Ay}$ *and* $\boldsymbol{\ell}'\mathbf{y}$ *are independent if and only if* $\mathbf{A}\boldsymbol{\ell} = \mathbf{0}$.

PROOF. We assume, without loss of generality, that $\|\boldsymbol{\ell}\| = 1$. Then $\mathbf{B} = \boldsymbol{\ell\ell}'$ is a symmetric idempotent matrix.

First suppose that $\mathbf{y'Ay}$ and $\boldsymbol{\ell}'\mathbf{y}$ are independent. Then using, as before, the fact that (measurable) functions of independent random variables are independent, we see that $\mathbf{y'Ay}$ and $\mathbf{y'By}$ are independent. It follows from **3.2** that $\mathbf{AB} = \mathbf{0}$, and then it is an easy exercise to show that $\mathbf{A}\boldsymbol{\ell} = \mathbf{0}$.

Conversely, if $\mathbf{A}\boldsymbol{\ell} = \mathbf{0}$, then by **2.3**, $\mathbf{Ay}$ and $\boldsymbol{\ell}'\mathbf{y}$ are independent. Hence $\mathbf{y'Ay} = (\mathbf{Ay})'(\mathbf{Ay})$ and $\boldsymbol{\ell}'\mathbf{y}$ are independent. That completes the proof. □

We now prove a matrix-theoretic formulation of one version of Cochran's theorem.

3.4. *Let* $\mathbf{A_1}, \ldots, \mathbf{A_k}$ *be* $n \times n$ *matrices with* $\sum_{i=1}^{k} \mathbf{A_i} = \mathbf{I}$. *Then the following conditions are equivalent:*

(i) $\sum_{i=1}^{k} R(\mathbf{A_i}) = n.$

(ii) $\mathbf{A_i^2} = \mathbf{A_i}, \quad i = 1, \ldots, k.$

(iii) $\mathbf{A_i}\mathbf{A_j} = \mathbf{0}, \quad i \neq j.$

PROOF. (i) $\Rightarrow$ (iii): Let $\mathbf{A_i} = \mathbf{B_i}\mathbf{C_i}$ be a rank factorization, $i = 1, \ldots, k$. Then

$$\mathbf{B_1}\mathbf{C_1} + \cdots + \mathbf{B_k}\mathbf{C_k} = \mathbf{I},$$

and hence

$$\begin{bmatrix} \mathbf{B_1} & \cdots & \mathbf{B_k} \end{bmatrix} \begin{bmatrix} \mathbf{C_1} \\ \vdots \\ \mathbf{C_k} \end{bmatrix} = \mathbf{I}.$$

Since $\sum_{i=1}^{k} R(\mathbf{A_i}) = n$, $\begin{bmatrix} \mathbf{B_1} & \cdots & \mathbf{B_k} \end{bmatrix}$ is a square matrix, and therefore

$$\begin{bmatrix} \mathbf{C_1} \\ \vdots \\ \mathbf{C_k} \end{bmatrix} \begin{bmatrix} \mathbf{B_1} & \cdots & \mathbf{B_k} \end{bmatrix} = \mathbf{I}.$$

Thus $\mathbf{C_i}\mathbf{B_j} = \mathbf{0}$, $i \neq j$. It follows that for $i \neq j$,

$$\mathbf{A_i}\mathbf{A_j} = \mathbf{B_i}\mathbf{C_i}\mathbf{B_j}\mathbf{C_j} = \mathbf{0}.$$

(iii) $\Rightarrow$ (ii): Since $\sum_{i=1}^{k} \mathbf{A_i} = \mathbf{I}$,

$$\mathbf{A_j}\left(\sum_{i=1}^{k} \mathbf{A_i}\right) = \mathbf{A_j}, \qquad j = 1, \ldots, k.$$

It follows that $\mathbf{A_j^2} = \mathbf{A_j}$.

(ii) $\Rightarrow$ (i): Since $\mathbf{A_i}$ is idempotent, $R(\mathbf{A_i}) = \text{trace}\mathbf{A_i}$. Now

$$\sum_{i=1}^{k} R(\mathbf{A_i}) = \sum_{i=1}^{k} \text{trace}\mathbf{A_i} = \text{trace}\left(\sum_{i=1}^{k} \mathbf{A_i}\right) = n.$$

That completes the proof. □

Problems

1. Let $\mathbf{A}$ be an $n \times n$ matrix. Using Cochran's theorem show that $\mathbf{A}$ is idempotent if and only if $R(\mathbf{A}) + R(\mathbf{I} - \mathbf{A}) = n$.

2. Let $X_1, \dots, X_n$ be a random sample from a standard normal distribution. Show, using Cochran's theorem, that

$$\overline{X} = \sum_{i=1}^{n} X_i \quad \text{and} \quad \sum_{i=1}^{n}(X_i - \overline{X})^2$$

are independently distributed.

3.4 One-Way and Two-Way Classifications

Suppose we have the vector of observations $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I_n})$. The quantity $\mathbf{y}'\mathbf{y} = \sum_{i=1}^{n} y_i^2$ is called the *crude sum of squares*. If we are able to decompose $\mathbf{y}'\mathbf{y}$ as

$$\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{A_1}\mathbf{y} + \cdots + \mathbf{y}'\mathbf{A_k}\mathbf{y},$$

where $\mathbf{A_i}$ are symmetric, and if we verify that $\mathbf{A_i}\mathbf{A_j} = \mathbf{0},\ i \neq j$, then by Cochran's theorem we may conclude that $\mathbf{y}'\mathbf{A_i}\mathbf{y}$ are independent chi-square random variables. The degrees of freedom are then given by the ranks of $\mathbf{A_i}$.

We first illustrate an application to the one-way classification model discussed earlier. The model is

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad i = 1, \dots, k, \quad j = 1, \dots, n_i,$$

where we now assume that ϵ_{ij} are independent $N(0, \sigma^2)$. Suppose we wish to test the hypothesis

$$H_0 : \alpha_1 = \cdots = \alpha_k.$$

Let

$$z_{ij} = \frac{y_{ij} - \mu - \alpha}{\sigma},$$

which is standard normal if H_0 is true and where α denotes the common value of $\alpha_1, \dots, \alpha_k$. Let $\mathbf{z}$ be the vector

$$(z_{11}, \dots, z_{1n_1}; z_{21}, \dots, z_{2n_2}; \cdots; z_{k1}, \dots, z_{kn_k}),$$

and let $n = \sum_{i=1}^{k} n_i$. We use the *dot notation*. Thus

$$z_{i.} = \sum_{j=1}^{n_i} z_{ij}, \qquad \overline{z}_{i.} = \frac{z_{i.}}{n_i},$$

and a similar notation is used when there are more than two subscripts.

We have the identity

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} z_{ij}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.} + \overline{z}_{i.} - \overline{z}_{..} + \overline{z}_{..})^2 \tag{16}$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.})^2 + \sum_{i=1}^{k} n_i(\overline{z}_{i.} - \overline{z}_{..})^2 + n\overline{z}_{..}^2 \ , \tag{17}$$

since the cross-product terms equal zero. For example,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.})(\overline{z}_{i.} - \overline{z}_{..}) = \sum_{i=1}^{k}(\overline{z}_{i.} - \overline{z}_{..})\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.}) = 0. \tag{18}$$

Let $\mathbf{A_1}$, $\mathbf{A_2}$, $\mathbf{A_3}$ be symmetric matrices such that the quadratic forms

$$\mathbf{z}'\mathbf{A_1}\mathbf{z}, \quad \mathbf{z}'\mathbf{A_2}\mathbf{z}, \quad \mathbf{z}'\mathbf{A_3}\mathbf{z}$$

equal the three forms in (17), respectively. Since each form is a sum of squares, $\mathbf{A_i}$ are, in fact, positive semidefinite. A moment's reflection will show that (18) implies that $\mathbf{A_1}\mathbf{A_2} = \mathbf{0}$ (why?). Similarly,

$$\mathbf{A_1}\mathbf{A_3} = \mathbf{A_2}\mathbf{A_3} = \mathbf{0}.$$

We conclude by Cochran's theorem that $\mathbf{z}'\mathbf{A_1}\mathbf{z}$, $\mathbf{z}'\mathbf{A_2}\mathbf{z}$ are independent chi-square variables.

It remains to find the degrees of freedom. These are $R(\mathbf{A_1})$, $R(\mathbf{A_2})$, respectively. Since $\mathbf{A_1}$, $\mathbf{A_2}$ are idempotent (this follows again by Cochran's theorem), it is sufficient to find trace$\mathbf{A_1}$, trace$\mathbf{A_2}$.

For $i = 1, \ldots, k, \ \ j = 1, \ldots, n_i$, let $\mathbf{z^{ij}}$ be the column vector of order n with 1 at coordinate (i, j) and zeros elsewhere. Then

$$\text{trace}\mathbf{A_1} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{z^{ij}})'\mathbf{A_1}\mathbf{z^{ij}}.$$

However,

$$\begin{aligned}(\mathbf{z^{ij}})'\mathbf{A_1}\mathbf{z^{ij}} &= \left(1 - \frac{1}{n_i}\right)^2 + \frac{n_i - 1}{n_i^2} \\ &= \frac{n_i - 1}{n_i^2} \times n_i = \frac{n_i - 1}{n_i}.\end{aligned}$$

Thus

$$\text{trace}\mathbf{A_1} = \sum_{i=1}^{k} n_i \times \frac{n_i - 1}{n_i} = n - k.$$

Similarly, trace$\mathbf{A_2} = k - 1$.

Alternatively, we observe that for each i, $\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.})^2$ is distributed as chi-square with $n_i - 1$ degrees of freedom, and $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.})^2$, being a sum of independent chi-squares, is distributed as chi-square with

$$\sum_{i=1}^{k}(n_i - 1) = n - k$$

degrees of freedom. Similarly, we can show that $\mathbf{z}'\mathbf{A_2}\mathbf{z}$ is distributed as chi-square with $k-1$ degrees of freedom.

Therefore, under H_0,

$$\frac{\sum_{i=1}^{k} n_i(\overline{z}_{i.} - \overline{z}_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_{i.})^2/(n-k)} \sim F(k-1, n-k).$$

In terms of y_{ij} we can write this as

$$\frac{\sum_{i=1}^{k} n_i(\overline{y}_{i.} - \overline{y}_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_{i.})^2/(n-k)} \sim F(k-1, n-k),$$

and it can be used to test H_0. This test statistic can be justified on intuitive grounds. If the difference *between* populations is large, in comparison to the difference *within* each population, then the statistic will be large, and that is when we reject H_0. The statistic can also be shown to have certain optimal properties.

We now describe two-way classification without interaction. Suppose there are two factors, one at a levels and the other at b levels. We have one observation for every combination of a level of the first factor and a level of the second factor. The model is

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \qquad i = 1, \dots, a, \quad j = 1, \dots, b,$$

where ϵ_{ij} are i.i.d. $N(0, \sigma^2)$. Here α_i denotes the effect of the ith level of the first factor and β_j the effect of the jth level of the second factor. Suppose we want to test the hypothesis $H_0 : \alpha_1 = \cdots = \alpha_a$. The subsequent discussion is under the assumption that H_0 is true. Let α be the common value of $\alpha_1, \dots, \alpha_a$. Let

$$z_{ij} = \frac{y_{ij} - \mu - \alpha - \beta_j}{\sigma}.$$

Then z_{ij} are i.i.d. $N(0, 1)$. Let z be the vector

$$(z_{11}, \dots, z_{1b}, z_{21}, \dots, z_{2b}, \dots, z_{a1}, \dots, z_{ab}).$$

We have

$$z_{ij} = (z_{ij} - \overline{z}_{i.} - \overline{z}_{.j} + \overline{z}_{..}) + (\overline{z}_{i.} - \overline{z}_{..}) + (\overline{z}_{.j} - \overline{z}_{..}) + \overline{z}_{..},$$

and as before,

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b} z_{ij}^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij} - \overline{z}_{i.} - \overline{z}_{.j} + \overline{z}_{..})^2 \\ &\quad + b\sum_{i=1}^{a}(\overline{z}_{i.} - \overline{z}_{..})^2 + a\sum_{j=1}^{b}(\overline{z}_{.j} - \overline{z}_{..})^2 + ab\overline{z}_{..}^2,\end{aligned}$$

since the cross-product terms vanish. Thus we can write

$$\mathbf{z}'\mathbf{z} = \mathbf{z}'\mathbf{A_1}\mathbf{z} + \mathbf{z}'\mathbf{A_2}\mathbf{z} + \mathbf{z}'\mathbf{A_3}\mathbf{z} + \mathbf{z}'\mathbf{A_4}\mathbf{z},$$

where $\mathbf{A_1}, \mathbf{A_2}, \mathbf{A_3}, \mathbf{A_4}$ are symmetric (in fact, positive semidefinite) matrices such that $\mathbf{A_i}\mathbf{A_j} = \mathbf{0}, i \neq j$. By Cochran's theorem $\mathbf{z}'\mathbf{A_i}\mathbf{z}$ are independent chi-square random variables. We leave it as an exercise to check that $R(\mathbf{A_1}) = (a-1)(b-1)$, $R(\mathbf{A_2}) = a-1$. Note that

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij} - \bar{z}_{i.} - \bar{z}_{.j} + \bar{z}_{..})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

and

$$\sum_{i=1}^{a}(\bar{z}_{i.} - \bar{z}_{..})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2.$$

Thus

$$\frac{b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2/(a-1)}{\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2/(a-1)(b-1)}$$

is distributed as F with $(a-1, (a-1)(b-1))$ degrees of freedom and can be used to test H_0. A test of $H_0 : \beta_1 = \cdots = \beta_b$ is constructed similarly.

If we take more than one, but an equal number of, observations per every level combination, then the model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

$i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, n$, where n denotes the number of observations per each level combination. The analysis in this case is similar, and one can show that under $H_0 : \alpha_1 = \cdots = \alpha_a$,

$$\frac{bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2/(a-1)}{n\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2/(abn - a - b + 1)}$$

is distributed as F with $(a-1, abn - a - b + 1)$ degrees of freedom.

If $k = 1, \dots, n_{ij}$, then the statistic, in general, is not expressible in a compact form. However, if n_{ij} satisfy the relation

$$n_{ij} = \frac{n_{i.}n_{.j}}{n_{..}},$$

then the F-statistic can be derived in a similar way as for the case of equal n_{ij}'s.

Two-way classification with interaction. Consider the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

$i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, n$, where $n > 1$ and ϵ_{ijk} are i.i.d. $N(0, \sigma^2)$. Then under the hypothesis that the γ_{ij}'s are all equal, it can be shown that the statistic

$$\frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2} \times \frac{ab(n-1)}{(a-1)(b-1)}$$

is distributed as $F((a-1)(b-1), ab(n-1))$. The proof is left as an exercise.

Problems

1. Three teaching methods, A, B, C, are to be compared. Each method was administered to a group of 4 students, and the scores obtained by the students on a test are given below. Carry out an F-test at level of significance 0.01 to decide whether the mean scores under the three methods are significantly different.

Method A: 75, 79, 71, 69

Method B: 82, 93, 86, 88

Method C: 78, 81, 76, 81.

2. The defective items produced on 3 machines M1, M2, M3 by 4 operators O1, O2, O3, O4 are given in the following table:

	O1	O2	O3	O4
M1	29	25	36	22
M2	28	19	40	28
M3	35	28	34	30

Carry out a test for significant differences between the machines as well as significant differences between the operators.

3.5 General Linear Hypothesis

We now bring in a normality assumption in our linear model and assume

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I_n}),$$

where $\mathbf{y}$ is $n \times 1$, $\mathbf{X}$ is $n \times p$, and $\boldsymbol{\beta}$ is $p \times 1$. Let $R(\mathbf{X}) = r$.

We have seen that

$$\text{RSS} = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

and the minimum is attained at

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}.$$

5.1.

$$\frac{RSS}{\sigma^2} \sim \chi^2_{n-r}.$$

Proof. As before, let $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$. From the proof of **5.2** of Chapter 2,

$$\text{RSS} = \mathbf{y}'\mathbf{P}\mathbf{y} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{P}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

since $\mathbf{PX} = \mathbf{0}$. Thus by **3.1**, it follows that RSS/σ^2 is distributed as χ^2. The degrees of freedom is $R(\mathbf{P})$, and this was seen to be $n - r$. □

Now consider the hypothesis $H : \mathbf{L}\boldsymbol{\beta} = \mathbf{z}$. We make the assumption, as before, that $\mathcal{R}(\mathbf{L}) \subset \mathcal{R}(\mathbf{X})$ and that the equation $\mathbf{L}\boldsymbol{\beta} = \mathbf{z}$ is consistent. Following Section 6 of Chapter 2, let $\mathbf{L} = \mathbf{W}\mathbf{X}'\mathbf{X}$, $\mathbf{W}\mathbf{X}' = \mathbf{T}$. Then

$$\mathrm{RSS}_H = \min_{\boldsymbol{\beta}:\mathbf{L}\boldsymbol{\beta}=\mathbf{z}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

is attained at $\tilde{\boldsymbol{\beta}}$, where

$$\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \mathbf{X}'\mathbf{X}\mathbf{L}'(\mathbf{T}\mathbf{T}')^-(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{z}).$$

Therefore,

$$\begin{aligned} \mathbf{X}\tilde{\boldsymbol{\beta}} &= (\mathbf{I} - \mathbf{P})\mathbf{y} - \mathbf{T}'(\mathbf{T}\mathbf{T}')^-(\mathbf{T}\mathbf{y} - \mathbf{z}) \\ &= (\mathbf{I} - \mathbf{P})\mathbf{y} - \mathbf{T}'(\mathbf{T}\mathbf{T}')^-(\mathbf{T}\mathbf{y} - \mathbf{T}\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{X}\boldsymbol{\beta} - \mathbf{z}). \end{aligned} \tag{19}$$

If H is true, then $\mathbf{T}\mathbf{X}\boldsymbol{\beta} = \mathbf{L}\boldsymbol{\beta} = \mathbf{z}$, and therefore by (19),

$$\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y} + \mathbf{U}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\mathbf{U} = \mathbf{T}'(\mathbf{T}\mathbf{T}')^-\mathbf{T}$. Thus

$$\mathrm{RSS}_H = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{P}\mathbf{y} + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{U}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

as $\mathbf{P}\mathbf{U} = \mathbf{0}$. Since $\mathbf{U}$ is idempotent, we conclude that

$$\frac{\mathrm{RSS}_H - \mathrm{RSS}}{\sigma^2} \sim \chi^2_{R(\mathbf{U})}. \tag{20}$$

Also, since $\mathbf{P}\mathbf{U} = \mathbf{0}$, RSS and $\mathrm{RSS}_H - \mathrm{RSS}$ are independently distributed.

We have

$$\begin{aligned} R(\mathbf{U}) &= R(\mathbf{T}'(\mathbf{T}\mathbf{T}')^-\mathbf{T}) = \mathrm{trace}\mathbf{T}'(\mathbf{T}\mathbf{T}')^-\mathbf{T} \\ &= \mathrm{trace}(\mathbf{T}\mathbf{T}')^-\mathbf{T}\mathbf{T}' = R((\mathbf{T}\mathbf{T}')^-\mathbf{T}\mathbf{T}') \\ &= R(\mathbf{T}\mathbf{T}') = R(\mathbf{T}). \end{aligned}$$

It follows from **6.2** of Chapter 2 that $R(\mathbf{U}) = R(\mathbf{L})$. We conclude that

$$\frac{(\mathrm{RSS}_H - \mathrm{RSS})/R(\mathbf{L})}{\mathrm{RSS}/(n-r)} \sim F(R(\mathbf{L}), n-r)$$

and can be used to test H.

Problems

1. Consider the model $y_1 = \theta_1 + \theta_2 + \epsilon_1$, $y_2 = 2\theta_1 + \epsilon_2$, $y_3 = \theta_1 - \theta_2 + \epsilon_3$, where ϵ_i, $i = 1, 2, 3$, are i.i.d. $N(0, \sigma^2)$. Derive the F-statistic to test $\theta_1 = \theta_2$.

2. Consider the model $Y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + e$ with the usual assumptions. Derive the test for the hypothesis $\beta_2 = 0$ and also for the hypothesis $\beta_3 = 0$ using the data given in Table 3.1.

TABLE 3.1.

Y	x_1	x_2
10	21	2.67
12	32	3.12
6	46	2.11
14	91	4.21
20	20	6.43
5	65	1.76
8	26	2.88
15	74	6.15
13	48	7.20
21	81	9.12
14	93	3.21
11	88	4.87
18	46	5.38
17	24	8.71
27	11	8.11

3.6 Extrema of Quadratic Forms

6.1. *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then*

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1,$$

and the maximum is attained at any eigenvector of $\mathbf{A}$ *corresponding to* λ_1.

PROOF. Write $\mathbf{A} = \mathbf{P}'\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{P}$, where $\mathbf{P}$ is orthogonal and where, as usual, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of $\mathbf{A}$. Then for $\mathbf{x} \neq \mathbf{0}$,

$$\begin{aligned}\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} &= \frac{\mathbf{y}'\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{y}}{\mathbf{y}'\mathbf{y}}, \qquad (\mathbf{y} = \mathbf{P}\mathbf{x}) \\ &= \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} \\ &\leq \frac{\lambda_1 y_1^2 + \cdots + \lambda_1 y_n^2}{y_1^2 + \cdots + y_n^2} = \lambda_1.\end{aligned}$$

Therefore,

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq \lambda_1.$$

Clearly, if $\mathbf{x}$ is an eigenvector corresponding to λ_1, then

$$\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1,$$

and the result is proved. □

6.2. *Let* $\mathbf{A}, \mathbf{B}$ *be* $n \times n$ *matrices where* $\mathbf{A}$ *is symmetric and* $\mathbf{B}$ *is positive definite. Then*

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mu,$$

where μ *is the largest eigenvalue of* $\mathbf{A}\mathbf{B}^{-1}$.

PROOF.

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{(\mathbf{x}'\mathbf{B}^{1/2})(\mathbf{B}^{1/2}\mathbf{x})} = \max_{\mathbf{y} \neq \mathbf{0}} \frac{\mathbf{y}'\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\mathbf{y}}{\mathbf{y}'\mathbf{y}}.$$

By **6.1** this last expression equals the maximum eigenvalue of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$. However, $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and $\mathbf{A}\mathbf{B}^{-1}$ have the same eigenvalues, and the proof is complete. □

In **6.2** the maximum is attained at any eigenvector of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ corresponding to the eigenvalue μ.

6.3. *Let* $\mathbf{B}$ *be a positive definite* $n \times n$ *matrix and let* $\mathbf{y} \in R^n$. *Then*

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{(\mathbf{x}'\mathbf{y})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{y}'\mathbf{B}^{-1}\mathbf{y}.$$

PROOF. We have

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{(\mathbf{x}'\mathbf{y})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{y}\mathbf{y}'\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}},$$

which is the largest eigenvalue of $\mathbf{y}\mathbf{y}'\mathbf{B}^{-1}$ by **6.1**. Again the eigenvalues of $\mathbf{y}\mathbf{y}'\mathbf{B}^{-1}$ and $\mathbf{B}^{-1/2}\mathbf{y}\mathbf{y}'\mathbf{B}^{-1/2}$ are equal. Since the latter matrix is symmetric and has rank 1, it has only one nonzero eigenvalue counting multiplicity. The eigenvalue equals

$$\operatorname{trace}\mathbf{B}^{-1/2}\mathbf{y}\mathbf{y}'\mathbf{B}^{-1/2} = \operatorname{trace}\mathbf{y}'\mathbf{B}^{-1}\mathbf{y} = \mathbf{y}'\mathbf{B}^{-1}\mathbf{y},$$

and the proof is complete. □

We may rewrite **6.3** as

$$(\mathbf{x}'\mathbf{y})^2 \leq (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{y}'\mathbf{B}^{-1}\mathbf{y}) \tag{21}$$

for any positive definite matrix $\mathbf{B}$. An alternative proof of (21) can be given using the Cauchy–Schwarz inequality:

$$\left(\sum u_i v_i\right)^2 \leq \left(\sum u_i^2\right)\left(\sum v_i^2\right).$$

Problems

1. If $\mathbf{A}$ is an $n \times n$ symmetric matrix, show that the largest eigenvalue of $\mathbf{A}$ cannot be less than $\frac{1}{n}\sum_{i,j=1}^{n} a_{ij}$.

2. Show that

$$\max_{(x_1,x_2)\neq(0,0)} \frac{(x_1 - x_2)^2}{5x_1^2 - 4x_1x_2 + x_2^2} = 4.$$

3.7 Multiple Correlation and Regression Models

Suppose the random vector (of order $p + 1$)

$$(y, x_1, \dots, x_p)'$$

has the dispersion matrix

$$\mathbf{V} = \begin{bmatrix} \sigma^2 & \mathbf{u}' \\ \mathbf{u} & \Sigma \end{bmatrix}, \tag{22}$$

where Σ is positive definite of order p.

We wish to find the linear combination

$$\boldsymbol{\alpha}'\mathbf{x} = \alpha_1 x_1 + \dots + \alpha_p x_p, \qquad \boldsymbol{\alpha} \neq \mathbf{0},$$

that has maximum correlation with y. The maximum value is called the *multiple correlation coefficient* between y and $x_1, \dots, x_p$, denoted by $r_{y(x_1,\dots,x_p)}$.

Thus

$$\begin{aligned} r^2_{y(x_1,\dots,x_p)} &= \max_{\boldsymbol{\alpha}\neq\mathbf{0}} \{\text{correlation}(y, \boldsymbol{\alpha}'\mathbf{x})\}^2 \\ &= \max_{\boldsymbol{\alpha}\neq\mathbf{0}} \frac{(\text{cov}(y, \boldsymbol{\alpha}'\mathbf{x}))^2}{\text{var}(y)\text{var}(\boldsymbol{\alpha}'\mathbf{x})} \\ &= \max_{\boldsymbol{\alpha}\neq\mathbf{0}} \frac{(\boldsymbol{\alpha}'\mathbf{u})^2}{\sigma^2 \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}} \\ &= \frac{\mathbf{u}'\Sigma^{-1}\mathbf{u}}{\sigma^2}, \end{aligned}$$

by **6.3**. The maximum is attained at $\boldsymbol{\alpha} = \Sigma^{-1}\mathbf{u}$.

We get another expression for $r^2_{y(x_1,\dots,x_p)}$ as follows. Let $\mathbf{Z} = \mathbf{V}^{-1}$. Then, by (5),

$$\frac{1}{z_{11}} = \sigma^2 - \mathbf{u}'\Sigma^{-1}\mathbf{u}.$$

Hence

$$r^2_{y(x_1,\dots,x_p)} = 1 - \frac{1}{\sigma^2 z_{11}}.$$

Suppose the vector $(y, x_1, \dots, x_p)'$ has the multivariate normal distribution with mean vector $(\tau, \boldsymbol{\mu}')'$ and dispersion matrix $\mathbf{V}$ partitioned as in (22). The conditional distribution of y given $x_1, \dots, x_p$ is

$$N(\tau + \mathbf{u}'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}), \sigma^2 - \mathbf{u}'\Sigma^{-1}\mathbf{u}).$$

Thus the conditional variance of y given $\mathbf{x} = (x_1, \ldots, x_p)'$ is

$$\sigma^2 \left(1 - r^2_{y(x_1,\ldots,x_p)}\right).$$

The conditional expectation of y given $\mathbf{x}$ (also known as the line of regression of y on $\mathbf{x}$) is

$$\mathbf{u}'\Sigma^{-1}\mathbf{x} + \tau - \mathbf{u}'\Sigma^{-1}\boldsymbol{\mu},$$

and recall that $\mathbf{u}'\Sigma^{-1}\mathbf{x}$ is precisely the linear combination of $x_1, \ldots, x_p$ that has maximum correlation with y. Thus the multiple correlation coefficient admits special interpretation if the distribution of the variables is multivariate normal.

