Generalized Inverse of Matrices and its Applications

Generalized Inverse of Matrices and its Applications

C. RADHAKRISHNA RAO, Sc.D., F.N.A., F.R.S.
Director, Research and Training School
Indian Statistical Institute

SUJIT KUMAR MITRA, Ph.D.
Professor of Statistics
Indian Statistical Institute

JOHN WILEY & SONS, INC.
New York London Sydney Toronto

Library of Congress Catalog Card Number: 74-158528

ISBN 0-471-70821-6

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Preface

This book is an attempt to bring together all the available results on "invertibility of singular matrices" under a unified theory and to discuss their applications.

It is well known that if **A** is a square non-singular matrix, then there exists a matrix **G**, such that $\mathbf{AG} = \mathbf{GA} = \mathbf{I}$, which is called the inverse of **A** and denoted by $\mathbf{A}^{-1}$. If **A** is a singular or a rectangular matrix, no such matrix **G** exists. However, Moore extended the notion of inverse to singular matrices in 1920 and discussed the concept at some length in 1935. Moore's definition of an inverse of **A** is equivalent to the existence of a matrix **G** such that

$$\mathbf{AG} = \mathbf{P_A}, \quad \mathbf{GA} = \mathbf{P_G}$$

where $\mathbf{P_X}$ stands for the projection operator onto $\mathcal{M}(\mathbf{X})$, the space generated by the columns of **X**. Unaware of Moore's work, Penrose defined in 1955 an inverse **G** of **A** as satisfying the conditions

$$\mathbf{AGA} = \mathbf{A}, \quad (\mathbf{AG})^* = \mathbf{AG}$$

$$\mathbf{GAG} = \mathbf{G}, \quad (\mathbf{GA})^* = \mathbf{GA}$$

which are equivalent to Moore's conditions (when the inner product between two vectors **x**, **y** is defined as $\mathbf{y}^*\mathbf{x}$, where * indicates conjugate transpose).

In three fundamental papers Tseng (1949a, 1949b and 1956) considered the problem of defining inverses of singular operators, which are more general than matrices. Attempts at defining and using an inverse of a singular matrix have been made from time to time (see Bjerhammer, 1951, 1957, 1958) but the results were less general or offered no systematic study.

In 1955 one of the authors, Rao, constructed an inverse of a singular matrix that occurs in normal equations in the least-squares theory, which he called a pseudoinverse, and showed that it serves the same purpose as the regular inverse of a nonsingular matrix in solving normal equations and also in computing standard errors of least-squares estimators. Rao's pseudoinverse did not satisfy all the conditions of Moore and Penrose, and the only property required of the inverse **G** was "that $\mathbf{x} = \mathbf{Gy}$ provides a solution of the equation $\mathbf{Ax} = \mathbf{y}$ for any **y**, such that $\mathbf{Ax} = \mathbf{y}$ is consistent." This is provided by a matrix **G** satisfying the only condition **AGA** = **A** in Penrose's definition. In 1962 Rao called a matrix **G** satisfying this single condition, **AGA** = **A**, a g-inverse (generalized inverse) of **A** and studied its properties in greater detail. In many practical applications, it is sufficient to work with a g-inverse satisfying this more general (weaker) definition, as demonstrated in two other publications by Rao in 1965 and 1966.

A g-inverse so defined is not unique and thus presents an interesting study in matrix algebra. In a publication in 1967, Rao showed how a variety of g-inverses could be constructed to suit different purposes and presented a classification (with nomenclature) of g-inverses.

The work was later pursued by Mitra (1968a and 1968b) who introduced some new classes of g-inverses. Further applications of g-inverses were considered in the joint publications by the authors (Mitra and Rao, 1968a, 1968b and 1969). The present book essentially describes the work of the authors and also within its framework brings all the important contributions by other authors on this subject up to date.

Some principal contributors to this subject since 1955 are Greville (1957), Bjerhammer (1957 and 1958), Ben-Israel and Charnes (1963), Chipman (1964), Chipman and Rao (1964) and Scroggs and Odell (1966). Bose (1959) mentions the use of g-inverse in his lecture notes on Analysis of Variance. Bott and Duffin (1953) introduced the concept of a constrained inverse of a square matrix, which is different from a g-inverse and is useful in some applications in network theory. Chernoff (1953) considered an inverse of a singular n.n.d. matrix, which is also not a g-inverse but is useful in discussing some problems in statistical estimation theory.

Chapter 1 of this book, *Generalized Inverse of Matrices and its Applications*, contains statements of certain results in matrix algebra which are used in the discussions of the later chapters. It also explains the notations used in the book. Properties of g-inverse based on the single condition **AGA** = **A**, and on the conditions **AGA** = **A**, **GAG** = **G** called reflexive g-inverse, are studied in Chapter 2. Solutions of some important matrix equations are also considered. It is shown that in many problems one needs a g-inverse satisfying the only condition **AGA** = **A**, and the other restraining conditions can only serve special purposes. Chapter 3 examines conditions on a g-inverse to obtain (i) minimum norm solution of a consistent equation **Ax** = **y**, (ii) least-squares solution of an inconsistent equation **Ax** = **y**, and (iii) minimum norm least-squares solution of an inconsistent equation **Ax** = **y**. It is shown that the g-inverse which produces a solution of the type (iii) is precisely the Moore–Penrose inverse. Other special types of g-inverses with reciprocal eigenvalue property, etc., are considered in Chapter 4. Chapter 5 contains a general discussion of projection operators and idempotent matrices and their explicit representations in terms of g-inverses of matrices. Simultaneous reduction of two hermitian forms when none of them need be positive definite is considered in Chapter 6. Applications of g-inverse in problems of estimation from linear models and robustness of statistical procedures under deviations from specified models are examined in Chapters 7 and 8. Very general results on the distribution of quadratic forms of random variables having a singular normal distribution are obtained in Chapter 9. Applications of g-inverse in network theory, mathematical programming and some problems in mathematical statistics (discriminant function when the dispersion matrix is singular and maximum likelihood estimation when the information matrix is singular) are considered in Chapter 10, while computational methods for obtaining a g-inverse are discussed in Chapter 11.

The new classes of constrained inverses introduced in Chapter 4 deserve special mention as they include all the types of g-inverses as special cases.

The material covered in the book relates to the research work