Suppose there are random variables $y, x_1, \ldots, x_p$ in a given situation and we want to study the relationship between y and the x_i's. In particular, we may want to predict the value of y given the values of the x_i's. We first observe $x_1, \ldots, x_p$. Then treating these as fixed, we take an observation on y, after conducting any experiment in the process that may be necessary. If we now stipulate the model

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

with the usual assumptions on $\mathrm{var}(y)$, then we have a linear model. The $E(y)$ term in the model is to be interpreted as the conditional expectation $E(y|x_1, \ldots, x_p)$. If we have n data points on the variables, then the model can be written as

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \qquad i = 1, \ldots, n, \tag{23}$$

and can be analyzed by methods developed for a linear model. Such a model is called a *regression model.* Since the x_{ij}'s are observations on a random variable, we can assume the model to be of full rank. In fact, under mild assumptions on the distribution of $x_1, \ldots, x_p$ it can be proved that the coefficient matrix of the model will have full column rank with probability 1. Thus the terms "full-rank model" and "regression model" are used interchangeably.

Consider the model (23) where $y_1, \ldots, y_n$ are independent $N(0, \sigma^2)$. The model can be expressed as

$$E(\mathbf{y}) = \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$$

where $\mathbf{e}$ is the column vector of all ones. We assume the model to be of full rank, and therefore the BLUEs of $\beta_0, \ldots, \beta_p$ are given by

$$\begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \vdots \\ \widehat{\beta_p} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{e}' \\ \mathbf{X}' \end{bmatrix} \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{e}' \\ \mathbf{X}' \end{bmatrix} \mathbf{y}$$

$$= \begin{bmatrix} n & \mathbf{e}'\mathbf{X} \\ \mathbf{X}'\mathbf{e} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{e}'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix}.$$

It can be verified that

$$\begin{bmatrix} n & \mathbf{e}'\mathbf{X} \\ \mathbf{X}'\mathbf{e} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{1}{n - \mathbf{e}'\mathbf{M}\mathbf{e}} & \mathbf{z}' \\ \mathbf{z} & (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1} \end{bmatrix},$$

where $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{Q} = \mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}'$ and

$$\mathbf{z} = -\frac{1}{n}(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}.$$

Thus the BLUEs of $\beta_1, \dots, \beta_p$ are given by

$$\begin{bmatrix} \widehat{\beta_1} \\ \vdots \\ \widehat{\beta_p} \end{bmatrix} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{y}. \tag{24}$$

The sample dispersion matrix of the vaiables $y, x_1, \dots, x_p$ is computed as follows. Let

$$\mathbf{S} = \begin{bmatrix} y_1 & \\ \vdots & \mathbf{X} \\ y_n & \end{bmatrix}.$$

Then the dispersion matrix is

$$\begin{aligned} \mathbf{S}'\mathbf{S} - \frac{1}{n}\mathbf{S}'\mathbf{e}\mathbf{e}'\mathbf{S} &= \mathbf{S}'\mathbf{Q}\mathbf{S} \\ &= \begin{bmatrix} \mathbf{y}'\mathbf{Q}\mathbf{y} & \mathbf{y}'\mathbf{Q}\mathbf{X} \\ \mathbf{X}'\mathbf{Q}\mathbf{y} & \mathbf{X}'\mathbf{Q}\mathbf{X} \end{bmatrix}. \end{aligned}$$

Thus the linear function of $x_1, \dots, x_p$ that has maximum correlation with y is obtained by taking

$$\boldsymbol{\alpha} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{y},$$

and this coincides with (24). To summarize, the linear function $\beta_1 x_1 + \dots + \beta_p x_p$ having maximum correlation with y is obtained by taking $\beta_i = \widehat{\beta_i}$, the least squares estimate of β_i.

Let

$$\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x_{i1} + \dots + \widehat{\beta_p}x_{p1}, \qquad i = 1, \dots, n,$$

be the predicted value of y_i. Let

$$\widehat{\mathbf{y}} = (\widehat{y_1}, \dots, \widehat{y_n})'$$

and let $\overline{\widehat{y}} = \frac{1}{n}\mathbf{e}'\widehat{\mathbf{y}}$. Then

$$\begin{aligned} r^2_{y(x_1,\ldots,x_p)} &= \{\text{correlation}(y, \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_p x_p)\}^2 \\ &= \{\text{correlation}(y, \widehat{y})\}^2 \\ &= \frac{\{\sum_{i=1}^n (y_i - \overline{y})(\widehat{y}_i - \overline{\widehat{y}})\}^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (\widehat{y}_i - \overline{\widehat{y}})^2}. \end{aligned}$$

The square root of the expression above is the multiple correlation coefficient calculated from a sample, and it is known as the *coefficient of determination*.

We now derive the F-statistic for the hypothesis $H : \beta_1 = \cdots = \beta_p = 0$. We have

$$\text{RSS} = \sum_{i=1}^n (y_i - \widehat{y}_i)^2,$$

and the corresponding degrees of freedom are $n - p - 1$.

To find RSS_H we must minimize

$$\sum_{i=1}^n (y_i - \beta_0)^2,$$

and this is achieved when $\beta_0 = \overline{y}$. The degrees of freedom now are p. Thus the statistic

$$\frac{(\text{RSS}_H - \text{RSS})/p}{\text{RSS}/(n-p-1)} = \frac{\sum_{i=1}^n (y_i - \overline{y})^2 - \sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \widehat{y}_i)^2} \times \frac{n-p-1}{p} \tag{25}$$

is $F(p, n-p-1)$ if H is true. For the relationship between this statistic and r^2 see Exercise 19.

Problems

1. Let (X_1, X_2, X_3) follow a trivariate normal distribution with mean $\mathbf{0}$ and dispersion matrix $\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$. Find the multiple correlation coefficient between X_1 and X_2, X_3.

2. Table 3.2 gives y, the score obtained in the final examination by 12 students; x_1, their IQs; x_2, the score in the midterm examination; and x_3, the score in the homework. Fit the model $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Find the multiple correlation coefficient (actually the coefficient of determination) between y and x_1, x_2, x_3. Test the hypothesis that $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$.

TABLE 3.2.

y	x_1	x_2	x_3
89	115	82	97
67	110	65	88
56	104	62	90
94	123	83	92
43	102	48	78
32	103	41	76
77	116	71	89
87	112	84	94
86	110	81	95
90	121	88	99
42	107	49	85
56	108	43	79

3.8 Exercises

1. Let $\mathbf{A}$ be a symmetric nonsingular matrix, let $\mathbf{X} = \mathbf{A}^{-1}$, and suppose $\mathbf{A}, \mathbf{X}$ are conformally partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{bmatrix}.$$

Then, assuming that the inverses exist, show that

$$\begin{aligned} \mathbf{U} &= (\mathbf{B} - \mathbf{C}\mathbf{D}^{-1}\mathbf{C}')^{-1} = \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{C}\mathbf{W}\mathbf{C}'\mathbf{B}^{-1}, \\ \mathbf{W} &= (\mathbf{D} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}'\mathbf{U}\mathbf{C}\mathbf{D}^{-1}, \\ \mathbf{V} &= -\mathbf{B}^{-1}\mathbf{C}\mathbf{W} = -\mathbf{U}\mathbf{C}\mathbf{D}^{-1}. \end{aligned}$$

2. Consider the partitioned matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

and suppose $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A})$, $\mathcal{R}(\mathbf{C}) \subset \mathcal{R}(\mathbf{A})$. Let $\tilde{\mathbf{D}} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-}\mathbf{B}$ be the "generalized Schur complement" of $\mathbf{D}$ in $\mathbf{X}$. Then show that $\tilde{\mathbf{D}}$ is well-defined and

$$R(\mathbf{X}) = R(\mathbf{A}) + R(\tilde{\mathbf{D}}).$$

Similarly, if $\mathcal{C}(\mathbf{C}) \subset \mathcal{C}(\mathbf{D})$, $\mathcal{R}(\mathbf{B}) \subset \mathcal{R}(\mathbf{D})$, then show that $\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-}\mathbf{C}$ is well-defined and $R(\mathbf{X}) = R(\mathbf{D}) + R(\tilde{\mathbf{A}})$.

3. Let $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$ and suppose $\mathbf{y}$, $\boldsymbol{\mu}$, and Σ are conformally partitioned as in (10). Obtain the conditional distribution of $\mathbf{y_2}$ given $\mathbf{y_1}$ when Σ is possibly singular.

4. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices. Show that

$$R\begin{bmatrix} \mathbf{A}+\mathbf{B} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} \end{bmatrix} = R(\mathbf{A}) + R(\mathbf{B}).$$

5. If $\mathbf{A}$ is a positive definite matrix and $\mathbf{A} \geq \mathbf{I}$, show that any Schur complement in $\mathbf{A}$ is also $\geq \mathbf{I}$ (of the appropriate order).

6. If $\mathbf{A}$ is an $n \times n$ matrix, then show that $R(\mathbf{A}) + R(\mathbf{I} - \mathbf{A}) = n + R(\mathbf{A} - \mathbf{A}^2)$. Conclude that $\mathbf{A}$ is idempotent if and only if $R(\mathbf{A}) + R(\mathbf{I} - \mathbf{A}) = n$.

7. Let $\mathbf{y} \sim N(\mathbf{0}, \Sigma)$, where Σ is a positive semidefinite $n \times n$ matrix of rank r. Show that there exists an $n \times r$ matrix $\mathbf{B}$ such that $\mathbf{y} = \mathbf{Bx}$, where $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I_r})$ and $\Sigma = \mathbf{BB}'$.

8. Let $\mathbf{y} \sim N(\mathbf{0}, \Sigma)$, where Σ is an $n \times n$ positive semidefinite matrix that is possibly singular. Let $\mathbf{A}$ be a symmetric matrix. Prove that $\mathbf{y}'\mathbf{Ay}$ has chi-square distribution if and only if $\Sigma\mathbf{A}\Sigma\mathbf{A}\Sigma = \Sigma\mathbf{A}\Sigma$, in which case the degrees of freedom is $R(\mathbf{A}\Sigma)$.

9. Let $\mathbf{A}, \mathbf{B}$ be symmetric idempotent matrices such that $\mathbf{A} \geq \mathbf{B}$. Then prove that $\mathbf{A} - \mathbf{B}$ is idempotent. State the analogous result for distribution of quadratic forms.

10. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. Show that $\mathbf{A}$ is idempotent if and only if $\mathbf{A}^4 = \mathbf{A}$.

11. Let $\mathbf{A}, \mathbf{B}$ be idempotent matrices such that $\mathbf{A} + \mathbf{B}$ is also idempotent. Does it follow that $\mathbf{AB} = \mathbf{0}$?

12. Let $\mathbf{A}$ be a square matrix. Show that $\mathbf{A}$ is *tripotent* (i.e., $\mathbf{A}^3 = \mathbf{A}$) if and only if

$$R(\mathbf{A}) = R(\mathbf{A} + \mathbf{A}^2) + R(\mathbf{A} - \mathbf{A}^2).$$

13. If $\mathbf{A} \geq \mathbf{B}$ and $R(\mathbf{A}) = R(\mathbf{B})$, then show that $\mathbf{A}^+ \leq \mathbf{B}^+$.

14. Let $\mathbf{y}$ be an $n \times 1$ vector with multivariate normal distribution. Show that $y_1, \dots, y_n$ are pairwise independent if and only if they are mutually independent.

15. Let $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I_n})$. Find the conditional distibution of $\mathbf{y}$ given $y_1 + \cdots + y_n = 0$.

16. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ be $n \times n$ matrices such that $\mathbf{A}$ is nonsingular and suppose $\mathbf{AC} = \mathbf{CA}$. Then show that

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{AD} - \mathbf{CB}|.$$

17. Let $\mathbf{P}$ be an orthogonal matrix and let $\mathbf{Q}$ be obtained by deleting the first row and column of $\mathbf{P}$. Then show that p_{11} and $|\mathbf{Q}|$ are equal in absolute value.

18. Consider the model

$$y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}, \qquad i = 1, \dots, a, \quad j = 1, \dots, b,$$

where ϵ_{ij} are i.i.d. $N(0, \sigma^2)$. Derive a necessary and sufficient condition for

$$\sum_{i=1}^{a} c_i \alpha_i + \sum_{j=1}^{b} d_j \beta_j$$

to be estimable.

19. In an agricultural experiment, 3 fertilizers are to be compared using 36 plots. Let α_i be the effect of the ith fertilizer, $i = 1, 2, 3$. Write the appropriate linear model. How many plots should be allocated to each fertilizer if we want to estimate $\alpha_1 + 2\alpha_2 + \alpha_3$ with maximum precision?

20. Seven varieties of guayule, a Mexican rubber plant, were compared with respect to yield of rubber. These were planted in five blocks of seven plots each. The yield, measured in suitable units, is given in the following table, where each row represents a variety and each column represents a block.

4.46	9.14	6.22	4.42	6.64
1.43	4.04	6.29	6.78	7.32
2.43	1.74	5.29	5.48	6.32
2.08	5.02	2.29	6.46	7.26
6.43	4.01	7.69	6.73	8.12
2.96	4.35	2.11	5.43	1.21
6.45	5.84	3.87	6.71	5.32

(i) Test whether there is a significant difference among the varieties.
(ii) Test whether the average yield of varieties 1, 5, 7 is different from that of varities 3, 4, 6.
(iii) Construct a 95 percent confidence interval for the yield of variety 2.

21. The angles $\theta_1, \theta_2, \theta_3$ of a triangular field were measured in an aerial survey, and the observations y_1, y_2, y_3 were obtained. Set up a linear model and making the necessary assumptions derive a test for the hypothesis that $\theta_1 = \theta_2 = \theta_3$.

22. We follow the notation of Section 7 here. Show that following assertions hold.

(i) $\mathbf{e}'\mathbf{y} = \mathbf{e}'\widehat{\mathbf{y}}$, and hence $\overline{y} = \overline{\widehat{y}}$.
(ii) $(\mathbf{y} - \widehat{\mathbf{y}})'\widehat{\mathbf{y}} = 0$.
(iii) $\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 + \sum_{i=1}^{n}(\widehat{y_i} - \overline{y})^2$.
(iv) The statistic in (25) is $\frac{r^2}{1-r^2} \times \frac{n-p-1}{p}$, where $r^2 = r^2_{y(x_1,\ldots,x_p)}$.

(Part (iv) can be interpreted as follows: If y bears a relationship with $x_1, \ldots, x_p$ that is close to linear, then r^2 is close to 1 and $\frac{r^2}{1-r^2}$ is large. This also indicates that $\beta_1, \ldots, \beta_p$ are significantly different from zero, and therefore we must reject H. Therefore, the fact that the F-statistic for H is $\frac{r^2}{1-r^2}$ (up to a constant) is intuitively justified.)

3.9 Hints and Solutions

Section 4

1. Answer: The F-statistic is $\frac{398/2}{138/9} = 12.97$. The Table value for (2, 9) degrees of freedom and level of significance 0.01 is 8.02. Hence we reject the hypothesis that the means under the three methods are the same.

2. Answer: The error sum of squares is 92.5. The F-statistic for the difference between the machines is 1.0216, which is not significant. The F-statistic for the difference between the operators is 5.9027, which is significant.

Section 5

1. Answer: $RSS = \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 4y_1y_2 + 4y_2y_3 + 8y_1y_3)$, $RSS_H = \frac{1}{2}(y_1 - y_2)^2 + y_3^2$. The test statistic is $\frac{RSS_H - RSS}{RSS}$, distributed as $F(1, 1)$.

2. Answer: $RSS = 153.86$. For the hypothesis $\beta_2 = 0$, $RSS_H = 158.89$ and $F = 0.12$, which is not significant. For the hypothesis $\beta_3 = 0$, $RSS_H = 2986.32$ and $F = 67.5$, which is significant.

Section 7

1. Answer: $\frac{1}{\sqrt{3}}$.

2. Answer: $\widehat{\beta_0} = -58.398$, $\widehat{\beta_1} = 0.793$, $\widehat{\beta_2} = 1.111$, $\widehat{\beta_3} = -0.397$, $r = 0.971$. The value of the F-statistic for $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$ is 43.410, which is significant compared to the table value for the $F(3, 8)$ statistic, which is 7.59 at 1 percent level of significance.

Section 8

2. Hint: Let

$$\mathbf{U} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}\mathbf{A}^- & \mathbf{I} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{I} & -\mathbf{A}^-\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

and consider $\mathbf{UXV}$.

6. Hint: Use Exercise 2 applied to the Schur complements of $\mathbf{I}$ and $\mathbf{A}$ in the matrix

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} \end{bmatrix}.$$

7. By the spectral theorem we may write $\Sigma = \mathbf{P}\begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{P}'$, where $\mathbf{P}$ is orthogonal and $\mathbf{D}$ is an $r \times r$ diagonal matrix with positive diagonal entries. Set

$\mathbf{B} = \mathbf{P}\begin{bmatrix} \mathbf{D}^{\frac{1}{2}} \\ \mathbf{0} \end{bmatrix}$. Then $\Sigma = \mathbf{B}\mathbf{B}'$, and if $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I_r})$, it can be seen that $\mathbf{y} \sim N(\mathbf{0}, \Sigma)$.

8. Hint: First make the transformation given in the previous exercise and then use **3.1**.

9. Since $\mathbf{A}$ is idempotent, all its eigenvalues are either 0 or 1. Using the spectral theorem we may assume, without loss of generality, that $\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Partition $\mathbf{B}$ conformally as $\begin{bmatrix} \mathbf{B_{11}} & \mathbf{B_{12}} \\ \mathbf{B_{21}} & \mathbf{B_{22}} \end{bmatrix}$. Since $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A} - \mathbf{B}$ are all positive semidefinite, we see that $\mathbf{B_{22}} = \mathbf{0}$. Then since $\mathbf{B}$ is positive semidefinite, it follows that $\mathbf{B_{12}}$, $\mathbf{B_{21}}$ are both zero matrices. Also, $\mathbf{B_{11}}$ is idempotent, and hence so is $\mathbf{I} - \mathbf{B_{11}}$. It follows that $\mathbf{A} - \mathbf{B}$ is idempotent. In terms of distribution of quadratic forms the result says the following: Suppose Q_1, Q_2 are quadratic forms in the vector $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I})$, which are both distributed as chi-square. Then $Q_1 - Q_2$ is also distributed as chi-square if and only if $Q_1 - Q_2$ is nonnegative with probability one.

12. Suppose $R(\mathbf{A}) = r$ and let $\mathbf{A} = \mathbf{BC}$ be a rank factorization, where $\mathbf{B}$, $\mathbf{C}$ are of order $n \times r$, $r \times n$, respectively. Let $\mathbf{B}^-$, $\mathbf{C}^-$ be a left inverse of $\mathbf{B}$ and a right inverse of $\mathbf{C}$, respectively. Let $\mathbf{X} = \frac{1}{2}(\mathbf{I} + \mathbf{CB})$, so that $\mathbf{I} - \mathbf{X} = \frac{1}{2}(\mathbf{I} - \mathbf{CB})$. Then

$$\begin{aligned} R(\mathbf{I_r}) &= R(\mathbf{X} + \mathbf{I} - \mathbf{X}) \\ &\le R(\mathbf{X}) + R(\mathbf{I} - \mathbf{X}) \\ &= R(\mathbf{B}^-(\mathbf{A} + \mathbf{A^2})\mathbf{C}^-) + R(\mathbf{B}^-(\mathbf{A} - \mathbf{A^2})\mathbf{C}^-) \\ &\le R(\mathbf{A} + \mathbf{A^2}) + R(\mathbf{A} - \mathbf{A^2}). \end{aligned}$$

Thus $R(\mathbf{A}) = R(\mathbf{A} + \mathbf{A^2}) + R(\mathbf{A} - \mathbf{A^2})$ holds if and only if $R(\mathbf{I_r}) = R(\mathbf{X}) + R(\mathbf{I} - \mathbf{X})$, and the latter condition is equivalent to $\mathbf{X}$ being idempotent. Now observe that $\mathbf{X}$ is idempotent if and only if $\mathbf{A}$ is tripotent.

15. Answer: Multivariate normal with mean $\mathbf{0}$ and dispersion matrix $\mathbf{I_n} - \mathbf{J_n}$, where $\mathbf{J_n}$ is the $n \times n$ matrix with each entry $\frac{1}{n}$.

21. Answer: The model is $y_i = \theta_i + \epsilon_i, i = 1, 2, 3$, where $\epsilon_i, i = 1, 2, 3$, are i.i.d. $N(0, 1)$ and $\theta_1 + \theta_2 + \theta_3 = \pi$. The test statistic for the hypothesis $\theta_1 = \theta_2 = \theta_3 (= \frac{\pi}{6})$ is $\frac{2(y_1+y_2+y_3)^2}{(y_1+y_2+y_3-\pi)^2}$, distributed as $F(1, 1)$.

22. (i)

$$\mathbf{e}'\widehat{\mathbf{y}} = \mathbf{e}'\begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix}\begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \vdots \\ \widehat{\beta_p} \end{bmatrix}$$

$$
\begin{aligned}
&= \begin{bmatrix} n & \mathbf{e}'\mathbf{X} \end{bmatrix} \begin{bmatrix} n & \mathbf{e}'\mathbf{X} \\ \mathbf{X}'\mathbf{e} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{e}'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix} \\
&= \mathbf{e}'\mathbf{y}.
\end{aligned}
$$

(ii)

$$
\begin{aligned}
(\mathbf{y} - \widehat{\mathbf{y}})'\widehat{\mathbf{y}} &= (\mathbf{y} - \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \widehat{\boldsymbol{\beta}})' \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \widehat{\boldsymbol{\beta}} \\
&= \mathbf{y}' \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}' \begin{bmatrix} \mathbf{e}' \\ \mathbf{X}' \end{bmatrix} \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \widehat{\boldsymbol{\beta}} \\
&= \mathbf{y}' \begin{bmatrix} \mathbf{e} & \mathbf{X} \end{bmatrix} \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}' \begin{bmatrix} \mathbf{e}' \\ \mathbf{X}' \end{bmatrix} \mathbf{y} \\
&= 0.
\end{aligned}
$$

(iii)

$$
\begin{aligned}
\sum_{i=1}^{n} (y_i - \overline{y})^2 &= \sum_{i=1}^{n} (y_i - \widehat{y_i} + \widehat{y_i} - \overline{y})^2 \\
&= \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 + \sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2,
\end{aligned}
$$

since the cross-product term, $(\mathbf{y} - \widehat{\mathbf{y}})'(\widehat{\mathbf{y}} - \overline{y}\mathbf{e})$, equals 0 by (i), (ii).
(iv) It follows from (ii) that $\sum_{i=1}^{n} (y_i - \overline{y})(\widehat{y_i} - \overline{y}) = \sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2$. Thus, since $\overline{y} = \overline{\widehat{y}}$ by (i), we have

$$
r^2 = \frac{\sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2},
$$

and hence, by (iii),

$$
r^2 = \frac{\sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2 + \sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2}.
$$

It follows that

$$
\frac{r^2}{1 - r^2} = \frac{\sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2},
$$

which equals

$$
\frac{\sum_{i=1}^{n} (y_i - \overline{y})^2 - \sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2},
$$

and the result is proved, in view of (25).

4

Singular Values and Their Applications

4.1 Singular Value Decomposition

Let $\mathbf{A}$ be an $n \times n$ matrix. The *singular values* of $\mathbf{A}$ are defined to be the eigenvalues of $(\mathbf{AA}')^{\frac{1}{2}}$. Since $\mathbf{AA}'$ is positive semidefinite, the singular values are nonnegative, and we denote them by

$$\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_n(\mathbf{A}).$$

If there is no possibility of confusion, then we will denote the singular values of $\mathbf{A}$ simply by $\sigma_1 \geq \cdots \geq \sigma_n$.

The singular values of a rectangular matrix are defined as follows. Suppose $\mathbf{A}$ is $m \times n$ with $m < n$. Augment $\mathbf{A}$ by $n - m$ zero rows to get a square matrix, say $\mathbf{B}$. Then the singular values of $\mathbf{A}$ are defined to be the singular values of $\mathbf{B}$. If $m > n$, then a similar definition can be given by augmenting $\mathbf{A}$ by zero columns, instead of rows. For convenience we will limit our discussion mostly to singular values of *square* matrices. The general case can be handled by making minor modifications.

The following assertions are easily verified. We omit the proof:

(i) The singular values of $\mathbf{A}$ and $\mathbf{PAQ}$ are identical for any orthogonal matrices $\mathbf{P}$, $\mathbf{Q}$.

(ii) The rank of a matrix equals the number of nonzero singular values of the matrix.

(iii) If $\mathbf{A}$ is symmetric, then the singular values of $\mathbf{A}$ are the absolute values of its eigenvalues. If $\mathbf{A}$ is positive semidefinite, then the singular values are the same as the eigenvalues.

1.1 (The Singular Value Decomposition). *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then there exist orthogonal matrices* $\mathbf{P}$, $\mathbf{Q}$ *such that*

$$\mathbf{PAQ} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix}. \tag{1}$$

Proof. Let $\mathbf{x}$ be an eigenvector of $\mathbf{AA}'$ corresponding to the eigenvalue σ_1^2, such that $\|\mathbf{x}\| = 1$. Thus

$$\mathbf{AA}'\mathbf{x} = \sigma_1^2\mathbf{x},$$

and hence $\mathbf{x}'\mathbf{AA}'\mathbf{x} = \sigma_1^2$. If we set

$$\mathbf{y} = \frac{1}{\|\mathbf{A}'\mathbf{x}\|}\mathbf{A}'\mathbf{x},$$

then it follows that $\mathbf{x}'\mathbf{Ay} = \sigma_1$. We can construct orthogonal matrices $\mathbf{U}$, $\mathbf{V}$ such that the first row of $\mathbf{U}$ is $\mathbf{x}'$ and the first column of $\mathbf{V}$ is $\mathbf{y}$. Then

$$\mathbf{UAV} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{bmatrix}.$$

Now we use induction as in the proof of the spectral theorem (**8.7** of Chapter 1) and get the result. □

1.2 Remark

In (1), let $\mathbf{y_1}, \ldots, \mathbf{y_n}$ denote the columns of $\mathbf{Q}$ and let $\mathbf{x_1}, \ldots, \mathbf{x_n}$ denote the columns of $\mathbf{P}'$. Then $\mathbf{y_i}$ is an eigenvector of $\mathbf{A}'\mathbf{A}$ and $\mathbf{x_i}$ is an eigenvector of $\mathbf{AA}'$ corresponding to the same eigenvalue. These vectors are called the *singular vectors* of $\mathbf{A}$.

Problems

1. Find the singular values and the singular vectors of the following matrices:

$$\begin{bmatrix} 2 & 0 & 0 \\ -1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

2. Let $\mathbf{A}$ be an $n \times n$ matrix with $\frac{n^2}{2}$ entries equal to 2 and the remaining entries equal to 4. Show that the sum of the squares of the singular values of $\mathbf{A}$ is $10n^2$.

3. For $1 \le i < j \le n$, let e^{ij} denote the $n \times 1$ column vector with 1, -1 at the i, j coordinates, respectively, and zeros elsewhere. Let $\mathbf{A}$ be the matrix with n

rows and with $\frac{n(n-1)}{2}$ columns given by e^{ij}, $1 \le i < j \le n$. Show that the only nonzero singular value of $\mathbf{A}$ is $\sqrt{n}$ with multiplicity $n-1$.

4.2 Extremal Representations

2.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then*

$$\max_{\|\mathbf{u}\|=1, \|\mathbf{v}\|=1} |\mathbf{u}'\mathbf{A}\mathbf{v}| = \sigma_1.$$

PROOF. We make use of (1). For any $\mathbf{u}$, $\mathbf{v}$ of norm 1,

$$\begin{aligned}
|\mathbf{u}'\mathbf{A}\mathbf{v}| &= |\mathbf{u}'\mathbf{P}'\mathrm{diag}(\sigma_1, \dots, \sigma_n)\mathbf{Q}'\mathbf{v}| \\
&= |\mathbf{w}'\mathrm{diag}(\sigma_1, \dots, \sigma_n)\mathbf{z}| \qquad (\mathbf{w} = \mathbf{P}\mathbf{u}, \mathbf{z} = \mathbf{Q}'\mathbf{v}) \\
&= |\sigma_1 w_1 z_1 + \cdots + \sigma_n w_n z_n| \\
&\le \sigma_1(|w_1 z_1| + \cdots + |w_n z_n|) \\
&\le \sigma_1 \|\mathbf{w}\| \|\mathbf{z}\| \quad \text{(by the Cauchy–Schwarz inequality)} \\
&= \sigma_1.
\end{aligned} \tag{2}$$

Therefore,

$$\max_{\|\mathbf{u}\|=1, \|\mathbf{v}\|=1} |\mathbf{u}'\mathbf{A}\mathbf{v}| \le \sigma_1. \tag{3}$$

Also, by **1.1**, $|\mathbf{x_1}'\mathbf{A}\mathbf{y_1}| = \sigma_1$, and hence equality is attained in (3). □

2.2. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then for* $2 \le k \le n$,

$$\max |\mathbf{u}'\mathbf{A}\mathbf{v}| = \sigma_k,$$

where the maximum is taken over the set

$$\{\mathbf{u}, \mathbf{v} : \|\mathbf{u}\| = \|\mathbf{v}\| = 1, \mathbf{u} \perp \mathbf{x_i}, \mathbf{v} \perp \mathbf{y_i};\ i = 1, \dots, k-1\}.$$

(Here $\mathbf{x_i}, \mathbf{y_i}$ *are the same as in Remark* **1.2**.*)*

PROOF. The proof proceeds along similar lines to that of **2.1**. Only observe that in (2), the first $k-1$ terms reduce to zero. □

If **2.1, 2.2** are applied to a positive semidefinite matrix, then we get a representation for the eigenvalues. This is known as Rayleigh's quotient expression and is given in the next result.

2.3. *Let* $\mathbf{A}$ *be a positive semidefinite* $n \times n$ *matrix. Then*

(i) $\max\limits_{\|\mathbf{u}\|=1} \mathbf{u}'\mathbf{A}\mathbf{u} = \lambda_1.$

(ii) *Let* $(\mathbf{x_1}, \dots, \mathbf{x_n})$ *be a set of orthonormal eigenvectors of* $\mathbf{A}$ *corresponding to* $\lambda_1, \dots, \lambda_n$ *respectively. Then*

$$\max \mathbf{u}'\mathbf{A}\mathbf{u} = \lambda_k, \quad k = 2, \dots, n,$$

where the maximum is taken over the set

$$\{\mathbf{u} : \|\mathbf{u}\| = 1,\ \mathbf{u} \perp \mathbf{x_1}, \ldots, \mathbf{x_{k-1}}\}.$$

PROOF. Since $\mathbf{A}$ is positive semidefinite,

$$|\mathbf{u}'\mathbf{A}\mathbf{v}| \leq (\mathbf{u}'\mathbf{A}\mathbf{u})^{\frac{1}{2}}(\mathbf{v}'\mathbf{A}\mathbf{v})^{\frac{1}{2}}.$$

This fact and **2.1, 2.2** give the result. □

Note that **2.3** holds for any symmetric matrix as well. This can be seen as follows. If $\mathbf{A}$ is symmetric, then we may choose $\delta > 0$ sufficiently large so that $\mathbf{A} + \delta\mathbf{I}$ is positive semidefinite and then apply **2.3**.

The next result is known as the Courant–Fischer minimax theorem. It is an important result with several interesting consequences.

2.4. *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then for* $k = 2, \ldots, n$,

$$\lambda_k(\mathbf{A}) = \min_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \max_{\|\mathbf{u}\|=1,\mathbf{u}\perp\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \mathbf{u}'\mathbf{A}\mathbf{u}.$$

PROOF. It is sufficient to show that

$$\lambda_k(\mathbf{A}) \leq \min_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \max_{\|\mathbf{u}\|=1,\mathbf{u}\perp\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \mathbf{u}'\mathbf{A}\mathbf{u},$$

since by **2.3**, equality is attained when $\mathbf{w_i} = \mathbf{x_i}$. Let $\mathbf{P}$ be orthogonal with $\mathbf{PAP}' = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$\mathbf{u}'\mathbf{A}\mathbf{u} = \sum_{i=1}^{n} \lambda_i z_i^2,$$

where $\mathbf{z} = \mathbf{P}\mathbf{u}$.

Consider the vector spaces

$$T_1 = \{\mathbf{z} \in R^n : \mathbf{z} \perp \mathbf{P}\mathbf{w_1}, \ldots, \mathbf{P}\mathbf{w_{k-1}}\}$$

and

$$T_2 = \{\mathbf{z} \in R^n : z_{k+1} = \cdots = z_n = 0\}.$$

Then $\dim(T_1) \geq n - k + 1$ and $\dim(T_2) = k$. Thus (see Exercise 28, Chapter 1) there exists a vector $\mathbf{z}$ of norm 1 in $T_1 \cap T_2$. For this $\mathbf{z}$,

$$\sum_{i=1}^{n} \lambda_i z_i^2 = \sum_{i=1}^{k} \lambda_i z_i^2 \geq \lambda_k.$$

Thus for any $\mathbf{w_1}, \ldots, \mathbf{w_{k-1}}$,

$$\max_{\|\mathbf{u}\|=1,\mathbf{u}\perp\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \mathbf{u}'\mathbf{A}\mathbf{u} = \max_{\|\mathbf{z}\|=1,\mathbf{z}\perp\mathbf{P}\mathbf{w_1},\ldots,\mathbf{P}\mathbf{w_{k-1}}} \sum_{i=1}^{n} \lambda_i z_i^2 \geq \lambda_k,$$

and the proof is complete. □

2.5 (Cauchy Interlacing Principle). *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix and let* $\mathbf{B}$ *be a principal submatrix of* $\mathbf{A}$ *of order* $(n-1) \times (n-1)$. *Then*

$$\lambda_k(\mathbf{A}) \geq \lambda_k(\mathbf{B}) \geq \lambda_{k+1}(\mathbf{A}), \qquad k = 1, \ldots, n-1.$$

PROOF. We assume, without loss of generality, that $\mathbf{B}$ is obtained by deleting the first row and column of $\mathbf{A}$. By **2.3**,

$$\lambda_1(\mathbf{A}) = \max_{\|\mathbf{u}\|=1} \mathbf{u}'\mathbf{A}\mathbf{u} \geq \max_{\|\mathbf{u}\|=1, u_1=0} \mathbf{u}'\mathbf{A}\mathbf{u} = \lambda_1(\mathbf{B}).$$

Similarly, for $k = 2, \ldots, n-1$, by **2.4**,

$$\begin{aligned}
\lambda_k(\mathbf{A}) &= \min_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \max_{\|\mathbf{u}\|=1, \mathbf{u}\perp\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \mathbf{u}'\mathbf{A}\mathbf{u} \\
&\geq \min_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \max_{\|\mathbf{u}\|=1, \mathbf{u}\perp\mathbf{w_1},\ldots,\mathbf{w_{k-1}}, u_1=0} \mathbf{u}'\mathbf{A}\mathbf{u} \\
&= \lambda_k(\mathbf{B}).
\end{aligned}$$

Now

$$\begin{aligned}
\lambda_k(\mathbf{B}) &= -\lambda_{n-k}(-\mathbf{B}) \\
&\geq -\lambda_{n-k}(-\mathbf{A}) \quad \text{(by the first part)} \\
&= \lambda_{k+1}(\mathbf{A}),
\end{aligned}$$

$k = 2, \ldots, n-1$. □

2.6. *Let* $\mathbf{A}, \mathbf{B}$ *be* $n \times n$ *matrices such that* $\mathbf{A} \geq \mathbf{B}$. *Then*

$$\lambda_k(\mathbf{A}) \geq \lambda_k(\mathbf{B}), \qquad k = 1, \ldots, n.$$

PROOF. It is possible to derive the result using **2.4**. However, we give another proof. Let $\mathbf{u_i}$, $\mathbf{v_i}$ be sets of orthonormal eigenvectors of $\mathbf{A}, \mathbf{B}$ corresponding to $\lambda_i(\mathbf{A})$, $\lambda_i(\mathbf{B})$, respectively. Fix $k \in \{1, \ldots, n\}$ and let

$$T_1 = \operatorname{span}\{\mathbf{u_k}, \ldots, \mathbf{u_n}\}$$

and

$$T_2 = \operatorname{span}\{\mathbf{v_1}, \ldots, \mathbf{v_k}\}.$$

Then $\dim(T_1) = n - k + 1$, $\dim(T_2) = k$, and therefore there exists a unit vector $\mathbf{z} \in T_1 \cap T_2$. Consider $\mathbf{z}'\mathbf{A}\mathbf{z} = \mathbf{z}'(\lambda_1(\mathbf{A})\mathbf{u_1}\mathbf{u_1'} + \cdots + \lambda_n(\mathbf{A})\mathbf{u_n}\mathbf{u_n'})\mathbf{z}$. Since $\mathbf{z} \in T_1$, it is orthogonal to $\{\mathbf{u_1}, \ldots, \mathbf{u_{k-1}}\}$, and hence we have

$$\mathbf{z}'\mathbf{A}\mathbf{z} = \mathbf{z}'(\lambda_k(\mathbf{A})\mathbf{u_k}\mathbf{u_k'} + \cdots + \lambda_n(\mathbf{A})\mathbf{u_n}\mathbf{u_n'})\mathbf{z}.$$

Now, using the fact that $\lambda_k(\mathbf{A}) \geq \lambda_i(\mathbf{A})$, $i \geq k$, and that

$$\mathbf{z}'(\mathbf{u_k}\mathbf{u_k'} + \cdots + \mathbf{u_n}\mathbf{u_n'})\mathbf{z} \leq \mathbf{z}'(\mathbf{u_1}\mathbf{u_1'} + \cdots + \mathbf{u_n}\mathbf{u_n'})\mathbf{z} = 1,$$

we get $\mathbf{z}'\mathbf{A}\mathbf{z} \leq \lambda_k(\mathbf{A})$. A similar argument, using $\mathbf{z} \in T_2$, gives

$$\begin{aligned}
\mathbf{z}'\mathbf{B}\mathbf{z} &= \mathbf{z}'(\lambda_1(\mathbf{B})\mathbf{v_1}\mathbf{v_1'} + \cdots + \lambda_n(\mathbf{B})\mathbf{v_n}\mathbf{v_n'})\mathbf{z} \\
&= \mathbf{z}'(\lambda_1(\mathbf{B})\mathbf{v_1}\mathbf{v_1'} + \cdots + \lambda_k(\mathbf{B})\mathbf{v_k}\mathbf{v_k'})\mathbf{z} \\
&\geq \lambda_k(\mathbf{B}).
\end{aligned}$$

The result now follows, since for any $\mathbf{z}$, $\mathbf{z}'\mathbf{A}\mathbf{z} \geq \mathbf{z}'\mathbf{B}\mathbf{z}$. □

Problems

1. Let $\mathbf{A}$ be an $n \times n$ matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_n$. Show that

$$\max_{i,j} |a_{ij}| \leq \sigma_1.$$

2. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ positive semidefinite matrices such that $\lambda_k(\mathbf{A}) \geq \lambda_k(\mathbf{B})$, $k = 1, 2, \ldots, n$. Does it follow that $\mathbf{A} \geq \mathbf{B}$?

3. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ such that only λ_n is negative. Suppose all the diagonal entries of $\mathbf{A}$ are negative. Show that any principal submatrix of $\mathbf{A}$ also has exactly one negative eigenvalue.

4.3 Majorization

If $\mathbf{x} \in R^n$, then we denote by

$$x_{[1]} \geq \cdots \geq x_{[n]}$$

the components of $\mathbf{x}$ arranged in nonincreasing order. If $\mathbf{x}, \mathbf{y} \in R^n$, then $\mathbf{x}$ is said to be *majorized* by $\mathbf{y}$, denoted $\mathbf{x} \prec \mathbf{y}$, if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \quad k = 1, \ldots, n-1,$$

and

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i.$$

Intuitively, $\mathbf{x} \prec \mathbf{y}$ if the components of $\mathbf{y}$ are more "spread out" than the components of $\mathbf{x}$. The concept finds applications in several areas in mathematics, statistics, and economics.

3.1. *Let $\mathbf{A} = (a_{ij})$ be a symmetric $n \times n$ matrix. Then*

$$(a_{11}, \ldots, a_{nn}) \prec (\lambda_1, \ldots \lambda_n).$$

Proof. The result is obvious for $n = 1$. Assume the result to be true for matrices of order $n - 1$ and proceed by induction. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. By a permutation of rows and an identical permutation of columns we can arrange the diagonal entries in nonincreasing order, and this operation does not change the eigenvalues. Therefore, we assume, without loss of generality, that

$$a_{11} \geq \cdots \geq a_{nn}.$$

Let $\mathbf{B}$ be the submatrix obtained by deleting the last row and column of $\mathbf{A}$. Then for $k = 1, \ldots, n-1$,

$$\sum_{i=1}^{k} \lambda_i(\mathbf{A}) \geq \sum_{i=1}^{k} \lambda_i(\mathbf{B}) \geq \sum_{i=1}^{k} a_{ii},$$

by **2.5** and the induction assumption. Also,

$$\sum_{i=1}^{n} \lambda_i(\mathbf{A}) = \text{trace}\mathbf{A} = \sum_{i=1}^{n} a_{ii},$$

and the proof is complete. □

3.2. *Let* **A** *be a symmetric* $n \times n$ *matrix. Then for* $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} \lambda_i = \max \ \text{trace } \mathbf{R}'\mathbf{A}\mathbf{R},$$

where the maximum is taken over all $n \times k$ *matrices* **R** *with orthonormal columns.*

PROOF. Let **R** be an $n \times k$ matrix with orthonormal columns. Complete **R** into an $n \times n$ orthogonal matrix **P**. Then trace$\mathbf{R}'\mathbf{A}\mathbf{R}$ is the sum of the first k diagonal entries of $\mathbf{P}'\mathbf{A}\mathbf{P}$. Since $\mathbf{P}'\mathbf{A}\mathbf{P}$ has the same eigenvalues as **A**, it follows by **3.1** that

$$\sum_{i=1}^{k} \lambda_i \geq \text{trace}\mathbf{R}'\mathbf{A}\mathbf{R}. \tag{4}$$

If the columns of **R** form an orthonormal set of eigenvectors corresponding to $\lambda_1, \ldots, \lambda_k$, then equality is obtained in (4), and the result is proved. □

3.3. *Let* **A**, **B** *be symmetric* $n \times n$ *matrices. Then*

$$(\lambda_1(\mathbf{A}+\mathbf{B}), \ldots, \lambda_n(\mathbf{A}+\mathbf{B}))$$

is majorized by

$$(\lambda_1(\mathbf{A}) + \lambda_1(\mathbf{B}), \ldots, \lambda_n(\mathbf{A}) + \lambda_n(\mathbf{B})).$$

PROOF. We use **3.2**. The maximum in the following argument is over $n \times k$ matrices **R** with orthonormal columns. We have

$$\begin{aligned}
\sum_{i=1}^{k} \lambda_i(\mathbf{A}+\mathbf{B}) &= \max \text{trace}\mathbf{R}'(\mathbf{A}+\mathbf{B})\mathbf{R} \\
&\leq \max \text{trace}\mathbf{R}'\mathbf{A}\mathbf{R} + \max \text{trace}\mathbf{R}'\mathbf{B}\mathbf{R} \\
&= \sum_{i=1}^{k} \lambda_i(\mathbf{A}) + \sum_{i=1}^{k} \lambda_i(\mathbf{B}),
\end{aligned}$$

and the proof is complete. □

3.4. *Let* **A** *be a symmetric* $n \times n$ *matrix and let* $\mathbf{P_1}, \ldots, \mathbf{P_m}$ *be* $n \times n$ *orthogonal matrices. Then the eigenvalues of*

$$\frac{1}{m} \sum_{i=1}^{m} \mathbf{P_i}'\mathbf{A}\mathbf{P_i}$$

are majorized by the eigenvalues of **A**.

PROOF. The result follows by a simple application of **3.3**. □

Problems

1. Show that the vector $(x_1, \ldots, x_p)$ majorizes the vector $(\overline{x}, \ldots, \overline{x})$, where $\overline{x}$ appears p times.

2. There are n players participating in a chess tournament. Each player plays a game against every other player. The winner gets 1 point and the loser gets zero. In case of a draw each player is awarded half a point. Let s_i denote the score of player i, $i = 1, 2, \ldots, n$. Show that the vector $(s_1, s_2, \ldots, s_n)$ is majorized by $(0, 1, \ldots, n-1)$.

3. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. By the spectral theorem there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{PDP}'$, where $\mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Let $\mathbf{S} = (s_{ij})$ be the $n \times n$ matrix where $s_{ij} = p_{ij}^2$, $i, j = 1, 2, \ldots, n$. Show that

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{nn} \end{bmatrix} = \mathbf{S} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}.$$

Also show that $\mathbf{S}$ has row and column sums equal to one. (Thus $\mathbf{S}$ is *doubly stochastic*, and it follows by a well-known result of Hardy, Littlewood, and Pólya that

$$(a_{11}, \ldots, a_{nn}) \prec (\lambda_1, \ldots \lambda_n),$$

leading to another proof of **3.1**.)

4.4 Principal Components

Suppose in a statistical survey we collect observations on a large number of random variables, $x_1, \ldots, x_n$, and we want to study the variability in the data. It is desirable to reduce the number of variables to a few variables that "explain" the total variation. Suppose the new variables are $y_1, \ldots, y_p$, $p \le n$; where each y_i is a function of $x_1, \ldots, x_n$. For mathematical simplicity it is convenient to let $y_1, \ldots, y_p$ be linear functions of $x_1, \ldots, x_n$.

The first *principal component* is defined to be the linear combination

$$\alpha_1 x_1 + \cdots + \alpha_n x_n, \quad \|\boldsymbol{\alpha}\| = 1,$$

with maximum variance.

Let Σ be the dispersion matrix of $\mathbf{x} = (x_1, \ldots, x_n)'$, which we assume to be positive definite. Then

$$\operatorname{var}(\boldsymbol{\alpha}'\mathbf{x}) = \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}.$$

Thus in order to find the first principal component we must maximize $\boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}$ subject to $\|\boldsymbol{\alpha}\| = 1$. By **6.1** of Chapter 3 this maximum is $\lambda_1(\Sigma)$ and is attained at a unit eigenvector corresponding to $\lambda_1(\Sigma)$.

Let $\mathbf{v_1}, \dots, \mathbf{v_n}$ be an orthonormal set of eigenvectors corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$ of Σ, respectively. The second principal component is a linear combination

$$\beta_1 x_1 + \cdots + \beta_n x_n, \quad \|\boldsymbol{\beta}\| = 1,$$

which is uncorrelated with the first principal component and has maximum variance. In order to find the second principal component we must maximize $\boldsymbol{\beta}'\Sigma\boldsymbol{\beta}$ subject to $\|\boldsymbol{\beta}\| = 1$ and $\boldsymbol{\beta}'\Sigma\boldsymbol{\alpha} = 0$, i.e., $\boldsymbol{\beta}'\mathbf{v_1} = 0$. By **2.3** this maximum is $\lambda_2(\Sigma)$, attained at $\beta = \mathbf{v_2}$.

In general, we may define the kth principal component as a linear combination

$$\gamma_1 x_1 + \cdots + \gamma_n x_n, \quad \|\boldsymbol{\gamma}\| = 1,$$

which is uncorrelated with the first $k-1$ principal components and which has maximum variance. By a similar analysis as before, the kth principal component is obtained when $\boldsymbol{\gamma} = \mathbf{v_k}$, and its variance is λ_k.

The sum of the variances of the principal components is $\sum_{i=1}^n \lambda_i$, which is the same as $\sum_{i=1}^n \operatorname{var}(x_i)$. The proportion of the total variability explained by the first k principal components can be defined to be

$$\frac{\sum\limits_{i=1}^{k} \lambda_i}{\sum\limits_{i=1}^{n} \lambda_i}.$$

Problems

1. Suppose the variance–covariance matrix of a sample from a bivariate distribution is

$$\begin{bmatrix} 60 & 20 \\ 20 & 60 \end{bmatrix}.$$

Find the two principal components and their associated variances.

2. The following variance–covariance matrix was computed from a sample of 50 observations, each consisting of measurements on four characteristics. What percentage of the total variability is accounted for by the last principal component?

$$\begin{bmatrix} 10 & -3 & 2 & 1 \\ -3 & 8 & -1 & 4 \\ 2 & -1 & 10 & -2 \\ 1 & 4 & -2 & 12 \end{bmatrix}.$$

4.5 Canonical Correlations

Consider a situation where we have two sets of variables, $x_1, \dots, x_p$ and $y_1, \dots, y_{n-p}$, and we want to study the correlation structure between the two sets. As an example, we might have observations on a group of students. The first set of variables may correspond to "physical" variables such as height and weight, whereas the second set may correspond to "mental" characteristics such as scores on various tests.

Let Σ be the dispersion matrix of

$$(x_1, \dots, x_p, y_1, \dots, y_{n-p})'.$$

We assume Σ to be positive definite and suppose it is partitioned as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where Σ_{11} is $p \times p$.

The first pair of *canonical variates* is a pair of linear combinations $\boldsymbol{\alpha}'\mathbf{x}$, $\boldsymbol{\beta}'\mathbf{y}$ with unit variance such that the correlation between them is maximum. The correlation is called the first *canonical correlation*. We have

$$\operatorname{cov}(\boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\beta}'\mathbf{y}) = \boldsymbol{\alpha}'\Sigma_{12}\boldsymbol{\beta},$$

and this is to be maximized subject to the conditions $\boldsymbol{\alpha}'\Sigma_{11}\boldsymbol{\alpha} = \boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta} = 1$. Let

$$\mathbf{u} = \Sigma_{11}^{1/2}\boldsymbol{\alpha}, \mathbf{v} = \Sigma_{22}^{1/2}\boldsymbol{\beta}.$$

Then the problem is to find

$$\max_{\|\mathbf{u}\|=1, \|\mathbf{v}\|=1} \mathbf{u}'A\mathbf{v},$$

and by **2.1** this is $\sigma_1(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2})$. The maximum is attained when $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are the eigenvectors of

$$\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2} \quad \text{and} \quad \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2},$$

respectively, corresponding to $\sigma_1^2(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2})$.

Let $\mathbf{v}_1'\mathbf{x}$, $\mathbf{w}_1'\mathbf{y}$ denote the first pair of canonical variates. The second pair of canonical variates is defined to be a pair of linear combinations $\boldsymbol{\gamma}'\mathbf{x}$, $\boldsymbol{\delta}'\mathbf{y}$ such that they have unit variance, and have a maximum correlation subject to the condition that $\boldsymbol{\gamma}'\mathbf{x}$ is uncorrelated with $\mathbf{v}_1'\mathbf{x}$ and $\boldsymbol{\delta}'\mathbf{y}$ is uncorrelated with $\mathbf{w}_1'\mathbf{y}$. This maximum correlation is called the second canonical correlation. Further canonical variates are defined similarly. By **2.2** it can be seen that the canonical correlations correspond to the singular values of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$.

Problems

1. The following variance–covariance matrix was computed from a sample of 50 observations, each consisting of measurements on four characteristics. Find the canonical correlations and canonical variates between the first two variables and the last two variables.

$$\begin{bmatrix} 6 & 2 & 2 & 7 \\ 2 & 30 & -3 & 20 \\ 2 & -3 & 12 & -2 \\ 7 & 20 & -2 & 22 \end{bmatrix}.$$

2. We follow the notation of Section 5 here. Suppose $\sigma_1, \dots, \sigma_p$ are the canonical correlations. Show that $\sigma_1^2, \dots, \sigma_p^2$ are the nonzero roots of the equation

$$\begin{vmatrix} -\lambda \boldsymbol{\Sigma}_{\mathbf{11}} & \boldsymbol{\Sigma}_{\mathbf{12}} \\ \boldsymbol{\Sigma}_{\mathbf{21}} & -\lambda \boldsymbol{\Sigma}_{\mathbf{22}} \end{vmatrix} = 0.$$

4.6 Volume of a Matrix

Let $\mathbf{A}$ be an $m \times n$ matrix and let $1 \le k \le \min\{m, n\}$. We denote by $Q_{k,n}$ the set of increasing sequences of k elements from $\{1, 2, \dots, n\}$. For indices $I \subset \{1, 2, \dots, m\}$, $J \subset \{1, 2, \dots, n\}$, $\mathbf{A_{IJ}}$ will denote the corresponding submatrix of $\mathbf{A}$. The kth *compound* of $\mathbf{A}$, denoted by $C_k(\mathbf{A})$, is an $\binom{m}{k} \times \binom{n}{k}$ matrix defined as follows. The rows and the columns of $C_k(\mathbf{A})$ are indexed by $Q_{k,m}$, $Q_{k,n}$, respectively, where the ordering is arbitrary but fixed. If $I \in Q_{k,m}$, $J \in Q_{k,n}$, then the (I, J)-entry of $C_k(\mathbf{A})$ is set to be $|\mathbf{A_{IJ}}|$.

The next result generalizes the familiar fact that for $n \times n$ matrices $\mathbf{A}$, $\mathbf{B}$, $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$.

6.1 (Cauchy–Binet Formula)**.** *Let* $\mathbf{A}$, $\mathbf{B}$ *be matrices of order* $m \times n$, $n \times m$, *respectively, where* $m \le n$. *Let* $S = \{1, 2, \dots, m\}$. *Then*

$$|\mathbf{AB}| = \sum_{I \in Q_{m,n}} |\mathbf{A_{SI}}||\mathbf{B_{IS}}|. \tag{5}$$

Proof. We only sketch a proof. By elementary row operations it can be verified that

$$\begin{vmatrix} -\mathbf{I_n} & \mathbf{B} \\ \mathbf{A} & \mathbf{0} \end{vmatrix} = \begin{vmatrix} -\mathbf{I_n} & \mathbf{B} \\ \mathbf{0} & \mathbf{AB} \end{vmatrix}.$$

The determinant on the right-hand side in the equation above is clearly $(-1)^n|\mathbf{AB}|$. It can be seen, by expanding along the first n columns, that the determinant on the left equals

$$(-1)^n \sum_{I \in Q_{m,n}} |\mathbf{A_{SI}}||\mathbf{B_{IS}}|,$$

and the result is proved. □

The next result follows immediately from **6.1**.

6.2. *Let* $\mathbf{A}, \mathbf{B}$ *be matrices of order* $m \times n$, $n \times p$, *respectively, and let* $1 \leq k \leq \min\{m, n, p\}$. *Then* $C_k(\mathbf{AB}) = C_k(\mathbf{A})C_k(\mathbf{B})$.

If $\mathbf{A}$ is an $n \times n$ matrix, then as usual, $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_n(\mathbf{A})$ will denote the singular values of $\mathbf{A}$. We will often write σ_i instead of $\sigma_i(\mathbf{A})$.

6.3. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix and let* $1 \leq k \leq n$. *Then the singular values of* $C_k(\mathbf{A})$ *are given by*

$$\left\{ \prod_{i \in I} \sigma_i(\mathbf{A}) : I \in Q_{k,n} \right\}. \tag{6}$$

PROOF. Let $\mathbf{P}, \mathbf{Q}$ be orthogonal matrices such that (see **1.1**)

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix} \mathbf{Q}.$$

Then by **6.2**,

$$C_k(\mathbf{A}) = C_k(\mathbf{P}) C_k \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix} C_k(\mathbf{Q}).$$

Since $\mathbf{P}, \mathbf{Q}$ are orthogonal, it follows by **6.2** that $C_k(\mathbf{P}), C_k(\mathbf{Q})$ are also orthogonal. Thus the singular values of $C_k(\mathbf{A})$ are given by the diagonal elements of $C_k(\operatorname{diag}(\sigma_1, \ldots, \sigma_n))$, which are precisely as in (6). □

Let $\mathbf{A}$ be an $n \times n$ matrix. The *k-volume* of $\mathbf{A}$, denoted by $\operatorname{vol}_k(\mathbf{A})$, is defined as

$$\operatorname{vol}_k(\mathbf{A}) = \left\{ \sum_{I, J \in Q_{k,n}} |\mathbf{A}_{\mathbf{IJ}}|^2| \right\}^{\frac{1}{2}} = \{\text{trace } C_k(\mathbf{AA}')\}^{\frac{1}{2}}.$$

Note that $\operatorname{vol}_k(\mathbf{A}) = 0$ if $k > R(\mathbf{A})$. Also, it follows from **6.3** that

$$\operatorname{vol}_k(\mathbf{A}) = \left\{ \sum_{I \in Q_{k,n}} \prod_{i \in I} \sigma_i^2 \right\}^{\frac{1}{2}}.$$

In particular, if $k = r = R(\mathbf{A})$, then

$$\operatorname{vol}_r(\mathbf{A}) = \sigma_1 \cdots \sigma_r.$$

We call the r-volume of $\mathbf{A}$ simply the *volume* of $\mathbf{A}$ and denote it by $\operatorname{vol}(\mathbf{A})$.

The term "volume" can be justified as follows. There is a close connection between determinant and (geometrical) volume. Suppose $\mathbf{A}$ is an $n \times n$ matrix and let $\mathcal{C}$ be the unit cube in R^n. Consider the linear transformation $\mathbf{x} \mapsto \mathbf{Ax}$ from R^n

to R^n and let $\mathbf{A}(\mathcal{C})$ be the image of $\mathcal{C}$ under this transformation. Then it turns out that the volume of $\mathbf{A}(\mathcal{C})$ is precisely $|\mathbf{A}|$. More generally, suppose $\mathbf{A}$ is an $n \times n$ matrix of rank r and let $\mathcal{C}$ be the unit cube in the column space of $\mathbf{A}'$. Then the volume of the image of $\mathcal{C}$ under the linear transformation $\mathbf{x} \mapsto \mathbf{Ax}$ is $\mathrm{vol}(\mathbf{A})$.

We now obtain some chracterizations of the Moore–Penrose inverse in terms of volume. For example, we will show that if $\mathbf{A}$ is an $n \times n$ matrix, then $\mathbf{A}^+$ is a g-inverse of $\mathbf{A}$ with minimum volume. First we prove some preliminary results. It is easily seen (see Exercise 1) that $\mathbf{A}^+$ can be determined from the singular value decomposition of $\mathbf{A}$. A more general result is proved next.

6.4. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix of rank* r *and let*

$$\mathbf{A} = \mathbf{P}\begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}$$

be the singular value decomposition of $\mathbf{A}$, *where* $\mathbf{P}, \mathbf{Q}$ *are orthogonal and* $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$. *Then the class of g-inverses of* $\mathbf{A}$ *is given by*

$$\mathbf{G} = \mathbf{Q}'\begin{bmatrix} \Sigma^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}\mathbf{P}', \tag{7}$$

where $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *are arbitrary matrices of appropriate dimension. The class of reflexive g-inverses* $\mathbf{G}$ *of* $\mathbf{A}$ *is given by (7) with the additional condition that* $\mathbf{Z} = \mathbf{Y}\Sigma\mathbf{X}$. *The class of least squares g-inverses* $\mathbf{G}$ *of* $\mathbf{A}$ *is given by (7) with* $\mathbf{X} = \mathbf{0}$. *The class of minimum norm g-inverses* $\mathbf{G}$ *of* $\mathbf{A}$ *is given by (7) with* $\mathbf{Y} = \mathbf{0}$. *Finally, the Moore–Penrose inverse of* $\mathbf{A}$ *is given by (7) with* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *all being zero.*

Proof. Let $\mathbf{B} = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and suppose $\mathbf{H} = \begin{bmatrix} \mathbf{U} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}$ is a g-inverse of $\mathbf{B}$. Then $\mathbf{BHB} = \mathbf{B}$ leads to

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

Thus with $\mathbf{U} = \Sigma^{-1}$ and with arbitrary $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, it follows that $\mathbf{H}$ is a g-inverse of $\mathbf{B}$. Since the class of g-inverses of $\mathbf{A}$ is given by

$$\left\{ \mathbf{Q}'\begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^{-}\mathbf{P}' \right\},$$

we have proved the first assertion. The remaining assertions are proved easily. For example, imposing both the conditions $\mathbf{BHB} = \mathbf{B}$ and $\mathbf{HBH} = \mathbf{H}$, we see that $\mathbf{U} = \Sigma^{-1}$ and $\mathbf{Z} = \mathbf{Y}\Sigma\mathbf{X}$. That completes the proof. □

We now show that the Moore–Penrose inverse enjoys a certain minimality property with respect to the singular values in the class of all g-inverses of a given matrix.

6.5. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix of rank* r *with singular values*

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0.$$

Let $\mathbf{G}$ *be a g-inverse of* $\mathbf{A}$ *with rank* s *and with singular values*

$$\sigma_1(\mathbf{G}) \geq \cdots \geq \sigma_s(\mathbf{G}) > \sigma_{s+1}(\mathbf{G}) = \cdots = \sigma_n(\mathbf{G}) = 0.$$

Then

$$\sigma_i(\mathbf{G}) \geq \sigma_i(\mathbf{A}^+), \quad i = 1, 2, \ldots, n.$$

PROOF. We assume, without loss of generality, that $\mathbf{A} = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$. By **6.4**, $\mathbf{G} = \begin{bmatrix} \Sigma^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}$ for some $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$. Thus

$$\mathbf{GG}' = \begin{bmatrix} \Sigma^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & \mathbf{Y}' \\ \mathbf{X}' & \mathbf{Z}' \end{bmatrix} = \begin{bmatrix} \Sigma^{-2} + \mathbf{XX}' & ? \\ ? & ? \end{bmatrix},$$

where ? denotes a submatrix not needed explicitly in the proof. Using the results of Section 4.2 we get

$$\begin{aligned} \sigma_i^2(\mathbf{G}) &= \lambda_i(\mathbf{GG}') \geq \lambda_i(\Sigma^{-2} + \mathbf{XX}') \\ &\geq \lambda_i(\Sigma^{-2}) = \sigma_i^2, \qquad i = 1, 2, \ldots, r. \end{aligned}$$

If $i > r$, then $\sigma_i = 0$, and the proof is complete. □

6.6. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix of rank* r *and let* $1 \leq k \leq r$. *Then* $\mathbf{A}^+$ *minimizes* $\operatorname{vol}_k(\mathbf{A})$ *in the class of g-inverses of* $\mathbf{A}$.

PROOF. Recall that the square of $\operatorname{vol}_k(\mathbf{A})$ equals the sum of squares of the singular values of $C_k(\mathbf{A})$. Now the result follows using (6) and **6.5**. □

Problems

1. The *Frobenius norm* of a matrix $\mathbf{A}$, denoted by $\|\mathbf{A}\|_F$, is defined as $(\sum_{i,j} a_{ij}^2)^{\frac{1}{2}}$. Show that if $\mathbf{A}$ is an $n \times n$ matrix, then for any g-inverse of $\mathbf{A}$, $\|\mathbf{G}\|_F \geq \|\mathbf{A}\|_F$ and that equality holds if and only if $\mathbf{G} = \mathbf{A}^+$.
2. Consider the matrix

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

Let $\mathcal{C}$ be the unit square with vertices

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Find the area of the image of $\mathcal{C}$ under the transformation $\mathbf{x} \mapsto \mathbf{Ax}$. Verify that the area equals vol($\mathbf{A}$).

4.7 Exercises

1. Let $\mathbf{A}$ be an $n \times n$ matrix of rank r with singular value decomposition as in (1). Then show that

$$\mathbf{A}^{+} = \mathbf{Q}\mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)\mathbf{P}.$$

2. Let $\mathbf{A}$, $\mathbf{B}$ be positive definite matrices of order $m \times m$, $n \times n$, respectively, and let $\mathbf{C}$ be $m \times n$. Prove that the matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B} \end{bmatrix}$$

is positive definite if and only if the largest singular value of $\mathbf{A}^{-1/2}\mathbf{C}\mathbf{B}^{-1/2}$ is less than 1.

3. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. Show that

$$\lambda_n = \min_{\|\mathbf{x}\|=1} \mathbf{x}'\mathbf{A}\mathbf{x}.$$

Obtain a similar expression for λ_k. Also obtain a max–min version of the Courant–Fischer theorem.

4. Let $\mathbf{A}$ be an $n \times n$ positive definite matrix partitioned as

$$\mathbf{A} = \begin{bmatrix} a_{11} & \mathbf{x}' \\ \mathbf{x} & \mathbf{B} \end{bmatrix}.$$

Show that

$$a_{11}|\mathbf{B}| - |\mathbf{A}| \geq \frac{\mathbf{x}'\mathbf{x}}{\lambda_1(\mathbf{B})}|\mathbf{B}|.$$

(Note that this inequality implies the Hadamard inequality.)

5. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix with $|\mathbf{A}| = 0$. Show that the $(n-1) \times (n-1)$ principal minors of $\mathbf{A}$ are either all nonnegative or all nonpositive.

6. Let $\mathbf{A}$ be an $m \times n$ matrix, $m < n$, and let $\mathbf{B}$ be obtained by deleting any column of $\mathbf{A}$. Show that the singular values of $\mathbf{A}$, $\mathbf{B}$ satisfy

$$\sigma_i(\mathbf{A}) \geq \sigma_i(\mathbf{B}) \geq \sigma_{i+1}(\mathbf{A}), \qquad i = 1, \ldots, m.$$

7. Let $\mathbf{A}$, $\mathbf{B}$ be positive semidefinite $n \times n$ matrices. Prove that

$$\lambda_1(\mathbf{AB}) \leq \lambda_1(\mathbf{A})\lambda_1(\mathbf{B}), \quad \lambda_n(\mathbf{AB}) \geq \lambda_n(\mathbf{A})\lambda_n(\mathbf{B}).$$

8. Let $\mathbf{A}$ be a positive definite matrix that is partitioned as

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A_{11}}$ is square. Show that the eigenvalues of $\mathbf{A}$ majorize the eigenvalues of

$$\begin{bmatrix} \mathbf{A_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_{22}} \end{bmatrix}.$$

9. Let $\mathbf{A}$ be an $n \times n$ matrix of rank r. Show that $C_r(\mathbf{A})$ has rank 1.

10. If $\mathbf{A}$ is an $m \times n$ matrix of rank 1, then show that

$$\mathbf{A}^+ = \frac{1}{\mathrm{vol}^2(\mathbf{A})}\mathbf{A}'.$$

(Note that since $\mathbf{A}$ has rank 1, $\mathrm{vol}^2(\mathbf{A}) = \sum_{i,j} a_{ij}^2$.)

11. Let $\mathbf{A}$ be an $n \times n$ matrix of rank r. Show that for any $I, J \in Q_{r,n}$,

$$|\mathbf{A}^+_{\mathbf{IJ}}| = \frac{1}{\mathrm{vol}^2(\mathbf{A})}|\mathbf{A_{JI}}|.$$

(Thus the $r \times r$ minors of $\mathbf{A}^+$ are proportional to the corresponding minors of $\mathbf{A}'$.)

12. Let $\mathbf{A}$ be an $n \times n$ matrix of rank r and let $\mathbf{A} = \mathbf{UV}$ be a rank factorization. Show that the following conditions are equivalent:

(i) $R(\mathbf{A}) = R(\mathbf{A^2})$.
(ii) $\mathbf{VU}$ is a nonsingular $r \times r$ matrix.
(iii) There exists a reflexive g-inverse $\mathbf{G}$ of $\mathbf{A}$ such that $\mathbf{AG} = \mathbf{GA}$. (Such a g-inverse is called a *group inverse* of $\mathbf{A}$.)

13. Let $\mathbf{A}$ be an $n \times n$ matrix of rank r and suppose $R(\mathbf{A}) = R(\mathbf{A^2})$. Prove the following propertries of group inverse:

a. The group inverse of $\mathbf{A}$ is unique. We will denote the group inverse of $\mathbf{A}$ by $\mathbf{A}^\#$.
b. $\mathcal{C}(\mathbf{A}^\#) = \mathcal{C}(\mathbf{A})$, $\mathcal{R}(\mathbf{A}^\#) = \mathcal{R}(\mathbf{A})$.
c. If $r = 1$, then $\mathbf{A}^\# = \frac{1}{(\mathrm{trace}\mathbf{A})^2}\mathbf{A}$.
d. for any $I, J \in Q_{r,n}$,

$$|\mathbf{A}^\#_{\mathbf{IJ}}| = \frac{1}{(\mathrm{trace}C_r(\mathbf{A}))^2}|\mathbf{A_{IJ}}|.$$

14. Let $\mathbf{A}$ be an $m \times n$ matrix of rank r. Show that $\mathbf{A}$ has a group inverse if and only if the sum of the $r \times r$ principal minors of $\mathbf{A}$ is nonzero.

4.8 Hints and Solutions

Section 4

1. The principal components are $\frac{1}{\sqrt{2}}(x_1 + x_2)$ and $\frac{1}{\sqrt{2}}(x_1 - x_2)$ with the corresponding variances 80 and 40, respectively.

2. The answer is given by (100 times) the ratio of the least eigenvalue of the matrix and the trace, and is seen to be approximately 9.5 percent.

Section 5

1. First canonical correlation: .9337.
First pair of canonical variates: $.0957X_1 + .9954X_2$, $.6169X_3 + .7870X_4$.
Second canonical correlation: .3054.
Second pair of canonical variates: $.9954X_1 - .0957X_2$, $.7870X_3 - .6169X_4$.

2. Hint: Use the formula for the determinant involving Schur complements.

Section 6

1. Note that $\|\mathbf{A}\|_F = (\sum_i \sigma_i^2(\mathbf{A}))^{\frac{1}{2}}$. Thus the first part of the result follows from **6.5**. We give another, easier, argument. Since $\|\mathbf{A}\|_F = \|\mathbf{PAQ}\|_F$ for orthogonal $\mathbf{P}, \mathbf{Q}$, we may assume, without loss of generality, that

$$\mathbf{A} = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ and r is the rank of $\mathbf{A}$. Then

$$\mathbf{G} = \begin{bmatrix} \Sigma^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix},$$

for some $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$. Furthermore, $\mathbf{G} = \mathbf{A}^+$ precisely when $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are all zero. Clearly,

$$\|\mathbf{G}\|_F^2 - \|\mathbf{A}\|_F^2 = \sum_{i,j} x_{ij}^2 + \sum_{i,j} y_{ij}^2 + \sum_{i,j} z_{ij}^2,$$

and the result, including the assertion about equality, follows.

2. The image of $\mathcal{C}$ is the parallelogram with vertices

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}.$$

The area of the parallelogram is $\sqrt{3}$, which also equals $\operatorname{vol}(\mathbf{A})$.

Section 7

2. Suppose σ_1 is the largest singular value of $\mathbf{A}^{-1/2}\mathbf{C}\mathbf{B}^{-1/2}$. Then

$$\sigma_1^2 = \lambda_1(\mathbf{A}^{-1/2}\mathbf{C}\mathbf{B}^{-1}\mathbf{C}'\mathbf{A}^{-1/2}).$$

Thus $\sigma_1^2 < 1$ if and only if

$$\mathbf{I} - \mathbf{A}^{-1/2}\mathbf{C}\mathbf{B}^{-1}\mathbf{C}'\mathbf{A}^{-1/2}$$

is positive definite, or equivalently, $\mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}'$ is positive definite. Since $\mathbf{B}$ is positive definite, the partitioned matrix is positive definite if and only if the Schur complement $\mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}'$ is positive definite, and the result is proved.

4. Hint: By the Schur complement formula, we have

$$|\mathbf{A}| = |\mathbf{B}|(a_{11} - \mathbf{x}'\mathbf{B}^{-1}\mathbf{x}).$$

Use extremal representation for the least eigenvalue of $\mathbf{B}^{-1}$.

5. Since $|\mathbf{A}| = 0$, then 0 is an eigenvalue of $\mathbf{A}$. If 0 has multiplicity greater than one, then by the interlacing principle, any $(n-1) \times (n-1)$ principal submatrix must have 0 as an eigenvalue and therefore has determinant zero. So suppose 0 has multiplicity one. Let $\mathbf{B}$, $\mathbf{C}$ be principal submatrices of order $(n-1) \times (n-1)$. we must show $|\mathbf{B}||\mathbf{C}| \geq 0$. This is trivial if either $\mathbf{B}$ or $\mathbf{C}$ is singular. So suppose $\mathbf{B}$, $\mathbf{C}$ are nonsingular. Again by the interlacing principle it follows that the number of positive eigenvalues (and hence the number of negative eigenvalues) of $\mathbf{B}$ and $\mathbf{C}$ are the same. Hence $|\mathbf{B}||\mathbf{C}| \geq 0$.

6. Hint: Observe that $\mathbf{B}'\mathbf{B}$ is a principal submatrix of $\mathbf{A}'\mathbf{A}$ and use the interlacing principle.

7.

$$\begin{aligned}
\lambda_1(\mathbf{AB}) &= \lambda_1(\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}) \\
&= \max_{\|\mathbf{x}\|=1} \mathbf{x}'\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}\mathbf{x} \\
&= \max_{\|\mathbf{x}\|=1} \frac{\mathbf{x}'\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}\mathbf{x}}{\mathbf{x}'\mathbf{A}\mathbf{x}} \cdot \mathbf{x}'\mathbf{A}\mathbf{x} \\
&\leq \max_{\|\mathbf{x}\|=1} \frac{\mathbf{x}'\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}\mathbf{x}}{\mathbf{x}'\mathbf{A}\mathbf{x}} \cdot \max_{\|\mathbf{x}\|=1} \mathbf{x}'\mathbf{A}\mathbf{x} \\
&= \lambda_1(\mathbf{B})\lambda_1(\mathbf{A}).
\end{aligned}$$

The second inequality is proved similarly.

8. Suppose $\mathbf{A}_{11}$ is $r \times r$, so that $\mathbf{A}_{22}$ is $(n-r) \times (n-r)$. Let $\mathbf{D}$ be the diagonal matrix with $d_{ii} = 1, 1 \leq i \leq r$, and $d_{ii} = -1$ otherwise. Then

$$\mathbf{A} + \mathbf{D}\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\mathbf{D} = 2\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}.$$

The result follows by **3.4**.

9. If $\mathbf{A}$ has singular values $\sigma_1, \ldots, \sigma_r$, then $C_r(\mathbf{A})$ has only one nonzero singular value, namely $\sigma_1 \cdots \sigma_r$. Therefore, $C_r(\mathbf{A})$ has rank 1.

10. Hint: Let $\mathbf{A} = \mathbf{x}\mathbf{y}'$ be a rank factorization and verify the definition of the Moore–Penrose inverse.

11. It is easily verified that $(C_r(\mathbf{A}))^+ = C_r(\mathbf{A}^+)$. Using the two preceding exerises, $C_r(\mathbf{A})$ has rank 1 and

$$C_r(\mathbf{A}^+) = (C_r(\mathbf{A}))^+ = \frac{1}{\mathrm{vol}^2(C_r(\mathbf{A}))} C_r(\mathbf{A}').$$

(Here we also used the simple fact that $(C_r(\mathbf{A}))' = C_r(\mathbf{A}')$.) Now observe that since $\mathbf{A}$ has rank r, $\mathrm{vol}^2(C_r(\mathbf{A})) = \mathrm{vol}^2(\mathbf{A})$.

12. If $R(\mathbf{A}) = R(\mathbf{A^2})$, then $\mathbf{A} = \mathbf{A^2X}$ for some matrix $\mathbf{X}$. Then $\mathbf{UV} = \mathbf{UVUVX}$. Note that $\mathbf{U}$ admits a left inverse, say $\mathbf{U}^-$, and $\mathbf{V}$ admits a right inverse, say $\mathbf{V}^-$. Using this in the previous equation we get $\mathbf{I_r} = \mathbf{VUVXV}^-$. Therefore $\mathbf{VU}$ is nonsingular. Thus (i) implies (ii). If $\mathbf{VU}$ is nonsingular, then it can be verified that $\mathbf{G} = \mathbf{U}(\mathbf{VU})^{-2}\mathbf{V}$ is a group inverse of $\mathbf{A}$, and hence (iii) holds. Finally, if (iii) holds, then $\mathbf{A} = \mathbf{AGA} = \mathbf{A^2G}$, and it follows that $R(\mathbf{A}) = R(\mathbf{A^2})$. Thus (iii) implies (i), and the proof is complete.

13. (i) If $\mathbf{G_1}$, $\mathbf{G_2}$ are both group inverses, then

$$\mathbf{AG_1A} = \mathbf{A}, \quad \mathbf{G_1AG_1} = \mathbf{G_1}, \quad \mathbf{AG_1} = \mathbf{G_1A}$$

and

$$\mathbf{AG_2A} = \mathbf{A}, \quad \mathbf{G_2AG_2} = \mathbf{G_2}, \quad \mathbf{AG_2} = \mathbf{G_2A}.$$

Then

$$\begin{aligned}\mathbf{G_1} &= \mathbf{G_1AG_1} = \mathbf{G_1AG_2AG_1} = \mathbf{AG_1AG_2G_1} = \mathbf{AG_2G_1}\\ &= \mathbf{G_2AG_1} = \mathbf{G_2G_1A} = \mathbf{G_2G_1AG_2A} = \mathbf{G_2AG_1AG_2}\\ &= \mathbf{G_2AG_2} = \mathbf{G_2}.\end{aligned}$$

(ii) We have $\mathbf{A}^{\#} = \mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#} = \mathbf{A}(\mathbf{A}^{\#})^{\mathbf{2}}$. Therefore, $\mathcal{C}(\mathbf{A}^+) \subset \mathcal{C}(\mathbf{A})$. Since $R(\mathbf{A}^{\#}) = R(\mathbf{A})$, the two spaces must be equal. The second part is proved similarly. The proof of (iii), (iv) is similar to the corresponding properties of the Moore–Penrose inverse (see Exercises 10,11).

14. Let $\mathbf{A} = \mathbf{UV}$ be a rank factorization. As observed in Exercise 12, $\mathbf{A}$ admits a group inverse if and only if $\mathbf{VU}$ is nonsingular. It can be seen by the Cauchy–Binet formula that the sum of the $r \times r$ principal minors of $\mathbf{A}$ equals $|\mathbf{UV}|$, and the result follows.

This page intentionally left blank

5

Block Designs and Optimality

5.1 Reduced Normal Equations

Suppose we want to compare v treatments, $1, \ldots, v$. The treatments could be different fertilizers in an agricultural experiment, drugs in medicine, machines in an industrial setting, teaching methods in education, and so on. The experimental material is available in the form of units generally referred to as plots. The plots are grouped into blocks such that within each block the plots are as similar as possible. Suppose there are b blocks of sizes $k_1, \ldots, k_b$. Let n denote the total number of plots, so that $n = k_1 + \cdots + k_b$.

A *design* (or a *block design*) is an allocation of the v treatments to the n plots. Suppose we get one observation from each plot. The parameters of the model are μ (general effect), $\tau_1, \ldots, \tau_v$ (effects due to treatments), and $\alpha_1, \ldots, \alpha_b$ (effects due to blocks). The linear model arising out of the design is as follows. The expected value of the observation from any plot is equal to μ plus the effect due to the corresponding block plus the effect due to the treatment that is applied to the plot. We assume that the observations are independent normal with variance σ^2.

Let $\mathbf{y}$ denote the $n \times 1$ vector of observations and let $\mathbf{X}$ be the coefficient matrix in the linear model. Instead of writing $\mathbf{X}$ explicitly, it is more convenient to write $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$; these are the only matrices needed to construct the normal equations.

Let $\mathbf{N} = (n_{ij})$ denote the $v \times b$ *incidence matrix* of treatments versus blocks. Thus n_{ij} is the number of times the ith treatment occurs in the jth block. It can be seen that

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & k_1 & \cdots & \cdots & k_b & r_1 & \cdots & \cdots & r_v \\ k_1 & k_1 & 0 & \cdots & 0 & & & & \\ \vdots & 0 & k_2 & & 0 & & \mathbf{N}' & & \\ \vdots & \vdots & & \ddots & & & & & \\ k_b & 0 & 0 & & k_b & & & & \\ r_1 & & & & & r_1 & 0 & \cdots & 0 \\ \vdots & & \mathbf{N} & & & 0 & r_2 & & 0 \\ \vdots & & & & & \vdots & & \ddots & \\ r_v & & & & & 0 & 0 & & r_v \end{bmatrix},$$

where r_i is the number of times treatment i occurs in the entire design, $i = 1, \ldots, v$. Also,

$$\mathbf{X}'\mathbf{y} = (G, B_1, \ldots, B_b, T_1, \ldots, T_v)',$$

where G is the total of all the observations, B_i is the total of the observations in the ith block, and T_j is the total of the observations corresponding to the jth treatment. The normal equations are

$$\mathbf{X}'\mathbf{X} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_b \\ \tau_1 \\ \vdots \\ \tau_v \end{bmatrix} = \mathbf{X}'\mathbf{y}. \tag{1}$$

Our interest is in comparing the treatments, and therefore we are not interested in μ or the block effects. We first obtain a general result on reducing the normal equations, eliminating some of the parameters.

1.1. *Consider the linear model* $E(\mathbf{y}) = \mathbf{A}\boldsymbol{\beta}$, $\quad D(\mathbf{y}) = \sigma^2\mathbf{I}$ *and suppose* $\boldsymbol{\beta}$, $\mathbf{A}$, *and* $\mathbf{z} = \mathbf{A}'\mathbf{y}$ *are conformally partitioned as*

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2], \quad \mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}.$$

Then the equations

$$(\mathbf{A}_2'\mathbf{A}_2 - \mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2)\boldsymbol{\beta}_2 = \mathbf{z}_2 - \mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{z}_1 \tag{2}$$

are the reduced normal equations for $\boldsymbol{\beta}_2$ in the following sense: A function $\boldsymbol{\ell}'\boldsymbol{\beta}_2$ is estimable if and only if

$$\boldsymbol{\ell}' \in \mathcal{R}(\mathbf{A}_2'\mathbf{A}_2 - \mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2), \tag{3}$$

and in that case its BLUE is $\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}_2}$, where $\widehat{\boldsymbol{\beta}_2}$ is any solution of (2).

PROOF. First observe that since $\mathcal{R}(\mathbf{A}_1) = \mathcal{R}(\mathbf{A}_1'\mathbf{A}_1)$, the matrix

$$\mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2$$

is invariant under the choice of the g-inverse.

Suppose that $\boldsymbol{\ell}'\boldsymbol{\beta}_2$ is estimable. Then there exists $\mathbf{c}$ such that

$$E(\mathbf{c}'\mathbf{y}) = \boldsymbol{\ell}'\boldsymbol{\beta}_2.$$

Let

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$$

be the partitioning of $\mathbf{y}$, where $\mathbf{c}$ is conformable with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. Then

$$E(\mathbf{c}_1'\mathbf{y}_1) + E(\mathbf{c}_2'\mathbf{y}_2) = \boldsymbol{\ell}'\boldsymbol{\beta}_2, \tag{4}$$

and hence

$$\mathbf{c}_1'\mathbf{A}_1'\mathbf{A}_1\boldsymbol{\beta}_1 + \mathbf{c}_1'\mathbf{A}_1'\mathbf{A}_2\boldsymbol{\beta}_2 + \mathbf{c}_2'\mathbf{A}_2'\mathbf{A}_1\boldsymbol{\beta}_1 + \mathbf{c}_2'\mathbf{A}_2'\mathbf{A}_2\boldsymbol{\beta}_2 = \boldsymbol{\ell}'\boldsymbol{\beta}_2.$$

Therefore

$$\mathbf{c}_1'\mathbf{A}_1'\mathbf{A}_1 + \mathbf{c}_2'\mathbf{A}_2'\mathbf{A}_1 = \mathbf{0}, \tag{5}$$

$$\mathbf{c}_1'\mathbf{A}_1'\mathbf{A}_2 + \mathbf{c}_2'\mathbf{A}_2'\mathbf{A}_2 = \boldsymbol{\ell}'. \tag{6}$$

By (5),

$$\mathbf{c}_1'\mathbf{A}_1'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2 + \mathbf{c}_2'\mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2 = \mathbf{0},$$

and hence

$$\mathbf{c}_1'\mathbf{A}_1'\mathbf{A}_2 + \mathbf{c}_2'\mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2 = \mathbf{0}.$$

Now from (6),

$$\boldsymbol{\ell}' = \mathbf{c}_2'(\mathbf{A}_2'\mathbf{A}_2 - \mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2),$$

and (3) is proved.

Conversely, if (3) holds, then for some matrix $\mathbf{M}$,

$$\boldsymbol{\ell}'\boldsymbol{\beta}_2 = \mathbf{M}(\mathbf{A}_2'\mathbf{A}_2 - \mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{A}_2)\boldsymbol{\beta}_2.$$

Set

$$\mathbf{c}_1' = -\mathbf{M}\mathbf{A}_2'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^-, \quad \mathbf{c}_2' = \mathbf{M}.$$

Then $\mathbf{c}_1, \mathbf{c}_2$ satisfy (5), (6), and hence (4) holds. Thus $\boldsymbol{\ell}'\boldsymbol{\beta}_2$ is estimable. Thus we have shown that $\boldsymbol{\ell}'\boldsymbol{\beta}_2$ is estimable if and only if (3) holds. The second part is left as an exercise. □

5.2 The C-Matrix

We apply **1.1** to the normal equations (1). A g-inverse of $\mathbf{A}_1'\mathbf{A}_1$ in this case is found to be

$$\begin{bmatrix} n & k_1 & \cdots & k_b \\ k_1 & k_1 & & \\ \vdots & & \ddots & \\ k_b & & & k_b \end{bmatrix}^{-} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{k_1} & & \\ \vdots & & \ddots & \\ 0 & & & \frac{1}{k_b} \end{bmatrix}.$$

The reduced normal equations for the treatment parameters are

$$\mathbf{C}\tau = \mathbf{Q};$$

where $\tau = (\tau_1, \dots, \tau_v)'$, $\mathbf{Q}$ is a function of $\mathbf{y}$, and the matrix $\mathbf{C}$, known as the C-matrix of the design, is given by

$$\mathbf{C} = \text{diag}\,(r_1, \dots, r_v) - \mathbf{N}\text{diag}(\frac{1}{k_1}, \cdots, \frac{1}{k_b})\mathbf{N}'.$$

Clearly, $\mathbf{C}$ is symmetric, and it is positive semidefinite, since it is the (generalized) Schur complement (see Exercise 2 in Chapter 3) of a principal submatrix in a positive semidefinite matrix.

The row sums (and hence the column sums) of $\mathbf{C}$ are zero. This is seen as follows. Denoting by $\mathbf{e}$ the vector of all ones,

$$\begin{aligned} \mathbf{Ce} &= \begin{bmatrix} r_1 \\ \vdots \\ r_v \end{bmatrix} - \mathbf{N} \begin{bmatrix} \frac{1}{k_1} & & \\ & \ddots & \\ & & \frac{1}{k_b} \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_b \end{bmatrix} \\ &= \begin{bmatrix} r_1 \\ \vdots \\ r_v \end{bmatrix} - \mathbf{Ne} = \mathbf{0}. \end{aligned}$$

It follows that $\mathbf{C}$ is singular and $R(\mathbf{C}) \leq v - 1$.

For convenience we will now consider designs with equal block sizes, $k_1 = \cdots = k_b = k$. We will denote the C-matrix of the design d by $\mathbf{C_d}$.

A function $\ell_1\tau_1 + \cdots + \ell_v\tau_v$ is estimable if and only if $\boldsymbol{\ell}' = (\ell_1, \dots, \ell_v)$ is in the row space of $\mathbf{C}$. Since $\mathbf{Ce} = \mathbf{0}$, a necessary condition for estimability is that

$$\ell_1 + \cdots + \ell_v = 0. \tag{7}$$

The function $\ell_1\tau_1 + \cdots + \ell_v\tau_v$ is called a *contrast* if (7) is satisfied. A contrast of the form $\tau_i - \tau_j,\ i \neq j$ is called an *elementary contrast*. A design is said to be *connected* if all contrasts are estimable.

2.1. *A design d is connected if and only if the rank of $\mathbf{C_d}$ is $v-1$.*

PROOF. If d is connected, then all contrasts are estimable. In particular, the contrasts

$$\tau_1 - \tau_2, \quad \tau_1 - \tau_3, \quad \ldots, \quad \tau_1 - \tau_v$$

are estimable. Thus the vectors

$$\begin{gathered}(1, -1, 0, \ldots, 0),\\ (1, 0, -1, \ldots, 0),\\ \cdots\\ (1, 0, 0, \ldots, -1)\end{gathered}$$

are in $\mathcal{R}(\mathbf{C_d})$. These vectors are linearly independent, and therefore $R(\mathbf{C_d}) \geq v-1$. But it is always true that $R(\mathbf{C_d}) \leq v-1$ (since the null space of $\mathbf{C_d}$ contains $\mathbf{e}$), and therefore $R(\mathbf{C_d}) = v-1$. The converse is proved similarly. □

5.3 E-, A-, and D-Optimality

Let $\mathcal{D}(v, b, k)$ denote the class of all connected designs with v treatments arranged in b blocks of size k. (It is possible that for some values of v, b, k, the class is empty.) We now consider the problem of choosing a design in $\mathcal{D}(v, b, k)$ that has some desirable properties. Let d be a design. The BLUE of an estimable function $\boldsymbol{\ell}'\boldsymbol{\tau}$ is $\boldsymbol{\ell}'\widehat{\boldsymbol{\tau}}$, where $\widehat{\boldsymbol{\tau}}$ is a solution of the reduced normal equations, and the variance of the BLUE is

$$\sigma^2 \boldsymbol{\ell}'\mathbf{C_d^-}\boldsymbol{\ell}$$

for any choice of the g-inverse. The design d is preferable if this variance is "small" for all contrasts. There are many different ways of making this precise, thus leading to different optimality criteria.

The design d is said to be *E-optimal* if it is connected and it minimizes

$$\max_{\boldsymbol{\ell}'\mathbf{e}=0, \|\boldsymbol{\ell}\|=1} \boldsymbol{\ell}'\mathbf{C}^-\boldsymbol{\ell}. \tag{8}$$

We will denote the eigenvalues of $\mathbf{C_d}$ by

$$0 = \mu_{0,d} \leq \mu_{1,d} \leq \cdots \leq \mu_{v-1,d}.$$

3.1. *Suppose $\mathcal{D}(v, b, k)$ is nonempty. Then $d^* \in \mathcal{D}(v, b, k)$ is E-optimal if*

$$\mu_{1,d^*} = \max_{d \in \mathcal{D}(v,b,k)} \mu_{1,d}.$$

PROOF. Let

$$\mathbf{x_0} = \frac{1}{\sqrt{v}}\mathbf{e}, \quad \mathbf{x_1}, \quad \ldots, \quad \mathbf{x_{v-1}}$$

be an orthonormal set of eigenvectors corresponding to $\mu_{0,d}, \mu_{1,d}, \ldots, \mu_{v-1,d}$, respectively. Thus

$$\mathbf{C_d} = \mu_{0,d}\mathbf{x_0}\mathbf{x_0'} + \mu_{1,d}\mathbf{x_1}\mathbf{x_1'} + \cdots + \mu_{v-1,d}\mathbf{x_{v-1}}\mathbf{x_{v-1}'}.$$

If d is connected, then

$$\mathbf{C_d^+} = \frac{1}{\mu_{1,d}}\mathbf{x_1}\mathbf{x_1'} + \cdots + \frac{1}{\mu_{v-1,d}}\mathbf{x_{v-1}}\mathbf{x_{v-1}'}.$$

A vector $\boldsymbol{\ell}$ with $\boldsymbol{\ell}'\mathbf{e} = 0$, $\|\boldsymbol{\ell}\| = 1$ can be expressed as

$$\boldsymbol{\ell} = \beta_1\mathbf{x_1} + \cdots + \beta_{v-1}\mathbf{x_{v-1}},$$

where $\beta_1^2 + \cdots + \beta_{v-1}^2 = 1$. Thus

$$\begin{aligned}
\boldsymbol{\ell}'\mathbf{C_d^+}\boldsymbol{\ell} &= (\beta_1\mathbf{x_1'} + \cdots + \beta_{v-1}\mathbf{x_{v-1}'})\mathbf{C}^+(\beta_1\mathbf{x_1} + \cdots + \beta_{v-1}\mathbf{x_{v-1}}) \\
&= \frac{\beta_1^2}{\mu_{1,d}}\mathbf{x_1'}\mathbf{x_1} + \cdots + \frac{\beta_{v-1}^2}{\mu_{v-1,d}}\mathbf{x_{v-1}'}\mathbf{x_{v-1}} \\
&\leq \frac{1}{\mu_{1,d}}.
\end{aligned}$$

Equality holds in the above inequality when $\boldsymbol{\ell} = \mathbf{x_1}$. The result now follows in view of the definition of E-optimality. □

A design is said to be *binary* if a treatment occurs in any block at most once. We will denote the treatment versus block incidence matrix of the design d by $\mathbf{N_d}$. Then for a binary design d, $\mathbf{N_d}$ consists of only zeros and ones.

3.2. *For any* $d \in \mathcal{D}(v, b, k)$, trace $\mathbf{C_d} \leq b(k-1)$. *Equality holds if and only if the design is binary.*

PROOF. By the definition of $\mathbf{C_d}$ it follows that

$$\operatorname{trace}\mathbf{C_d} = \sum_{i=1}^{v} r_i - \frac{1}{k}\sum_{i=1}^{v}\sum_{j=1}^{b} n_{ij}^2, \tag{9}$$

where $\mathbf{N_d} = (n_{ij})$. It is an easy exercise to show that if $k < v$,

$$\sum_{i=1}^{v}\sum_{j=1}^{b} n_{ij}^2$$

is minimized subject to the conditions that each n_{ij} is a nonnegative integer and

$$\sum_{i=1}^{v}\sum_{j=1}^{b} n_{ij} = bk$$

when each n_{ij} is 0 or 1, in which case

$$\sum_{i=1}^{v}\sum_{j=1}^{b} n_{ij}^2 = \sum_{i=1}^{v}\sum_{j=1}^{b} n_{ij} = bk.$$

Thus we get from (9),

$$\text{trace}\mathbf{C_d} \le \sum_{i=1}^{v} r_i - \frac{1}{k} bk = b(k-1).$$

Equality holds if and only if each n_{ij} is 0 or 1, i.e., the design is binary. □

A design $d^* \in \mathcal{D}(v, b, k)$ is said to be A-optimal if

$$\sum_{i=1}^{v-1} \frac{1}{\mu_{i,d^*}} \le \sum_{i=1}^{v-1} \frac{1}{\mu_{i,d}} \tag{10}$$

for any $d \in \mathcal{D}(v, b, k)$.

The statistical significance of A-optimality is given in the following:

3.3. *A design $d^* \in \mathcal{D}(v, b, k)$ is A-optimal if it minimizes the average variance of the BLUE of an elementary contrast.*

PROOF. Consider the elementary contrast $\tau_i - \tau_j$. We can write $\tau_i - \tau_j = \mathbf{z}'\boldsymbol{\tau}$, where $\mathbf{z}$ is a vector of order v with 1 at position i, -1 at position j, and zeros elsewhere. Let $\mathbf{C_d^+} = \mathbf{T}$. It can be seen that $\mathbf{Te} = \mathbf{0}$; use, for example, the representation for $\mathbf{C_d^+}$ given in the proof of **3.1**. Then the variance of the BLUE of $\tau_i - \tau_j$ is

$$\sigma^2 \mathbf{z}'\mathbf{C_d^+}\mathbf{z} = \sigma^2 \mathbf{z}'\mathbf{T}\mathbf{z} = \sigma^2(t_{ii} + t_{jj} - 2t_{ij}).$$

Hence the average variance of the BLUE of an elementary contrast is $\frac{2\sigma^2}{v(v-1)}$ times

$$\begin{aligned}
\sum_{i<j}(t_{ii} + t_{jj} - 2t_{ij}) &= \frac{1}{2}\sum_{i=1}^{v}\sum_{j=1}^{v}(t_{ii} + t_{jj} - 2t_{ij}) \\
&= \frac{1}{2}(2\text{trace }\mathbf{T} - 2\mathbf{e}'\mathbf{Te}) \\
&= \text{trace }\mathbf{T} \qquad \text{since } \mathbf{Te} = \mathbf{0} \\
&= \sum_{i=1}^{v-1} \frac{1}{\mu_{i,d}},
\end{aligned}$$

and the result is proved. □

A design $d \in \mathcal{D}(v, b, k)$ is said to be D-optimal if it maximizes $\prod_{i=1}^{v-1} \mu_{i,d}$.

To obtain the statistical interpretation of D-optimality we first establish the following:

3.4. *Let $d \in \mathcal{D}(v, b, k)$. Let $\mathbf{P}$ be a $v \times (v-1)$ matrix whose columns form an orthonormal basis for the row space of $\mathbf{C_d}$. Then the eigenvalues of $\mathbf{P}'\mathbf{C_d^-}\mathbf{P}$ are $\frac{1}{\mu_{1,d}}, \dots, \frac{1}{\mu_{v-1,d}}$.*

PROOF. First observe that $\mathbf{P}'\mathbf{C_d^-}\mathbf{P}$ is invariant under the choice of the g-inverse. So we will consider $\mathbf{P}'\mathbf{C_d^+}\mathbf{P}$. Let

$$\mathbf{z} = (\frac{1}{\sqrt{v}}, \dots, \frac{1}{\sqrt{v}})'.$$

Then $\mathbf{Q} = (\mathbf{P}, \mathbf{z})$ is an orthogonal matrix. We have

$$\begin{bmatrix} \mathbf{P}' \\ \mathbf{z}' \end{bmatrix} \mathbf{C}_\mathbf{d}^{+}[\mathbf{P}, \mathbf{z}] = \begin{bmatrix} \mathbf{P}'\mathbf{C}_\mathbf{d}^{+}\mathbf{P} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix},$$

since $\mathbf{C}_\mathbf{d}^{+}\mathbf{z} = \mathbf{0}$. Thus the eigenvalues of $\mathbf{P}'\mathbf{C}_\mathbf{d}^{+}\mathbf{P}$ are the same as the nonzero eigenvalues of

$$\begin{bmatrix} \mathbf{P}' \\ \mathbf{z}' \end{bmatrix} \mathbf{C}_\mathbf{d}^{+}[\mathbf{P}, \mathbf{z}],$$

which, in turn, are the same as the nonzero eigenvalues of $\mathbf{C}_\mathbf{d}^{+}$. □

As an immediate consequence of **3.4** we have

$$|\mathbf{P}'\mathbf{C}_\mathbf{d}^{-}\mathbf{P}| = \prod_{i=1}^{v-1} \frac{1}{\mu_{i,d}}.$$

Let $\boldsymbol{\beta} = \mathbf{P}'\boldsymbol{\tau}$. Then the BLUE of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = \mathbf{P}'\widehat{\boldsymbol{\tau}}$, and it has the dispersion matrix $\sigma^2\mathbf{P}'\mathbf{C}_\mathbf{d}^{-}\mathbf{P}$. Thus

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2\mathbf{P}'\mathbf{C}_\mathbf{d}^{-}\mathbf{P}).$$

Therefore, a confidence ellipsoid for $\boldsymbol{\beta}$ is of the form

$$\left\{\boldsymbol{\beta} \in R^{v-1} : (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{P}'\mathbf{C}_\mathbf{d}^{-}\mathbf{P})^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le c\sigma^2\right\}. \tag{11}$$

The volume of this ellipsoid is proportional to $|\mathbf{P}'\mathbf{C}_\mathbf{d}^{-}\mathbf{P}|$ (see Exercise 9).

Thus a design is D-optimal if it minimizes the volume of the ellipsoid (11) over $\mathcal{D}(v, b, k)$ for any c, σ^2.

A design d is called a *balanced incomplete block design (BIBD)* if it is a binary design such that

(i) $r_1 = \cdots = r_v = r$, say, and
(ii) any two treatments occur together in the same number, say λ, of blocks.

There does not necessarily exist a BIBD in $\mathcal{D}(v, b, k)$ for every given choice of v, b, k. Some necessary conditions for existence can be derived as follows.

If $d \in \mathcal{D}(v, b, k)$ is a BIBD, then

$$\mathbf{N_d e} = \begin{bmatrix} r \\ \vdots \\ r \end{bmatrix},$$

and hence $\mathbf{e}'\mathbf{N_d e} = vr$. Similarly, $\mathbf{e}'\mathbf{N_d e} = (\mathbf{e}'\mathbf{N_d})\mathbf{e} = (k, \dots, k)\mathbf{e} = bk$. Thus

$$bk = vr. \tag{12}$$

We have

$$\mathbf{N_d N_d'} = \begin{bmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \vdots & & \ddots & \\ \lambda & \lambda & \cdots & r \end{bmatrix}$$

and hence $\mathbf{N_d N_d' e} = (r + \lambda(v-1))\mathbf{e}$. Also, $\mathbf{N_d N_d' e} = \mathbf{N_d}(\mathbf{N_d' e}) = kr\mathbf{e}$. Thus

$$\lambda(v-1) = r(k-1). \tag{13}$$

We will use (12), (13) in the subsequent discussion.

A simple way to construct a BIBD is to take v treatments and let every possible pair of distinct treatments be a block. Then $k = 2, r = v - 1, \lambda = 1$, and $b = \frac{1}{2}v(v-1)$. Another example of a BIBD is the following design. Here the columns denote blocks:

$$\begin{matrix} 1 & 3 & 1 & 2 & 1 & 2 & 5 \\ 2 & 4 & 3 & 4 & 4 & 3 & 6 \\ 5 & 5 & 6 & 6 & 7 & 7 & 7 \end{matrix}$$

3.5. *Let d^* be a BIBD with parameters v, b, k, r, λ. Then the eigenvalues of $\mathbf{C_{d^*}}$ are 0 with multiplicity 1 and $\frac{v\lambda}{k}$ with multiplicity $v - 1$.*

PROOF. Since d^* is a BIBD,

$$\mathbf{C_d} = \frac{v\lambda}{k}(\mathbf{I} - \frac{1}{v}\mathbf{ee'}).$$

The result follows by (12), (13). □

We now prove an optimality property of a BIBD when it exists.

3.6. *Let d^* be a BIBD with parameters v, b, k, r, λ. Then d^* is E-, A-, and D-optimal in $\mathcal{D}(v, b, k)$.*

PROOF. Let $d \in \mathcal{D}(v, b, k)$. Then

$$\begin{aligned} \mu_{1,d} &\le \frac{1}{v-1}\sum_{i=1}^{v-1}\mu_{i,d} \\ &= \frac{\text{trace}\mathbf{C_d}}{v-1} \\ &\le \frac{b(k-1)}{v-1} \qquad \text{(by } \mathbf{3.2}) \\ &= \frac{v\lambda}{k} \qquad \text{(by (12), (13))} \\ &= \mu_{1,d^*} \qquad \text{(by } \mathbf{3.5}) \end{aligned}$$

and hence d^* is E-optimal.

Let $d \in \mathcal{D}(v, b, k)$. By the arithmetic mean–harmonic mean inequality we have

$$\begin{aligned}
\sum_{i=1}^{v-1} \frac{1}{\mu_{i,d}} &\geq \frac{(v-1)^2}{\sum_{i=1}^{v-1} \mu_{i,d}} \\
&= \frac{(v-1)^2}{\text{trace } \mathbf{C_d}} \\
&\geq \frac{(v-1)^2}{b(k-1)} \quad \text{(by } \mathbf{3.2}\text{)} \\
&= \sum_{i=1}^{v-1} \frac{1}{\mu_{i,d^*}},
\end{aligned}$$

since $\mu_{i,d^*} = \frac{v\lambda}{k}, i = 1, \ldots, v-1$, and since (12), (13) hold. Thus (10) is satisfied and d^* is A-optimal.

In order to establish the D-optimalty of d^*, we must show that for any $d \in \mathcal{D}(v, b, k)$,

$$\prod_{i=1}^{v-1} \mu_{i,d^*} \geq \prod_{i=1}^{v-1} \mu_{i,d}.$$

The proof is similar to that of the A-optimality of d^*, except that we now use the arithmetic mean–geometric mean inequality. That completes the proof. □

5.4 Exercises

1. Consider the linear model $E(\mathbf{y}) = \mathbf{A}\boldsymbol{\beta}, \quad D(\mathbf{y}) = \sigma^2\mathbf{I}$, where

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

Obtain the coefficient matrix in the reduced normal equations for $\beta_3, \beta_4, \beta_5$. Hence determine the estimable functions of $\beta_3, \beta_4, \beta_5$.

2. A *randomized block design* is a design in which every treatment appears once in each block. Consider a randomized block design with b blocks and v treatments. Obtain the variance of the BLUE of $\sum_{i=1}^{v} \ell_i \tau_i$, where $\tau_1, \ldots, \tau_v$ are the treatment effects, $\sum_{i=1}^{v} \ell_i = 0$, and $\sum_{i=1}^{v} \ell_i^2 = 1$.

3. Consider the following design, where each column represents a block. Write down the C-matrix of the design and find its rank. Is the design connected?

$$\begin{matrix} 1 & 2 & 1 & 2 & 3 & 1 \\ 2 & 4 & 4 & 7 & 7 & 4 \\ 3 & 5 & 6 & 8 & 9 & 10 \end{matrix}$$

4. Consider a design with v treatments, block size k, and replication numbers $r_1, \dots, r_v$. Show that $\lambda_1(\mathbf{C})$, the largest eigenvalue of the C-matrix, satisfies $\lambda_1(\mathbf{C}) \geq \frac{k-1}{k} \max_i r_i$.

5. Let d be a binary design in $\mathcal{D}(v, b, k)$. Suppose the treatment set can be partitioned into p groups, where p divides v and $\frac{v}{p} = m$, say, such that the following conditions are satisfied: (i) Each group has m treatments (ii). Two treatments from the same group occur together in α blocks, whereas two treatments from different groups occur together in β blocks. Find the C-matrix of the design and determine its eigenvalues.

6. Let $\mathbf{A}$ be an $n \times n$ matrix with every row sum and column sum equal to zero. Show that the cofactors of $\mathbf{A}$ are all equal.

7. For any design d, show that the cofactors of $\mathbf{C_d}$ are all equal and their common value is at most $\prod_{i=1}^{v-1} \mu_{i,d}$.

8. Let d be a binary design with v treatments $\{1, 2, \dots, v\}$ and with block size k. For $1 \leq i < j \leq v$, let e^{ij} denote the $v \times 1$ column vector with $1, -1$ at the i, j coordinates, respectively, and zeros elsewhere. Suppose treatements i, j occur together in α_{ij} blocks in d. Let $\mathbf{Q}$ be the matrix in which e^{ij} appears as a column α_{ij} times, $1 \leq i < j \leq v$. Show that $k\mathbf{C_d} = \mathbf{QQ}'$. Deduce the fact that $\mathbf{C_d}$ is positive semidefinite.

9. Let $\mathbf{V}$ be an $m \times m$ positive definite matrix. Prove that the volume of the ellipsoid

$$\{\mathbf{y} \in R^m : \mathbf{y}'\mathbf{V}\mathbf{y} \leq 1\}$$

in R^m is

$$\frac{1}{|\mathbf{V}|} \times \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n+2}{2}\right)}.$$

10. Consider the following designs d_1, d_2 in $\mathcal{D}(5, 5, 3)$. Which of the two designs would you prefer according to the criteria of E-, A-, and D-optimality?

$$\begin{array}{ccccccccccccc} & 1 & 2 & 3 & 4 & 5 & & 1 & 2 & 3 & 4 & 5 \\ d_1: & 5 & 1 & 2 & 3 & 4 & d_2: & 2 & 1 & 1 & 3 & 3 \\ & 4 & 5 & 1 & 2 & 3 & & 4 & 5 & 2 & 5 & 1 \end{array}$$

11. Let d be a design and let d' be obtained by deleting some blocks of d. Show that d is better than d' according to the E-, A-, and D-optimality criteria.

12. Let ϕ denote a real-valued function defined on the set of vectors in R^{v-1} with nonnegative coordinates such that

(i) $\phi(x_1, \dots, x_{v-1}) = \phi(x_{\sigma(1)}, \dots, x_{\sigma(v-1)})$ for any permutation σ.
(ii) $\phi(x_1, \dots, x_{v-1}) \leq \phi(\overline{x}, \dots, \overline{x})$.
(iii) $\phi(x, \dots, x) \leq \phi(y, \dots, y)$ if $x \leq y$.

Call a design $d \in \mathcal{D}(v, b, k)$ "ϕ-optimal" if it minimizes

$$\phi(\mu_{1,d}, \dots, \mu_{v-1,d})$$

over $\mathcal{D}(v, b, k)$. Prove the following:

(i) If $d^* \in \mathcal{D}(v, b, k)$ is a BIBD, then d^* is ϕ-optimal.
(ii) The criteria of E-, A-, and D-optimality can be seen as special cases of ϕ-optimality.

13. If d^* is a BIBD with parameters v, b, k, r, λ, then show that $k^r \geq v^\lambda$.

5.5 Hints and Solutions

4. Hint: First show that the ith diagonal entry of the C-matrix cannot be less than $\frac{k-1}{k} r_i$. The result follows, since the largest eigenvalue of a symmetric matrix cannot be less than any diagonal element.

5. Answer: The eigenvalues of k times the C-matrix are 0 with multiplicity 1, $m(\alpha-\beta)+\beta v$ with multiplicity $v-p$, and $m\beta(p-1)$ with multiplicity $p-1$.

7. Since the row and column sums of $\mathbf{C_d}$ are all zero, the first part follows from the previous exercise. Let $\mathbf{B}$ be the matrix obtained by deleting the first row and column of $\mathbf{C_d}$. Let $\lambda_1 \leq \cdots \leq \lambda_{v-1}$ be the eigenvalues of $\mathbf{B}$. By the interlacing principle we see that $\lambda_i \leq \mu_{i,d}, i = 1, \ldots, v-1$. Thus $|\mathbf{B}| = \prod_{i=1}^{v-1} \lambda_i \leq \prod_{i=1}^{v-1} \mu_{i,d}$.

10. The C-matrices are given by

$$3\mathbf{C_{d_1}} = \begin{bmatrix} 6 & -2 & -1 & -1 & -2 \\ -2 & 6 & -2 & -1 & -1 \\ -1 & -2 & 6 & -2 & -1 \\ -1 & -1 & -2 & 6 & -2 \\ -2 & -1 & -1 & -2 & 6 \end{bmatrix}, \quad 3\mathbf{C_{d_2}} = \begin{bmatrix} 8 & -3 & -2 & -1 & -2 \\ -3 & 6 & -1 & -1 & -1 \\ -2 & -1 & 6 & -1 & -2 \\ -1 & -1 & -1 & 4 & -1 \\ -2 & -1 & -2 & -1 & 6 \end{bmatrix}.$$

The eigenvalues (rounded to four decimal places) are

$$\mathbf{C_{d_1}} : 0, \quad 2.1273, \quad 2.1273, \quad 2.8727, \quad 2.8727,$$
$$\mathbf{C_{d_2}} : 0, \quad 1.6667, \quad 2.1461, \quad 2.6667, \quad 3.5205.$$

It can be verified that d_1 is better than d_2 according to each of the E-, A-, and D-optimality criteria.

11. Hint: First show that the C-matrices of d, d' satisfy $\mathbf{C_d} \geq \mathbf{C_{d'}}$. Now use **2.6** of Chapter 4.

12. (i) Let $d \in \mathcal{D}(v, b, k)$ and let the corresponding C-matrix $\mathbf{C_d}$ have eigenvalues $0 \leq \mu_{1,d} \leq \cdots \leq \mu_{v-1,d}$. Suppose there exists a BIBD $d^* \in \mathcal{D}(v, b, k)$. By **3.4**, all the nonzero eigenvalues of the corresponding C-matrix equal $\frac{v\lambda}{k}$. By **3.2** we have trace$\mathbf{C_d} \leq b(k-1) = \frac{\lambda v(v-1)}{k}$, using (9), (10). Now, using the properties of ϕ, we get

$$\phi(\mu_{1,d}, \ldots, \mu_{v-1,d}) \leq \phi\left(\frac{\text{trace}\mathbf{C_d}}{v-1}, \ldots, \frac{\text{trace}\mathbf{C_d}}{v-1}\right) \leq \phi\left(\frac{\lambda v}{k}, \ldots, \frac{\lambda v}{k}\right),$$

and the result follows.

(ii) For E-, A-, and D-optimality we set $-\phi(\mu_{1,d}, \ldots, \mu_{v-1,d})$ to be $\mu_{1,d}$, $\sum_{i=1}^{v-1} \frac{1}{\mu_{i,d}}$ and $\prod_{i=1}^{v-1} \mu_{i,d}$, respectively. It can be verified that each of these functions satisfies conditions (i),(ii),(iii).

This page intentionally left blank

6

Rank Additivity

6.1 Preliminaries

If $\mathbf{A}, \mathbf{B}$ are $m \times n$ matrices, then since $\mathbf{B} = \mathbf{A} + (\mathbf{B} - \mathbf{A})$, we have

$$R(\mathbf{B}) \leq R(\mathbf{A}) + R(\mathbf{B} - \mathbf{A}). \tag{1}$$

When does equality hold in (1)? It turns out that there exist a number of different equivalent necessary and sufficient conditions for this to happen. In this chapter we study several such conditions and their interrelationships. We then illustrate an application to general linear models.

1.1. *Let* $\mathbf{A}, \mathbf{B}$ *be nonzero matrices. Then* $\mathbf{A}\mathbf{C}^{-}\mathbf{B}$ *is invariant under the choice of the g-inverse if and only if* $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{C})$ *and* $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{C})$.

PROOF. The *if* part was seen earlier in Chapter 2, Section 3. We now prove the converse. Thus suppose that $\mathbf{A}\mathbf{C}^{-}\mathbf{B}$ is invariant under the choice of the g-inverse and suppose $\mathcal{C}(\mathbf{B})$ is not contained in $\mathcal{C}(\mathbf{C})$. Then

$$\mathbf{D} = (\mathbf{I} - \mathbf{C}\mathbf{C}^{-})\mathbf{B}$$

is a nonzero matrix. Let $\mathbf{A} = \mathbf{X}\mathbf{Y}$, $\mathbf{D} = \mathbf{P}\mathbf{Q}$ be rank factorizations. Let

$$\mathbf{C}^{=} = \mathbf{C}^{-} + \mathbf{Y}_{\mathbf{r}}^{-}\mathbf{P}_{\ell}^{-}(\mathbf{I} - \mathbf{C}\mathbf{C}^{-}),$$

where $\mathbf{Y}_{\mathbf{r}}^{-}$, $\mathbf{P}_{\ell}^{-}$ are respectively a right inverse of $\mathbf{Y}$ and a left inverse of $\mathbf{P}$. Clearly, $\mathbf{C}^{=}$ is also a g-inverse of $\mathbf{C}$, and

$$\mathbf{A}\mathbf{C}^{=}\mathbf{B} = \mathbf{A}\mathbf{C}^{-}\mathbf{B} + \mathbf{A}\mathbf{Y}_{\mathbf{r}}^{-}\mathbf{P}_{\ell}^{-}\mathbf{D} = \mathbf{A}\mathbf{C}^{-}\mathbf{B} + \mathbf{X}\mathbf{Q}.$$

Since $\mathbf{X}$ admits a left inverse and $\mathbf{Q}$ admits a right inverse, $\mathbf{XQ}$ is a nonzero matrix, and we get a contradiction. Thus $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{C})$. Similarly, we can show that $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{C})$. $\square$

Suppose $\mathbf{A}, \mathbf{B}$ are $n \times n$ positive definite matrices. In the context of parallel connections of two electrical networks, the following sum, called the *parallel sum* of $\mathbf{A}, \mathbf{B}$, is defined:

$$P(\mathbf{A}, \mathbf{B}) = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}.$$

Note that

$$(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B},$$

and this suggests the following definition. Call matrices $\mathbf{A}, \mathbf{B}$ (not necessarily positive definite) *parallel summable* if

$$\mathbf{A}(\mathbf{A} + \mathbf{B})^{-}\mathbf{B} \tag{2}$$

is invariant under the choice of g-inverse, in which case call (2) the parallel sum of $\mathbf{A}, \mathbf{B}$, denoted by $P(\mathbf{A}, \mathbf{B})$. As we shall see, the concept of parallel sum is closely linked with rank additivity.

6.2 Characterizations of Rank Additivity

As usual, $\mathbf{A}^{-}$ will denote an arbitrary g-inverse of $\mathbf{A}$. We say that two vector spaces are virtually disjoint if they have only the zero vector in common.

2.1. *Let* $\mathbf{A}, \mathbf{B}$ *be* $m \times n$ *matrices. Then the following conditions are equivalent:*

(i) $R(\mathbf{B}) = R(\mathbf{A}) + R(\mathbf{B} - \mathbf{A})$.
(ii) $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B} - \mathbf{A}) = \{\mathbf{0}\}$, $\mathcal{R}(\mathbf{A}) \cap \mathcal{R}(\mathbf{B} - \mathbf{A}) = \{\mathbf{0}\}$.
(iii) *Every* $\mathbf{B}^{-}$ *is a g-inverse of* $\mathbf{A}$.
(iv) *There exists* $\mathbf{A}^{-}$ *such that* $\mathbf{A}^{-}\mathbf{A} = \mathbf{A}^{-}\mathbf{B}$, $\mathbf{A}\mathbf{A}^{-} = \mathbf{B}\mathbf{A}^{-}$.
(v) *There exists* $\mathbf{A}^{-}$ *such that* $\mathbf{A}\mathbf{A}^{-}\mathbf{B} = \mathbf{B}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$.
(vi) $\mathbf{A}, \mathbf{B} - \mathbf{A}$ *are parallel summable and* $P(\mathbf{A}, \mathbf{B} - \mathbf{A}) = \mathbf{0}$.
(vii) $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$ *and there exists* $\mathbf{A}^{-}$ *such that* $\mathbf{A}^{-}\mathbf{B} = \mathbf{A}^{-}\mathbf{A}$.
(viii) $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{B})$ *and there exists* $\mathbf{A}^{-}$ *such that* $\mathbf{B}\mathbf{A}^{-} = \mathbf{A}\mathbf{A}^{-}$.
(ix) *There exist g-inverses* $\mathbf{B}^{-}, \mathbf{B}^{*}, \mathbf{B}^{=}$ *such that* $\mathbf{A} = \mathbf{B}\mathbf{B}^{-}\mathbf{A} = \mathbf{A}\mathbf{B}^{*}\mathbf{B} = \mathbf{A}\mathbf{B}^{=}\mathbf{A}$.
(x) *For any* $\mathbf{B}^{-}$, $\mathbf{A} = \mathbf{B}\mathbf{B}^{-}\mathbf{A} = \mathbf{A}\mathbf{B}^{-}\mathbf{B} = \mathbf{A}\mathbf{B}^{-}\mathbf{A}$.
(xi) *There exist* $\mathbf{K}, \mathbf{L}$, *at least one of which is idempotent, such that* $\mathbf{A} = \mathbf{K}\mathbf{B} = \mathbf{B}\mathbf{L}$.
(xii) *There exists a* $\mathbf{B}^{-}$ *that is a g-inverse of both* $\mathbf{A}$ *and* $\mathbf{B} - \mathbf{A}$.
(xiii) *Every* $\mathbf{B}^{-}$ *is a g-inverse of both* $\mathbf{A}$ *and* $\mathbf{B} - \mathbf{A}$.
(xiv) *There exist nonsingular matrices* $\mathbf{P}, \mathbf{Q}$ *such that*

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{P}(\mathbf{B} - \mathbf{A})\mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y} \end{bmatrix}.$$

(xv) *There exist g-inverses* $\mathbf{A}^-$, $\mathbf{A}^=$ *such that* $\mathbf{A}^-\mathbf{A} = \mathbf{A}^-\mathbf{B}$, $\mathbf{A}\mathbf{A}^= = \mathbf{B}\mathbf{A}^=$.
(xvi) *There exist g-inverses* $\mathbf{A}^-$, $\mathbf{A}^=$ *such that* $\mathbf{A}\mathbf{A}^-\mathbf{B} = \mathbf{B}\mathbf{A}^=\mathbf{A} = \mathbf{A}$.

PROOF. We will assume that both $\mathbf{A}$ and $\mathbf{B}-\mathbf{A}$ are nonzero matrices, for otherwise, the result is trivial. Let $\mathbf{A} = \mathbf{X}\mathbf{Y}$, $\mathbf{B}-\mathbf{A} = \mathbf{U}\mathbf{V}$ be rank factorizations.

(i) $\Rightarrow$ (iii): We have

$$\mathbf{B} = \mathbf{A} + (\mathbf{B}-\mathbf{A}) = [\mathbf{X}, \mathbf{U}]\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix}. \tag{3}$$

Since (i) holds, (3) is a rank factorization of $\mathbf{B}$. Now for any $\mathbf{B}^-$,

$$[\mathbf{X}, \mathbf{U}]\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix}\mathbf{B}^-[\mathbf{X}, \mathbf{U}]\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix} = [\mathbf{X}, \mathbf{U}]\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix}.$$

Hence

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix}\mathbf{B}^-[\mathbf{X}, \mathbf{U}] = \mathbf{I_r}.$$

Then

$$\begin{bmatrix} \mathbf{Y}\mathbf{B}^-\mathbf{X} & \mathbf{Y}\mathbf{B}^-\mathbf{U} \\ \mathbf{V}\mathbf{B}^-\mathbf{X} & \mathbf{V}\mathbf{B}^-\mathbf{U} \end{bmatrix} = \mathbf{I_r},$$

and therefore

$$\mathbf{Y}\mathbf{B}^-\mathbf{X} = \mathbf{I}, \quad \mathbf{Y}\mathbf{B}^-\mathbf{U} = \mathbf{0}, \quad \mathbf{V}\mathbf{B}^-\mathbf{X} = \mathbf{0}.$$

Since $\mathbf{A} = \mathbf{X}\mathbf{Y}$, it follows that $\mathbf{A}\mathbf{B}^-\mathbf{A} = \mathbf{A}$. We also conclude that

$$(\mathbf{B}-\mathbf{A})\mathbf{B}^-\mathbf{A} = \mathbf{0} \quad \text{and} \quad \mathbf{A}\mathbf{B}^-(\mathbf{B}-\mathbf{A}) = \mathbf{0}.$$

and thus (i) $\Rightarrow$ (ix),(x). Furthermore,

$$(\mathbf{B}-\mathbf{A})\mathbf{B}^-(\mathbf{B}-\mathbf{A}) = (\mathbf{B}-\mathbf{A})\mathbf{B}^-\mathbf{B} = \mathbf{B}-\mathbf{A},$$

and therefore (i) $\Rightarrow$ (xii),(xiii) as well.

(iii) $\Rightarrow$ (ii): For any $\mathbf{B}^-$, we have $\mathbf{A}\mathbf{B}^-\mathbf{A} = \mathbf{A}$, and thus $\mathbf{A}\mathbf{B}^-\mathbf{A}$ is invariant under the choice of g-inverse. Thus by **1.1**, $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$, $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{B})$. In particular, $\mathbf{A} = \mathbf{U}\mathbf{B}$ for some $\mathbf{U}$, and therefore $\mathbf{A}\mathbf{B}^-\mathbf{B} = \mathbf{A}$. Suppose

$$\mathbf{x} \in \mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}-\mathbf{A}). \tag{4}$$

Then

$$\mathbf{x} = \mathbf{A}\mathbf{y} = (\mathbf{B}-\mathbf{A})\mathbf{z} \tag{5}$$

for some $\mathbf{y}$, $\mathbf{z}$. Then

$$\begin{aligned} \mathbf{x} &= \mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{B}^-\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{B}^-(\mathbf{B}-\mathbf{A})\mathbf{z} \\ &= \mathbf{A}\mathbf{B}^-\mathbf{B}\mathbf{z} - \mathbf{A}\mathbf{B}^-\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{z} - \mathbf{A}\mathbf{z} = \mathbf{0}. \end{aligned}$$

Thus $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}-\mathbf{A}) = \{\mathbf{0}\}$. Similarly, it can be shown that the row spaces of $\mathbf{A}$ and $\mathbf{B}-\mathbf{A}$ are also virtually disjoint.

(ii) $\Rightarrow$ (i): We make use of (3). Since (ii) holds, $[\mathbf{X}, \mathbf{U}]$ must have full column rank, and $\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix}$ must have full row rank. It follows that the rank of $\mathbf{B}$ must equal the number of columns in $\mathbf{X}$, $\mathbf{U}$, and hence (i) is true.

(i) $\Rightarrow$ (iv): Since (i) $\Rightarrow$ (iii),(x), for any $\mathbf{B}^-$, we have $\mathbf{AB}^-\mathbf{B} = \mathbf{AB}^-\mathbf{A} = \mathbf{A}$. Since $\mathbf{B}^-\mathbf{AB}^-$ is a g-inverse of $\mathbf{A}$ as well, setting $\mathbf{A}^- = \mathbf{B}^-\mathbf{AB}^-$ we see that $\mathbf{A}^-\mathbf{B} = \mathbf{A}^-\mathbf{A}$. Similarly, using $(\mathbf{B}-\mathbf{A})\mathbf{B}^-\mathbf{A} = \mathbf{0}$ it follows that $\mathbf{AA}^- = \mathbf{BA}^-$.

The proof of (i) $\Rightarrow$ (v) is similar.

(iv) $\Rightarrow$ (ii): Suppose (4) holds. Then (5) is true for some $\mathbf{y}$, $\mathbf{z}$. Now,

$$\begin{aligned} \mathbf{x} &= \mathbf{Ay} = \mathbf{AA}^-\mathbf{Ay} = \mathbf{AA}^-(\mathbf{B}-\mathbf{A})\mathbf{z} \\ &= \mathbf{A}^-\mathbf{Bz} - \mathbf{Az} = \mathbf{AA}^-\mathbf{Az} - \mathbf{Az} = \mathbf{0}. \end{aligned}$$

Thus $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}-\mathbf{A}) = \{\mathbf{0}\}$. Similarly, it can be shown that the row spaces of $\mathbf{A}$ and $\mathbf{B}-\mathbf{A}$ are also virtually disjoint.

The proof of (v) $\Rightarrow$ (ii) is similar.

(i) $\Rightarrow$ (vi): This follows from (i) $\Rightarrow$ (x).

(vi) $\Rightarrow$ (iii): By **1.1**, $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{B})$. Thus for any $\mathbf{B}^-$, $\mathbf{AB}^-\mathbf{B} = \mathbf{A}$. Now $\mathbf{AB}^-(\mathbf{B}-\mathbf{A}) = \mathbf{0}$ implies that $\mathbf{AB}^-\mathbf{A} = \mathbf{A}$, and thus (iii) holds.

(iv) $\Rightarrow$ (vii): We must show only that $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$. Let $\mathbf{y} \in \mathcal{C}(\mathbf{A})$, so that $\mathbf{y} = \mathbf{Ax}$ for some $\mathbf{x}$. Then $\mathbf{y} = \mathbf{AA}^-\mathbf{Ax} = \mathbf{BA}^-\mathbf{Ax} \in \mathcal{C}(\mathbf{B})$.

(vii) $\Rightarrow$ (iii): Since $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$, then $\mathbf{BB}^-\mathbf{A} = \mathbf{A}$. Now for any $\mathbf{B}^-$,

$$\mathbf{AB}^-\mathbf{A} = \mathbf{AA}^-\mathbf{AB}^-\mathbf{A} = \mathbf{AA}^-\mathbf{BB}^-\mathbf{A} = \mathbf{AA}^-\mathbf{A} = \mathbf{A}.$$

The proofs of (iv) $\Rightarrow$ (viii) and of (viii) $\Rightarrow$ (iii) are similar. We have already seen that (i) $\Rightarrow$ (ix),(x).

(ix) $\Rightarrow$ (ii): Suppose (4) holds. Then (5) is true for some $\mathbf{y}$, $\mathbf{z}$. Now

$$\begin{aligned} \mathbf{x} &= \mathbf{Ay} = \mathbf{AB}^=\mathbf{Ay} = \mathbf{AB}^=(\mathbf{B}-\mathbf{A})\mathbf{z} \\ &= \mathbf{AB}^=\mathbf{Bz} - \mathbf{AB}^=\mathbf{Az} \\ &= \mathbf{AB}^*\mathbf{BB}^=\mathbf{Bz} - \mathbf{Az} = \mathbf{AB}^*\mathbf{Bz} - \mathbf{Az} = \mathbf{Az} - \mathbf{Az} = \mathbf{0}. \end{aligned}$$

Thus $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}-\mathbf{A}) = \{\mathbf{0}\}$. Similarly it can be shown that the row spaces of $\mathbf{A}$ and $\mathbf{B}-\mathbf{A}$ are also virtually disjoint.

Clearly, (x) $\Rightarrow$ (ix).

(x) $\Rightarrow$ (xi): Set $\mathbf{K} = \mathbf{AB}^-$, $\mathbf{L} = \mathbf{B}^-\mathbf{A}$ for some $\mathbf{B}^-$.

(xi) $\Rightarrow$ (x): Suppose $\mathbf{L}$ is idempotent. We have

$$\begin{aligned} \mathbf{BB}^-\mathbf{A} &= \mathbf{BB}^-\mathbf{BL} = \mathbf{BL} = \mathbf{A}, \\ \mathbf{AB}^-\mathbf{B} &= \mathbf{KBB}^-\mathbf{B} = \mathbf{KB} = \mathbf{A}, \end{aligned}$$

and

$$\mathbf{AB}^-\mathbf{A} = \mathbf{KBB}^-\mathbf{BL} = \mathbf{KBL} = \mathbf{AL} = \mathbf{BLL} = \mathbf{BL} = \mathbf{A}.$$

If $\mathbf{K}$ is idempotent, then the proof is similar.

We have already seen that (i) $\Rightarrow$ (xii),(xiii).

(xii) $\Rightarrow$ (i): Since $\mathbf{BB}^-$, $\mathbf{AB}^-$, $(\mathbf{B}-\mathbf{A})\mathbf{B}^-$ are idempotent,

$$\begin{aligned} R(\mathbf{B}) &= R(\mathbf{BB}^-) \\ &= \operatorname{trace}\mathbf{BB}^- \\ &= \operatorname{trace}\mathbf{AB}^- + \operatorname{trace}(\mathbf{B}-\mathbf{A})\mathbf{B}^- \\ &= R(\mathbf{AB}^-) + R((\mathbf{B}-\mathbf{A})\mathbf{B}^-) \\ &= R(\mathbf{A}) + R(\mathbf{B}-\mathbf{A}). \end{aligned}$$

Clearly, (xiii) $\Rightarrow$ (xii).

The proof of (xiv) $\Rightarrow$ (i) is easy. We now prove (ii) $\Rightarrow$ (xiv). By **7.3** of Chapter 1 we may assume, without loss of generality, that

$$\mathbf{A} = \begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where r is the rank of $\mathbf{A}$. Let $R(\mathbf{B}-\mathbf{A}) = s$ and let

$$\mathbf{B}-\mathbf{A} = \mathbf{UV} = \begin{bmatrix} \mathbf{U_1} \\ \mathbf{U_2} \end{bmatrix} [\mathbf{V_1}, \mathbf{V_2}]$$

be a rank factorization of $\mathbf{B}-\mathbf{A}$, where $\mathbf{U_1}$ is $r \times s$ and $\mathbf{V_1}$ is $s \times r$. We first claim that $R(\mathbf{U_2}) = s$. Suppose $R(\mathbf{U_2}) < s$. Then the null space of $\mathbf{U_2}$ contains a nonzero vector, say $\mathbf{x}$. If $\mathbf{U_1x} \neq \mathbf{0}$, then

$$\begin{bmatrix} \mathbf{U_1} \\ \mathbf{U_2} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{U_1x} \\ \mathbf{0} \end{bmatrix}$$

is a nonzero vector in $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}-\mathbf{A})$, which is not possible. Thus $\mathbf{U_1x} = \mathbf{0}$. Then $\mathbf{Ux} = \mathbf{0}$, which contradicts the fact that $\mathbf{U}$ has full column rank. Thus the claim is proved. It follows that $\mathbf{U_1} = \mathbf{MU_2}$ for some $\mathbf{M}$. Similarly, $\mathbf{V_1} = \mathbf{V_2N}$ for some $\mathbf{N}$. Set

$$\mathbf{P} = \begin{bmatrix} \mathbf{I_r} & -\mathbf{M} \\ \mathbf{0} & \mathbf{I_{s-r}} \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{I_r} & \mathbf{0} \\ -\mathbf{N} & \mathbf{I_{s-r}} \end{bmatrix}.$$

Then $\mathbf{PAQ} = \mathbf{A}$ and

$$\mathbf{P}(\mathbf{B}-\mathbf{A})\mathbf{Q} = \mathbf{P} \begin{bmatrix} \mathbf{U_1} \\ \mathbf{U_2} \end{bmatrix} [\mathbf{V_1}, \mathbf{V_2}]\mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U_2V_2} \end{bmatrix}.$$

This proves (ii) $\Rightarrow$ (xiv).

The treatment of cases (xv), (xvi) is similar to that of (iv), (v), respectively.

We have thus proved the following implications (or a hint is provided toward the proof). It is left to the reader to verify that all the other implications then follow.

$$
\begin{aligned}
\text{(i)} &\Rightarrow \text{(iii)-(vi), (ix), (x), (xii), (xiii), (xv), (xvi);}\\
\text{(ii)} &\Rightarrow \text{(i), (xiv);} \quad \text{(iii)} \Rightarrow \text{(ii);}\\
\text{(iv)} &\Rightarrow \text{(ii), (vii), (viii);} \quad \text{(v)} \Rightarrow \text{(ii);} \quad \text{(vi)} \Rightarrow \text{(iii);}\\
\text{(vii)} &\Rightarrow \text{(iii);} \quad \text{(viii)} \Rightarrow \text{(iii);} \quad \text{(ix)} \Rightarrow \text{(ii);}\\
\text{(x)} &\Rightarrow \text{(ix), (xi);} \quad \text{(xi)} \Rightarrow \text{(x);} \quad \text{(xii)} \Rightarrow \text{(i);}\\
\text{(xiii)} &\Rightarrow \text{(xii);} \quad \text{(xiv)} \Rightarrow \text{(i);} \quad \text{(xv)} \Rightarrow \text{(vii);} \quad \text{(xvi)} \Rightarrow \text{(ii).}
\end{aligned}
$$

That completes the proof. □

6.3 General Linear Model

3.1. *Let* $\mathbf{X}$ *be an* $n \times m$ *matrix and let* $\mathbf{V}$ *be a positive semidefinite* $n \times n$ *matrix. Then*

$$R\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} = R\begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} + R\begin{bmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = R[\mathbf{V}, \mathbf{X}] + R(\mathbf{X}).$$

PROOF. By **2.1** the result will be proved if we show that the column spaces of

$$\begin{bmatrix} \mathbf{V} \\ \mathbf{X}' \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{X} \\ \mathbf{0} \end{bmatrix}$$

are virtually disjoint. Suppose there exists a nonzero vector in the intersection of the two column spaces. Then there exist vectors $\mathbf{a}$, $\mathbf{b}$ such that

$$\mathbf{Va} = \mathbf{Xb} \neq \mathbf{0} \tag{6}$$

and $\mathbf{X}'\mathbf{a} = \mathbf{0}$. Then $\mathbf{a}'\mathbf{V} = \mathbf{b}'\mathbf{X}'$, and hence $\mathbf{a}'\mathbf{Va} = \mathbf{b}'\mathbf{X}'\mathbf{a} = \mathbf{0}$. Since $\mathbf{V}$ is positive semidefinite, it follows that $\mathbf{Va} = \mathbf{0}$, and this contradicts (6). □

In the remainder of this section we assume that $\mathbf{X}$, $\mathbf{V}$ are as in **3.1** and that

$$\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}^{-} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} \tag{7}$$

is one possible g-inverse.

3.2. *The following assertions are true:*

(i) $\mathbf{X}\mathbf{C}_2'\mathbf{X} = \mathbf{X}$, $\mathbf{X}\mathbf{C}_3\mathbf{X} = \mathbf{X}$.
(ii) $\mathbf{X}\mathbf{C}_4\mathbf{X}' = \mathbf{X}\mathbf{C}_4'\mathbf{X}' = \mathbf{V}\mathbf{C}_3'\mathbf{X}' = \mathbf{X}\mathbf{C}_3\mathbf{V} = \mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{X}\mathbf{C}_2'\mathbf{V}$.
(iii) $\mathbf{X}'\mathbf{C}_1\mathbf{X}$, $\mathbf{X}'\mathbf{C}_1\mathbf{V}$ *and* $\mathbf{V}\mathbf{C}_1\mathbf{X}$ *are zero matrices.*
(iv) $\mathbf{V}\mathbf{C}_1\mathbf{V}\mathbf{C}_1\mathbf{V} = \mathbf{V}\mathbf{C}_1\mathbf{V} = \mathbf{V}\mathbf{C}_1'\mathbf{V}\mathbf{C}_1\mathbf{V} = \mathbf{V}\mathbf{C}_1'\mathbf{V}$.
(v) *trace* $\mathbf{V}\mathbf{C}_1 = R[\mathbf{V}, \mathbf{X}] - R(\mathbf{X}) =$ *trace* $\mathbf{V}\mathbf{C}_1'$.

PROOF. We have

$$\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} \begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}. \tag{8}$$

This gives, after simplification,

$$\mathbf{V}\mathbf{C}_1\mathbf{V} + \mathbf{X}\mathbf{C}_3\mathbf{V} + \mathbf{V}\mathbf{C}_2\mathbf{X}' - \mathbf{X}\mathbf{C}_4\mathbf{X}' = \mathbf{V}, \tag{9}$$

$$\mathbf{V}\mathbf{C}_1\mathbf{X} + \mathbf{X}\mathbf{C}_3\mathbf{X} = \mathbf{X}, \quad \mathbf{X}'\mathbf{C}_1\mathbf{V} + \mathbf{X}'\mathbf{C}_2\mathbf{X}' = \mathbf{X}', \tag{10}$$

$$\mathbf{X}'\mathbf{C}_1\mathbf{X} = \mathbf{0}. \tag{11}$$

By **3.1** and **2.1** we see that the matrix on the right-hand side of (7) is a g-inverse of

$$\begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

as well. Thus we can write two more equations similar to (8), which give, after simplification,

$$\mathbf{V}\mathbf{C}_1\mathbf{V} + \mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{V}, \quad \mathbf{X}'\mathbf{C}_1\mathbf{V} + \mathbf{X}'\mathbf{C}_2\mathbf{X}' = \mathbf{X}', \tag{12}$$

and

$$\mathbf{X}\mathbf{C}_3\mathbf{X} = \mathbf{X}. \tag{13}$$

Since

$$\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}$$

is symmetric,

$$\begin{bmatrix} \mathbf{C}_1' & \mathbf{C}_3' \\ \mathbf{C}_2' & -\mathbf{C}_4' \end{bmatrix}$$

is also a possible g-inverse. Therefore, we can replace $\mathbf{C}_1$ by $\mathbf{C}_1'$, $\mathbf{C}_2$ by $\mathbf{C}_3'$, $\mathbf{C}_3$ by $\mathbf{C}_2'$, and $\mathbf{C}_4$ by $\mathbf{C}_4'$ in (9)–(13).

Assertions (i),(ii),(iii) follow from these equations by trivial manipulations. We will prove (iv),(v). We have

$$\mathbf{V}\mathbf{C}_1\mathbf{V} + \mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{V}, \quad \mathbf{X}'\mathbf{C}_1\mathbf{V} = \mathbf{0}. \tag{14}$$

Thus

$$\mathbf{V}\mathbf{C}_1\mathbf{V}\mathbf{C}_1\mathbf{V} + \mathbf{V}\mathbf{C}_1\mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{V}\mathbf{C}_1\mathbf{V},$$

and since $\mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{X}\mathbf{C}_2'\mathbf{V}$ and $\mathbf{V}\mathbf{C}_1\mathbf{X} = \mathbf{0}$, we get $\mathbf{V}\mathbf{C}_1\mathbf{V}\mathbf{C}_1\mathbf{V} = \mathbf{V}\mathbf{C}_1\mathbf{V}$. Equations (14) also imply

$$\mathbf{V}\mathbf{C}_1'\mathbf{V}\mathbf{C}_1\mathbf{V} + \mathbf{V}\mathbf{C}_1'\mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{V}\mathbf{C}_1'\mathbf{V},$$

and again $\mathbf{V}\mathbf{C}_2\mathbf{X}' = \mathbf{X}\mathbf{C}_2'\mathbf{V}$, $\mathbf{V}\mathbf{C}_1'\mathbf{X} = \mathbf{0}$ gives

$$\mathbf{V}\mathbf{C}_1'\mathbf{V}\mathbf{C}_1\mathbf{V} = \mathbf{V}\mathbf{C}_1'\mathbf{V}. \tag{15}$$

Taking the transpose on both sides of (15), we get $\mathbf{V}\mathbf{C}_1'\mathbf{V}\mathbf{C}_1\mathbf{V} = \mathbf{V}\mathbf{C}_1\mathbf{V}$, and (iv) is proved. Now,

$$\begin{aligned} R\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} &= R\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} \\ &= \operatorname{trace}\begin{bmatrix} \mathbf{V}\mathbf{C}_1 + \mathbf{X}\mathbf{C}_3 & \mathbf{V}\mathbf{C}_2 - \mathbf{X}\mathbf{C}_4 \\ \mathbf{X}'\mathbf{C}_1 & \mathbf{X}'\mathbf{C}_2 \end{bmatrix} \\ &= \operatorname{trace}(\mathbf{V}\mathbf{C}_1 + \mathbf{X}\mathbf{C}_3) + \operatorname{trace}\mathbf{X}'\mathbf{C}_2 \\ &= \operatorname{trace}\mathbf{V}\mathbf{C}_1 + R(\mathbf{X}) + R(\mathbf{X}'), \end{aligned} \tag{16}$$

where we have made use of (i). Also, by **2.1**,

$$R\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} = R[\mathbf{V}, \mathbf{X}] - R(\mathbf{X}). \tag{17}$$

It follows from (16), (17) that

$$\operatorname{trace}\mathbf{V}\mathbf{C}_1 = R[\mathbf{V}, \mathbf{X}] - R(\mathbf{X}).$$

Similarly,

$$\operatorname{trace}\mathbf{V}\mathbf{C}_1' = R[\mathbf{V}, \mathbf{X}] - R(\mathbf{X}),$$

and (v) is proved. □

Consider the linear model $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $D(\mathbf{y}) = \sigma^2\mathbf{V}$, where $\mathbf{y}$ is $n \times 1$, $\mathbf{X}$ is $n \times p$, and $\mathbf{V}$ is a (known) positive semidefinite $p \times p$ matrix. (The setup that we have considered so far is a special case with $\mathbf{V} = \mathbf{I}$.) We do not make any assumptions about the rank of $\mathbf{X}$.

We first remark that if $\mathbf{V}$ is positive definite, then making the transformation $\mathbf{z} = \mathbf{V}^{-1/2}\mathbf{y}$, we get the model $E(\mathbf{z}) = \mathbf{V}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta}$, $D(\mathbf{z}) = \sigma^2\mathbf{I}$, which can be treated by methods developed earlier. When $\mathbf{V}$ is singular, such a simple transformation is not available.

3.3. *The BLUE of an estimable function $\boldsymbol{\ell}'\boldsymbol{\beta}$ is $\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}} = \mathbf{C}_3\mathbf{y}$ or $\widehat{\boldsymbol{\beta}} = \mathbf{C}_2'\mathbf{y}$.*

Proof. Since $\boldsymbol{\ell}'\boldsymbol{\beta}$ is estimable, there exists a linear function $\mathbf{u}'\mathbf{y}$ such that $E(\mathbf{u}'\mathbf{y}) = \boldsymbol{\ell}'\boldsymbol{\beta}$. Thus $\mathbf{u}'\mathbf{X} = \boldsymbol{\ell}'$. If $\widehat{\boldsymbol{\beta}} = \mathbf{C}_3\mathbf{y}$, then

$$\begin{aligned} E(\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}) &= E(\boldsymbol{\ell}'\mathbf{C}_3\mathbf{y}) = \boldsymbol{\ell}'\mathbf{C}_3\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{u}'\mathbf{X}\mathbf{C}_3\mathbf{X}\boldsymbol{\beta} = \mathbf{u}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\ell}'\boldsymbol{\beta}, \end{aligned}$$

since $\mathbf{X}\mathbf{C}_3\mathbf{X} = \mathbf{X}$ by **3.2**. Thus $\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}$ is unbiased for $\boldsymbol{\ell}'\boldsymbol{\beta}$.

Let $\mathbf{w}'\mathbf{y}$ be any other unbiased estimate of $\boldsymbol{\ell}'\boldsymbol{\beta}$. Then $\mathbf{w}'\mathbf{X} = \boldsymbol{\ell}'$. We have

$$\begin{aligned} \frac{1}{\sigma^2}\operatorname{var}(\mathbf{w}'\mathbf{y}) &= \mathbf{w}'\mathbf{V}\mathbf{w} \\ &= (\mathbf{w} - \mathbf{C}_3'\boldsymbol{\ell} + \mathbf{C}_3'\boldsymbol{\ell})'\mathbf{V}(\mathbf{w} - \mathbf{C}_3'\boldsymbol{\ell} + \mathbf{C}_3'\boldsymbol{\ell}) \\ &= (\mathbf{w} - \mathbf{C}_3'\boldsymbol{\ell})'\mathbf{V}(\mathbf{w} - \mathbf{C}_3'\boldsymbol{\ell}) + \boldsymbol{\ell}'\mathbf{C}_3\mathbf{V}\mathbf{C}_3'\boldsymbol{\ell} + 2\boldsymbol{\ell}'\mathbf{C}_3\mathbf{V}(\mathbf{w} - \mathbf{C}_3'\boldsymbol{\ell}). \end{aligned} \tag{18}$$

Observe that

$$\begin{aligned}\boldsymbol{\ell}'\mathbf{C}_3\mathbf{V}(\mathbf{w}-\mathbf{C}_3'\boldsymbol{\ell}) &= 2\mathbf{w}'\mathbf{X}\mathbf{C}_3\mathbf{V}(\mathbf{w}-\mathbf{C}_3'\mathbf{X}'\mathbf{w})\\ &= 2\mathbf{w}'\mathbf{X}\mathbf{C}_3\mathbf{V}(\mathbf{I}-\mathbf{C}_3'\mathbf{X}')\mathbf{w}\\ &= 0,\end{aligned}$$

since by **3.2**,

$$\mathbf{X}\mathbf{C}_3\mathbf{V}\mathbf{C}_3'\mathbf{X}' = \mathbf{X}\mathbf{C}_3\mathbf{X}\mathbf{C}_3\mathbf{V} = \mathbf{X}\mathbf{C}_3\mathbf{V}.$$

Substituting in (18) we get

$$\frac{1}{\sigma^2}\operatorname{var}(\mathbf{w}'\mathbf{y}) \geq \boldsymbol{\ell}'\mathbf{C}_3\mathbf{V}\mathbf{C}_3'\boldsymbol{\ell} = \frac{1}{\sigma^2}\operatorname{var}(\boldsymbol{\ell}'\mathbf{C}_3\mathbf{y}).$$

The case $\widehat{\boldsymbol{\beta}} = \mathbf{C}_2'\mathbf{y}$ can be handled similarly. □

3.4. *Let* $\widehat{\boldsymbol{\beta}} = \mathbf{C}_3\mathbf{y}$ *or* $\widehat{\boldsymbol{\beta}} = \mathbf{C}_2'\mathbf{y}$. *If* $\boldsymbol{\ell}'\boldsymbol{\beta}$, $\mathbf{m}'\boldsymbol{\beta}$ *are estimable functions, then*

$$\operatorname{cov}(\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}, \mathbf{m}'\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\ell}'\mathbf{C}_4\mathbf{m} = \sigma^2\mathbf{m}'\mathbf{C}_4\boldsymbol{\ell}.$$

In particular, $\operatorname{var}(\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\ell}'\mathbf{C}_4\boldsymbol{\ell}$.

Proof. Since $\boldsymbol{\ell}'\boldsymbol{\beta}$, $\mathbf{m}'\boldsymbol{\beta}$ are estimable, $\boldsymbol{\ell}' = \mathbf{u}'\mathbf{X}$, $\mathbf{m}' = \mathbf{w}'\mathbf{X}$ for some $\mathbf{u}$, $\mathbf{w}$. If $\widehat{\boldsymbol{\beta}} = \mathbf{C}_3\mathbf{y}$, then

$$\begin{aligned}\operatorname{cov}(\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}, \mathbf{m}'\widehat{\boldsymbol{\beta}}) &= \operatorname{cov}(\mathbf{u}'\mathbf{X}\mathbf{C}_3\mathbf{y}, \mathbf{w}'\mathbf{X}\mathbf{C}_3\mathbf{y})\\ &= \sigma^2\mathbf{u}'\mathbf{X}\mathbf{C}_3\mathbf{V}\mathbf{C}_3'\mathbf{X}'\mathbf{w}\\ &= \sigma^2\mathbf{u}'\mathbf{X}\mathbf{C}_3\mathbf{V}\mathbf{w}\\ &= \sigma^2\mathbf{u}'\mathbf{X}\mathbf{C}_4\mathbf{X}'\mathbf{w}\\ &= \sigma^2\boldsymbol{\ell}'\mathbf{C}_4\mathbf{m}\end{aligned}$$

by **3.2**. Since the transpose of the matrix on the right-hand side of (7) is also a possible g-inverse, we have

$$\operatorname{cov}(\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}, \mathbf{m}'\widehat{\boldsymbol{\beta}}) = \sigma^2\mathbf{m}'\mathbf{C}_4\boldsymbol{\ell}.$$

The case $\widehat{\boldsymbol{\beta}} = \mathbf{C}_2'\mathbf{y}$ also follows by the same observation. □

3.5. *An unbiased estimate of* σ^2 *is* $\alpha^{-1}\mathbf{y}'\mathbf{C}_1\mathbf{y}$, *where* $\alpha = R[\mathbf{V}, \mathbf{X}] - R(\mathbf{X})$.

Proof. Suppose $\mathbf{u}'\mathbf{V} = \mathbf{0}$ for some $\mathbf{u}$. Then

$$\begin{aligned}\operatorname{var}(\mathbf{u}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})) &= \mathbf{u}'[E(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})']\mathbf{u}\\ &= \mathbf{u}'\mathbf{V}\mathbf{u} = 0,\end{aligned}$$

and hence $\mathbf{u}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = 0$ (with probability one). Thus $\mathbf{y}-\mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{V})$ and therefore $\mathbf{y}-\mathbf{X}\boldsymbol{\beta} = \mathbf{V}\mathbf{w}$ for some $\mathbf{w}$. We have

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{C}_1(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{C}_1(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) - \boldsymbol{\beta}'\mathbf{X}'\mathbf{C}_1\mathbf{V}\mathbf{w}\\ &= \mathbf{y}'\mathbf{C}_1(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\end{aligned}$$

since by **3.2**, $\mathbf{X}'\mathbf{C}_1\mathbf{V} = \mathbf{0}$. Thus

$$\begin{aligned} E(\mathbf{y} - \mathbf{X}\beta)'\mathbf{C}_1(\mathbf{y} - \mathbf{X}\beta) &= E(\mathbf{y}'\mathbf{C}_1\mathbf{y}) - \beta'\mathbf{X}'\mathbf{C}_1\mathbf{X}\beta \\ &= E(\mathbf{y}'\mathbf{C}_1\mathbf{y}), \end{aligned} \tag{19}$$

since by **3.2**, $\mathbf{X}'\mathbf{C}_1\mathbf{X} = \mathbf{0}$. However,

$$\begin{aligned} &E(\mathbf{y} - \mathbf{X}\beta)'\mathbf{C}_1(\mathbf{y} - \mathbf{X}\beta) \\ &\quad = E\{\text{trace}(\mathbf{y} - \mathbf{X}\beta)'\mathbf{C}_1(\mathbf{y} - \mathbf{X}\beta)\} \\ &\quad = E\{\text{trace}\mathbf{C}_1(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)'\} \\ &\quad = \text{trace}\mathbf{C}_1 E(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)' \\ &\quad = \sigma^2\text{trace}\mathbf{C}_1\mathbf{V} \\ &\quad = \sigma^2\{R[\mathbf{V}, \mathbf{X}] - R(\mathbf{X})\}, \end{aligned}$$

where again we have used (3.2). Substituting in (19), the result is proved. □

6.4 The Star Order

Let $M_{m\times n}$ denote the set of $m \times n$ matrices. A binary relation $\prec$ on $M_{m\times n}$ is said to be a *partial order* if it satisfies the following conditions:

(i) $\mathbf{A} \prec \mathbf{A}$ for every $\mathbf{A}$ (reflexivity).
(ii) $\mathbf{A} \prec \mathbf{B}, \mathbf{B} \prec \mathbf{A} \Rightarrow \mathbf{A} = \mathbf{B}$ (antisymmetry).
(iii) $\mathbf{A} \prec \mathbf{B}, \mathbf{B} \prec \mathbf{C} \Rightarrow \mathbf{A} \prec \mathbf{C}$ (transitivity).

Note that not every pair of matrices is necessarily comparable. Thus there exist pairs $\mathbf{A}, \mathbf{B}$ for which neither $\mathbf{A} \prec \mathbf{B}$ nor $\mathbf{B} \prec \mathbf{A}$ is true.

The *minus partial order*, denoted by $<^-$, is defined as follows. If $\mathbf{A}, \mathbf{B}$ are $m \times n$ matrices, then $\mathbf{A} <^- \mathbf{B}$ if $R(\mathbf{B}) = R(\mathbf{A}) + R(\mathbf{B} - \mathbf{A})$. The fact that this is a partial order can be seen as follows.

Clearly, $<^-$ is reflexive. If $\mathbf{A} <^- \mathbf{B}$ and $\mathbf{B} <^- \mathbf{A}$, then $R(\mathbf{B} - \mathbf{A}) = 0$ and hence $\mathbf{A} = \mathbf{B}$. Thus $<^-$ is antisymmetric. Finally, by (iii) of **2.1**, $\mathbf{A} <^- \mathbf{B}$ if and only if every g-inverse of $\mathbf{B}$ is a g-inverse of $\mathbf{A}$, and hence it follows that $<^-$ is transitive. Thus $<^-$ is a partial order.

The *star order* can be viewed as a refinement of the minus order, although historically the star order was introduced earlier. The definition is as follows. If $\mathbf{A}, \mathbf{B}$ are $m \times n$ matrices, then $\mathbf{A}$ is dominated by $\mathbf{B}$ under the star order, denoted by $\mathbf{A} <^* \mathbf{B}$, if $(\mathbf{B} - \mathbf{A})\mathbf{A}' = \mathbf{0}$ and $\mathbf{A}'(\mathbf{B} - \mathbf{A}) = \mathbf{0}$. Note that for complex matrices these conditions would be reformulated as $(\mathbf{B} - \mathbf{A})\mathbf{A}^* = \mathbf{0}$ and $\mathbf{A}^*(\mathbf{B} - \mathbf{A}) = \mathbf{0}$, where $\mathbf{A}^*$ denotes the complex conjugate of $\mathbf{A}$, and this explains the term "star order."

The star order is closely related to the Moore–Penrose inverse. This is mainly due to the following property of the Moore–Penrose inverse.

4.1. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix. Then* $\mathcal{C}(\mathbf{A}^+) = \mathcal{C}(\mathbf{A}')$ *and* $\mathcal{R}(\mathbf{A}^+) = \mathcal{R}(\mathbf{A}')$.

PROOF. We have $\mathbf{A}^+ = \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}'(\mathbf{A}^+)'\mathbf{A}^+$, and thus $\mathcal{C}(\mathbf{A}^+) \subset \mathcal{C}(\mathbf{A}')$. However, $R(\mathbf{A}^+) = R(\mathbf{A})$, and hence the two spaces must be equal. The second part is proved similarly. □

We now provide some alternative definitions of the star order.

4.2. *Let* $\mathbf{A}, \mathbf{B}$ *be* $m \times n$ *matrices. Then the following conditions are equivalent.*

(i) $\mathbf{A} <^* \mathbf{B}$, *i.e.,* $(\mathbf{B} - \mathbf{A})\mathbf{A}' = \mathbf{0}$, $\mathbf{A}'(\mathbf{B} - \mathbf{A}) = \mathbf{0}$.
(ii) $(\mathbf{B} - \mathbf{A})\mathbf{A}^+ = \mathbf{0}$, $\mathbf{A}^+(\mathbf{B} - \mathbf{A}) = \mathbf{0}$.
(iii) $\mathcal{C}(\mathbf{A}) \perp \mathcal{C}(\mathbf{B} - \mathbf{A})$, $\mathcal{R}(\mathbf{A}) \perp \mathcal{R}(\mathbf{B} - \mathbf{A})$.

PROOF. The equivalence of (i) and (ii) follows from **4.1**, while the equivalence of (i) and (iii) is trivial. □

As observed in **2.1**, $\mathbf{A} <^- \mathbf{B}$ if and only if every g-inverse of $\mathbf{B}$ is a g-inverse of $\mathbf{A}$. We wish to obtain an analogous statement for the star order. We first prove some preliminary results. The next result explains our earlier remark that the star order can be viewed as a refinement of the minus order.

4.3. *Let* $\mathbf{A}, \mathbf{B}$ *be* $m \times n$ *matrices and suppose* $\mathbf{A} <^* \mathbf{B}$. *Then* $\mathbf{A} <^- \mathbf{B}$.

PROOF. If $\mathbf{A} <^* \mathbf{B}$, then by **4.2**, $(\mathbf{B} - \mathbf{A})\mathbf{A}^+ = \mathbf{0}$, $\mathbf{A}^+(\mathbf{B} - \mathbf{A}) = \mathbf{0}$. It follows by (iv) of **2.1** that $\mathbf{A} <^- \mathbf{B}$. □

4.4. *Let* $\mathbf{A}, \mathbf{B}$ *be* $m \times n$ *matrices and suppose* $\mathbf{A} <^* \mathbf{B}$. *Then* $\mathbf{B}^+ = \mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+$.

PROOF. If $\mathbf{A} <^* \mathbf{B}$, then clearly $(\mathbf{B} - \mathbf{A}) <^* \mathbf{B}$. Now using **4.2**, we have

$$\mathbf{B}(\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+) = \mathbf{A}\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})(\mathbf{B} - \mathbf{A})^+$$

and

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

Thus both $\mathbf{B}(\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+)$ and $(\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+)\mathbf{B}$ are symmetric. Also,

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

Thus $\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+$ is a least squares, minimum norm g-inverse of $\mathbf{B}$. To show that it is also reflexive, it is sufficient to show that it has the same rank as $\mathbf{B}$. This is seen as follows. We have

$$\begin{aligned} R(\mathbf{B}) &\le R(\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+) \\ &\le R(\mathbf{A}^+) + R((\mathbf{B} - \mathbf{A})^+) \\ &= R(\mathbf{A}) + R(\mathbf{B} - \mathbf{A}) \\ &= R(\mathbf{B}), \end{aligned}$$

where the last equality follows using **4.3**. Thus $\mathbf{B}$ has the same rank as $\mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+$, and the proof is complete. □

We now present a characterization of the star order in terms of g-inverses.

4.5. *Let* $\mathbf{A}, \mathbf{B}$ *be* $m \times n$ *matrices. Then* $\mathbf{A} <^* \mathbf{B}$ *if and only if every minimum norm g-inverse of* $\mathbf{B}$ *is a minimum norm g-inverse of* $\mathbf{A}$ *and every least squares g-inverse of* $\mathbf{B}$ *is a least squares g-inverse of* $\mathbf{A}$.

PROOF. First suppose $\mathbf{A} <^* \mathbf{B}$. Let $\mathbf{B}_{\mathbf{m}}^-$ be an arbitrary minimum norm g-inverse of $\mathbf{B}$. Then (see Exercise 6)

$$\mathbf{B}_{\mathbf{m}}^- = \mathbf{B}^+ + \mathbf{V}(\mathbf{I} - \mathbf{B}\mathbf{B}^+) \tag{20}$$

for some matrix $\mathbf{V}$. First note that by **4.3**, $\mathbf{B}_{\mathbf{m}}^-$ is a g-inverse of $\mathbf{A}$.

Since $\mathbf{B}^+ = \mathbf{A}^+ + (\mathbf{B} - \mathbf{A})^+$ by **4.4**, then $\mathbf{B}^+\mathbf{A} = \mathbf{A}^+\mathbf{A} + (\mathbf{B} - \mathbf{A})^+\mathbf{A}$. However, by **4.2**, $\mathcal{C}(\mathbf{A}) \perp \mathcal{C}(\mathbf{B} - \mathbf{A})$, and by **4.1**, this latter space is the same as the row space of $(\mathbf{B} - \mathbf{A})^+$. Thus $(\mathbf{B} - \mathbf{A})^+\mathbf{A} = \mathbf{0}$, and hence $\mathbf{B}^+\mathbf{A} = \mathbf{A}^+\mathbf{A}$.

Also, $\mathbf{A} - \mathbf{B}\mathbf{B}^+\mathbf{A} = \mathbf{A} - \mathbf{B}\mathbf{A}^+\mathbf{A} = \mathbf{A} - \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{0}$, since $\mathbf{B}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^+$. These observations, together with (20), yield $\mathbf{B}_{\mathbf{m}}^-\mathbf{A} = \mathbf{A}^+\mathbf{A}$, which is symmetric. Therefore, $\mathbf{B}_{\mathbf{m}}^-$ is a minimum norm g-inverse of $\mathbf{A}$. Similarly, we can show that any least squares g-inverse of $\mathbf{B}$ is a least squares g-inverse of $\mathbf{A}$.

We now show the "if" part. For any matrix $\mathbf{V}$, $\mathbf{B}^+ + \mathbf{V}(\mathbf{I} - \mathbf{B}\mathbf{B}^+)$ is a minimum norm g-inverse of $\mathbf{B}$ and hence, by hypothesis, of $\mathbf{A}$. Thus

$$\mathbf{A}(\mathbf{B}^+ + \mathbf{V}(\mathbf{I} - \mathbf{B}\mathbf{B}^+))\mathbf{A} = \mathbf{A}.$$

Note that $\mathbf{B}^+$ is a minimum norm g-inverse of $\mathbf{B}$, and hence it is a g-inverse of $\mathbf{A}$. Thus $\mathbf{A}\mathbf{B}^+\mathbf{A} = \mathbf{A}$. Therefore, $\mathbf{A}\mathbf{V}(\mathbf{I} - \mathbf{B}\mathbf{B}^+)\mathbf{A} = \mathbf{A}$. Since $\mathbf{V}$ is arbitrary, we conclude that $(\mathbf{I} - \mathbf{B}\mathbf{B}^+)\mathbf{A} = \mathbf{0}$. Similarly, using the fact that any least squares g-inverse of $\mathbf{B}$ is a least squares g-inverse of $\mathbf{A}$, we conclude that $\mathbf{A}(\mathbf{I} - \mathbf{B}^+\mathbf{B}) = \mathbf{0}$. Thus for arbitrary $\mathbf{U}, \mathbf{V}$,

$$\mathbf{B}^+ + (\mathbf{I} - \mathbf{B}^+\mathbf{B})\mathbf{U} + \mathbf{V}(\mathbf{I} - \mathbf{B}\mathbf{B}^+)$$

is a g-inverse of $\mathbf{A}$. It follows (see Exercise 1 of Chapter 2) that any g-inverse of $\mathbf{B}$ is a g-inverse of $\mathbf{A}$ and hence $\mathbf{A} <^- \mathbf{B}$. Thus by **2.1**, $(\mathbf{B} - \mathbf{A})\mathbf{B}^-\mathbf{A} = \mathbf{0}$ for any g-inverse $\mathbf{B}^-$ of $\mathbf{B}$. Therefore, for any minimum norm g-inverse $\mathbf{B}_{\mathbf{m}}^-$ of $\mathbf{B}$,

$$\mathbf{0} = (\mathbf{B} - \mathbf{A})\mathbf{B}_{\mathbf{m}}^-\mathbf{A} = (\mathbf{B} - \mathbf{A})\mathbf{A}'(\mathbf{B}_{\mathbf{m}}^-)',$$

and hence

$$(\mathbf{B} - \mathbf{A})\mathbf{A}' = (\mathbf{B} - \mathbf{A})\mathbf{A}'(\mathbf{B}_{\mathbf{m}}^-)'\mathbf{A}' = \mathbf{0}.$$

Similarly, we can show that $\mathbf{A}'(\mathbf{B} - \mathbf{A}) = \mathbf{0}$ and hence $\mathbf{A} <^* \mathbf{B}$. That completes the proof. □

6.5 Exercises

1. If $\mathbf{A}, \mathbf{B}$ are positive semidefinite matrices, then show that they are parallel summable.

2. Let $\mathbf{A_1}, \dots, \mathbf{A_k}$ be $m \times n$ matrices and let $\mathbf{A} = \sum_{i=1}^{k} \mathbf{A_i}$. Show that $R(\mathbf{A}) = \sum_{i=1}^{k} R(\mathbf{A_i})$ only if for every $S \subset \{1, \dots, k\}$,

$$R(\mathbf{A}) = R\left(\sum_{i \in S} \mathbf{A_i}\right) + R\left(\sum_{i \notin S} \mathbf{A_i}\right).$$

3. Let $\mathbf{A}$ be a positive semidefinite matrix. Show that for any choice of the g-inverse,

$$\mathbf{x}'(\mathbf{A} + \mathbf{x}\mathbf{x}')^{-}\mathbf{x}$$

equals

$$\frac{\mathbf{x}'\mathbf{A}^{-}\mathbf{x}}{1 + \mathbf{x}'\mathbf{A}^{-}\mathbf{x}}$$

if $\mathbf{x} \in \mathcal{C}(\mathbf{A})$ and 1 otherwise.

4. Let d be a connected design and let $\mathbf{e}$ denote the column vector of all ones. Show that for any $\alpha \neq 0$, $\mathbf{C_d} + \alpha\mathbf{e}\mathbf{e}'$ is nonsingular and its inverse is a g-inverse of $\mathbf{C_d}$.

5. Let $\mathbf{A}$ be an $m \times n$ matrix and let $\mathbf{b}$ be an $m \times 1$ vector. If $\mathbf{G}$ is a least squares g-inverse of $\mathbf{A}$, then show that $\|\mathbf{A}^{+}\mathbf{b}\| \leq \|\mathbf{G}\mathbf{b}\|$. (Thus $x^0 = \mathbf{A}^{+}\mathbf{b}$ has minimum norm among all least squares solutions of the (not necessarily consistent) system $\mathbf{A}\mathbf{x} = \mathbf{b}$.)

6. Let $\mathbf{B}$ be an $m \times n$ matrix. Show that the class of minimum norm g-inverses of $\mathbf{B}$ is given by $\mathbf{B}^{+} + \mathbf{V}(\mathbf{I} - \mathbf{B}\mathbf{B}^{+})$, where $\mathbf{V}$ is arbitrary.

7. Let $\mathbf{A}, \mathbf{B}$ be $m \times n$ matrices. Show that $\mathbf{A} <^{*} \mathbf{B}$ if and only if $\mathbf{A}^{+} <^{*} \mathbf{B}^{+}$.

8. Let $\mathbf{A_1}, \dots, \mathbf{A_k}$ be $m \times n$ matrices and let $\mathbf{A} = \sum_{i=1}^{k} \mathbf{A_i}$. Give equivalent conditions for $R(\mathbf{A})$ to equal $\sum_{i=1}^{k} R(\mathbf{A_i})$, imitating **2.1**.

9. If $\mathbf{A}, \mathbf{B}$ are parallel summable, then show that

(i) $P(\mathbf{A}, \mathbf{B}) = P(\mathbf{B}, \mathbf{A})$.
(ii) $\mathbf{A}', \mathbf{B}'$ are parallel summable and $P(\mathbf{A}', \mathbf{B}') = [P(\mathbf{A}, \mathbf{B})]'$.
(iii) $\mathcal{C}[P(\mathbf{A}, \mathbf{B})] = \mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B})$.

10. Let $\mathbf{A_1}, \dots, \mathbf{A_k}$ be $m \times n$ matrices and let $\mathbf{A} = \sum_{i=1}^{k} \mathbf{A_i}$. Consider the following statements:

(a) Every $\mathbf{A}^{-}$ is a g-inverse of each $\mathbf{A_i}$.
(b) Every $\mathbf{A}^{-}$ satisfies $\mathbf{A_i}\mathbf{A}^{-}\mathbf{A_j} = \mathbf{0}$, $i \neq j$.
(c) For every $\mathbf{A}^{-}$, $R(\mathbf{A_i}\mathbf{A}^{-}\mathbf{A_i}) = R(\mathbf{A_i})$, $i = 1, \dots, k$.
(d) $R(\mathbf{A}) = \sum_{i=1}^{k} R(\mathbf{A_i})$.

Prove that (a) $\Rightarrow$ (b),(c),(d); (b),(c) $\Rightarrow$ (a),(d); (d) $\Rightarrow$ (a),(b),(c).

11. Let $\mathbf{N}, \mathbf{U}, \mathbf{V}$ be matrices of orders $m \times n$, $m \times p$, $q \times n$, respectively, let

$$\mathbf{F} = \begin{bmatrix} \mathbf{N} & \mathbf{U} \\ \mathbf{V} & \mathbf{0} \end{bmatrix},$$

and suppose

$$R(\mathbf{F}) = R[\mathbf{N}, \mathbf{U}] + R(\mathbf{V}) = R\begin{bmatrix} \mathbf{N} \\ \mathbf{V} \end{bmatrix} + R(\mathbf{U}).$$

Let

$$\begin{bmatrix} \mathbf{C_1} & \mathbf{C_2} \\ \mathbf{C_3} & -\mathbf{C_4} \end{bmatrix}$$

be a g-inverse of **F**. Show that

$$\mathbf{U C_3 U} = \mathbf{U}, \quad \mathbf{V C_2 V} = \mathbf{V}$$

and

$$\mathbf{U C_3 N} = \mathbf{N C_2 V} = \mathbf{U C_4 V}.$$

Furthermore, show that the common matrix in the equation above is invariant under the choice of the g-inverse of **F**. (This matrix is known as the *"shorted matrix"* **N** relative to the column space of **U** and the row space of **V**.)

12. A square matrix **V** is said to be *almost definite* if $\mathbf{x}'\mathbf{V}\mathbf{x} = 0$ implies $\mathbf{V}\mathbf{x} = 0$ for any $\mathbf{x} \in R^n$. Prove that any positive semidefinite matrix is almost definite. Let **V** be an almost definite $n \times n$ matrix and let **X** be $n \times m$. Then show that

$$R\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} = R[\mathbf{V}, \mathbf{X}] + R(\mathbf{X}).$$

13. Let $\mathbf{A_i}$, $i = 1, \dots, k$ be $m \times m$ matrices and let $\mathbf{A} = \mathbf{A_1} + \cdots + \mathbf{A_k}$. Consider the following statements:

(1) Each $\mathbf{A_i}$ is idempotent.
(2) $\mathbf{A_i A_j} = \mathbf{0}, i \neq j$, and $R(\mathbf{A_i^2}) = R(\mathbf{A_i})$ for all i.
(3) **A** is idempotent.
(4) $R(\mathbf{A}) = R(\mathbf{A_1}) + \cdots + R(\mathbf{A_k})$.

Prove that

(i) (1) and (2) together imply (3), (4).
(ii) (1) and (3) together imply (2), (4).
(iii) (2) and (3) together imply (1), (4).

6.6 Hints and Solutions

1. Let $\mathbf{A} = \mathbf{X}\mathbf{X}'$, $\mathbf{B} = \mathbf{Y}\mathbf{Y}'$. Then $\mathbf{A} + \mathbf{B} = [\mathbf{X}, \mathbf{Y}][\mathbf{X}, \mathbf{Y}]'$, and thus the column space of **B** is contained in the column space of $\mathbf{A} + \mathbf{B}$, while the row space of **A** is contained in the row space of $\mathbf{A} + \mathbf{B}$. Thus **A**, **B** are parallel summable.

2. For any $S \subset \{1, \ldots, k\}$, we have

$$R(\mathbf{A}) \leq R\left(\sum_{i \in S} \mathbf{A_i}\right) + R\left(\sum_{i \notin S} \mathbf{A_i}\right) \leq \sum_{i=1}^{k} R(\mathbf{A_i}).$$

Thus if $R(\mathbf{A}) = \sum_{i=1}^{k} R(\mathbf{A_i})$, then equality must occur in the inequality above, and the result is proved.

3. If $\mathbf{x} \notin \mathcal{C}(\mathbf{A})$, then $R(\mathbf{A} + \mathbf{xx}') = R(\mathbf{A}) + R(\mathbf{xx}')$, and thus any g-inverse of $\mathbf{A} + \mathbf{xx}'$ is a g-inverse of $\mathbf{xx}'$. Thus $\mathbf{x}'(\mathbf{A} + \mathbf{xx}')^{-}\mathbf{x} = \mathbf{x}'(\mathbf{xx}')^{-}\mathbf{x} = \mathbf{x}'(\mathbf{xx}')^{+}\mathbf{x} = 1$. If $\mathbf{x} \in \mathcal{C}(\mathbf{A})$, then using the spectral theorem we can reduce the problem to the case where $\mathbf{A}$ is positive definite. Then it is easily verified that

$$(\mathbf{A} + \mathbf{xx}')^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{x}'\mathbf{A}^{-1}\mathbf{x}}\mathbf{A}^{-1}\mathbf{xx}'\mathbf{A}^{-1}.$$

Using this identity we can simplify $\mathbf{x}'(\mathbf{A} + \mathbf{xx}')^{-1}\mathbf{x}$, and the result is proved.

4. Suppose $\mathbf{x}$ is a nonzero vector such that $(\mathbf{C_d} + \alpha\mathbf{ee}')\mathbf{x} = \mathbf{0}$. Premultiplying by $\mathbf{e}'$ we see that $\mathbf{e}'\mathbf{x} = 0$, and hence $\mathbf{C_d}\mathbf{x} = \mathbf{0}$. Since d is connected, it follows that $\mathbf{x}$ is a multiple of $\mathbf{e}$, which contradicts $\mathbf{e}'\mathbf{x} = 0$. Thus $\mathbf{x} = \mathbf{0}$, and $\mathbf{C_d} + \alpha\mathbf{ee}'$ is nonsingular. Also, $R(\mathbf{C_d} + \alpha\mathbf{ee}') = R(\mathbf{C_d}) + R(\mathbf{ee}')$, and hence the inverse of $\mathbf{C_d} + \alpha\mathbf{ee}'$ is a g-inverse of $\mathbf{C_d}$.

5. Let $\mathbf{A}^{-} = \mathbf{A}^{+} + \mathbf{X}$. Then $\mathbf{AXA} = \mathbf{0}$ and $\mathbf{AX}$ is symmetric. Therefore, $\mathbf{X}'\mathbf{A}'\mathbf{A} = \mathbf{0}$, and it follows that $\mathbf{AX} = \mathbf{0}$. Since $\mathcal{R}(\mathbf{A}^{+}) = \mathcal{R}(\mathbf{A}')$, then $(\mathbf{A}^{+})'\mathbf{X} = \mathbf{0}$. Thus

$$(\mathbf{A}^{-})'\mathbf{A}^{-} = (\mathbf{A}^{+} + \mathbf{X})'(\mathbf{A}^{+} + \mathbf{X}) = (\mathbf{A}^{+})'\mathbf{A}^{+} + \mathbf{X}'\mathbf{X} \geq (\mathbf{A}^{+})'\mathbf{A}^{+}.$$

Therefore, $(\mathbf{A}^{+}\mathbf{b})'\mathbf{A}^{+}\mathbf{b} \geq (\mathbf{A}^{-}\mathbf{b})'\mathbf{A}^{-}\mathbf{b}$, and the proof is complete.

6. Clearly, for any $\mathbf{V}$, $\mathbf{B}^{+} + \mathbf{V}(\mathbf{I} - \mathbf{BB}^{+})$ is a minimum norm g-inverse of $\mathbf{B}$. Conversely, suppose $\mathbf{G}$ is a minimum norm g-inverse of $\mathbf{B}$. Setting $\mathbf{V} = \mathbf{B}^{+} + \mathbf{G}$, we see that

$$\mathbf{B}^{+} + \mathbf{V}(\mathbf{I} - \mathbf{BB}^{+}) = \mathbf{B}^{+} + \mathbf{G} - \mathbf{GBB}^{+}.$$

Now, $\mathbf{GBB}^{+} = \mathbf{B}'\mathbf{G}'\mathbf{B}^{+}$, which equals $\mathbf{B}^{+}$, as can be seen using $\mathcal{C}(\mathbf{B}^{+}) = \mathcal{C}(\mathbf{B}')$.

7. Suppose $\mathbf{A} <^{*} \mathbf{B}$. Then by **4.2**, $\mathbf{AA}^{+} = \mathbf{BA}^{+}$, $\mathbf{A}^{+}\mathbf{A} = \mathbf{A}^{+}\mathbf{B}$. Now, $\mathbf{B}^{+}(\mathbf{A}^{+})^{+} = \mathbf{B}^{+}\mathbf{A} = (\mathbf{A}^{+} + (\mathbf{B} - \mathbf{A})^{+})\mathbf{A}$ by **4.4**, which equals $\mathbf{A}^{+}\mathbf{A}$, as observed in the proof of **4.5**. Similarly, we can show $(\mathbf{A}^{+})^{+}\mathbf{B}^{+} = \mathbf{AA}^{+}$, and hence $\mathbf{A}^{+} <^{*} \mathbf{B}^{+}$. The converse is immediate.

The solution of Exercise 8 is left to the reader. For solutions to the remaining exercises, see comments on Chapter 6 in the Notes section following this chapter.

This page intentionally left blank

Notes

Chapter 1

From among the numerous books dealing with linear algebra and matrix theory we particularly mention Horn and Johnson (1985), Mehta (1989), Mirsky (1955), Rao and Rao (1998), and Strang (1980). Several problems complementing the material in this chapter are found in Zhang (1996).

The proof of **4.1** is taken from Bhimasankaram (1988); this paper contains several applications of rank factorization as well.

The proof of **8.5** is based on an idea suggested by N. Ekambaram. Some readers may find the development in this section a bit unusual. But this approach seems necessary if one wants to avoid the use of complex vector spaces and lead toward the spectral theorem.

Chapter 2

The books by Rao and Mitra (1971), Ben-Israel and Greville (1974), and Campbell and Meyer (1979) contain a vast amount of material on the generalized inverse.

Assertions **1.4–1.6** follow the treatment in Rao (1973), pp. 48–50.

Exercise 20 is based on Bhaskara Rao (1983). Bapat et al. (1990) and Prasad et al. (1991) constitute a generalization of this work.

For a survey of Hadamard matrices we refer to Hedayat and Wallis (1978).

For some inequalities related to the Hadamard inequality, see Bapat and Raghavan (1997).

Chapter 3

The term "Schur complement" was coined by E. Haynsworth (1968) (see also Carlson (1986)).

We refer to Muirhead (1982) for a relatively modern treatment of multivariate analysis.

There are numerous results in the literature on the distributions of quadratic forms; see the discussion in Searle (1971), Anderson and Styan (1982), and the references contained therein.

The proof of Cochran's theorem given here is not widely known.

Our treatment in this as well as the previous chapter is clearly influenced by the books by Searle (1971), Seber (1977), and Rao (1973, Chapter 4). In deriving the F-test for a linear hypothesis we have adopted a slightly different method.

Christensen (1987) is a nice book emphasizing the projections approach. Sen and Srivastava (1990) is highly recommended for an account of applications of linear models. For some optimal properties of the F-statistic used in one-way and two-way classifications we refer to the discussion in Scheffe (1959, Chapter 2).

For more applications of generalized Schur complement and for several results related to Exercise 13, see Nordström (1989).

Chapter 4

Most of the material in the first three sections of this chapter has been treated in greater detail in Horn and Johnson (1985); where more inequalities on singular values and eigenvalues can be found.

The proof of **2.6** given here is due to Ikebe et al. (1987), where some related inequalities are proved using the same technique.

The standard reference for majorization is Marshall and Olkin (1979). Arnold (1987) is another entertaining book on the subject.

Section 4.6 is based on Bapat and Ben-Israel (1995); the notion of volume was introduced in Ben-Israel (1992).

Results more general than those in Exercises 11, 12 are given in Bapat et al. (1990) and Bapat (1994).

The group inverse, introduced in Exercise 12, finds important applications in several areas, particularly in the theory of Markov chains; see Berman and Plemmons (1994). A result similar to that in Exercise 14 has been proved in Prasad et al. (1991), where the group inverse of a matrix over an integral domain is studied.

Chapter 5

We refer to the books by Dey (1986), Joshi (1987), and John (1971), where much more material on block designs and further references can be found.

Exercise 12 is essentially taken from Constantine (1987), which is recommended for a readable account of optimality.

Chapter 6

Many of the conditions in **2.1** are contained in the papers due to S.K. Mitra and his coauthors. We refer to Carlson (1987) and Mitra (1991) for more information.

Result **3.2** is part of the "inverse partitioned matrix method"; see Rao (1973), p. 294. The proof given here, using rank additivity, is due to Mitra (1982). Results **3.3–3.5** can be found in Rao (1973), p. 298.

The definition of the minus partial order is attributed to Hartwig (1980) and Nambooripad (1980). Several extensions of this order have been considered; see Mitra (1991) for a unified treatment.

The star order was introduced by Drazin (1978). The result in **4.5** is due to Mitra (1986).

Exercise 9 and further properties of the parallel sum can be found in Rao and Mitra (1971). Exercise 10 and related results are in Anderson and Styan (1982). Exercises 11, 12 are based on Mitra (1986) and Mitra and Puri (1983), respectively. A solution to Exercise 13 is found in Rao (1973), p. 28.

This page intentionally left blank

References

Anderson, T.W. and Styan, G.P.H. (1982). Cochran's theorem, rank additivity and tripotent matrices, 1–23, *Statistics and Probability: Essays in honor of C.R. Rao*, G. Kallianpur, P.R. Krishnaiah, J.K. Ghosh eds. North-Holland Publ. Co.

Arnold, B.C. (1987). *Majorization and the Lorentz Order: A Brief Introduction*, Springer-Verlag, Berlin.

Bapat, R.B. (1994). Generalized inverses with proportional minors, *Linear Algebra and Its Applications*, 211:27–33.

Bapat, R.B. and Ben-Israel, Adi (1995). Singular values and maximum rank minors of generalized inverses, *Linear and Multilinear Algebra*, 40:153–161.

Bapat, R.B., Bhaskara Rao, K.P.S., and Prasad, K.M. (1990). Generalized inverses over integral domains, *Linear Algebra and Its Applications*, 140:181–196.

Bapat, R.B. and Raghavan, T.E.S. (1997). *Nonnegative Matrices and Applications*, Encyclopedia of Mathematical Sciences, Vol. 64, Cambridge University Press, Cambridge.

Ben-Israel, Adi (1992). A volume associated with $m \times n$ matrices, *Linear Algebra and Its Applications*, 167:87–111.

Ben-Israel, A. and Greville, T.N.E. (1974). *Generalized Inverses: Theory and Applications*, Wiley-Interscience.

Berman, A. and Plemmons, R.J. (1994). *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia.

Bhaskara Rao, K.P.S. (1983). On generalized inverses of matrices over integral domains, *Linear Algebra and Its Applications*, 49:179–189.

Bhimasankaram, P. (1988). Rank factorization of a matrix and its applications, *Mathematical Scientist*, 13:4–14.

Campbell, S.L. and Meyer, C.D., Jr. (1979). *Generalized Inverses of Linear Transformations*, Pitman, London.

Carlson, D. (1986). What are Schur complements, anyway? *Linear Algebra and Its Applications*, 74:257–275.

Carlson, D. (1987). Generalized inverse invariance, partial orders, and rank-minimization problems for matrices, 81–87,*Current Trends in Matrix Theory*, Frank Uhlig, Robert Groine, eds., Elsevier Science Publishing Co., Inc., New York.

Christensen, R. (1987). *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer-Verlag, Berlin.

Constantine, G.M. (1987). *Combinatorial Theory and Statistical Design*, Wiley, New York.

Dey, A. (1986). *Theory of Block Designs*, Wiley Eastern Ltd., New Delhi.

Drazin, M.P. (1978). Natural structures on semigroups with involutions, *Bull. Amer. Math. Soc.* 84:139–141.

Hartwig, R.E. (1980). How to order regular elements? *Mathematica Japonica*, 25:1–13.

Haynsworth, E.V. (1968). Determination of the inertia of a partitioned Hermitian matrix, *Linear Algebra and Its Applications*, 1:73–82.

Hedayat, A. and Wallis, W.D. (1978). Hadamard matrices and their applications, *Annals of Statistics*, 6:1184–1238.

Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*, Cambridge University Press.

Ikebe, Y., Inagaki, T., and Miyamoto, S. (1987). The monotonicity theorem, Cauchy interlace theorem and the Courant–Fischer theorem, *American Mathematical Monthly*, 94(4):352–354.

John, P.W.M.(1971). *Statistical Design and Analysis of Experiments*, Macmillan, New York.

Joshi, D.D. (1987). *Linear Estimation and Design of Experiments*, Wiley Eastern Ltd., New Delhi.

Marshall, A.W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

Mehta, M.L. (1989). *Matrix Theory: Selected Topics and Useful Results*, enlarged re-edition, Hindustan Publishing Corporation, Delhi.

Mirsky, L. (1955). *An Introduction to Linear Algebra*, Oxford University Press.

Mitra, S.K. (1982). Simultaneous diagonalization of rectangular matrices, *Linear Algebra and Its Applications*, 47:139–150.

Mitra, S.K. (1986). The minus partial order and the shorted matrix, *Linear Algebra and Its Applications*, 83:1–27.

Mitra, S.K. (1991). Matrix partial orders through generalized inverses: Unified Theory, *Linear Algebra and Its Applications*, 148:237–263.

Mitra, S.K. and Puri, M.L. (1983). The fundamental bordered matrix of linear estimation and the Duffin–Morley general linear electromechanical system, *Applicable Analysis*, 14:241–258.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Nambooripad, K.S.S. (1980). The natural partial order on a regular semigroup, *Proceedings Edinburgh Mathematical Society*, 23:249–260.

Nordström, K. (1989). Some further aspects of the Löwner—ordering antitonicity of the Moore–Penrose inverse, *Communications in Statistics—Theory and Methods*, 18(12):4471–4489.

Prasad, K.M., Bhaskara Rao, K.P.S. and Bapat, R.B. (1991). Generalized inverses over integral domains II. Group inverses and Drazin inverses, *Linear Algebra and Its Applications*, 146:31–47.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

Rao, C.R. and Mitra, S.K. (1971). *Generalized Inverse of Matrices and Its Applications*, Wiley, New York.

Rao, C.R. and Rao, M.B. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics*, World Scientific, Singapore.

Scheffe, H. (1959). *The Analysis of Variance*, Wiley, New York.

Searle, S.R. (1971). *Linear Models*, Wiley, New York.

Seber, G.A.F. (1977). *Linear Regression Analysis*, Wiley, New York.

Sen, Ashish and Srivastava, Muni (1990). *Regression Analysis; Theory, Methods and Applications*, Springer-Verlag, New York.

Strang, G. (1980). *Linear Algebra and Its Applications*, 2nd ed., Academic Press, New York.

Zhang, Fuzhen, (1996). *Linear Algebra, Challenging Problems for Students*, The Johns Hopkins University Press, Baltimore and London.

This page intentionally left blank

Index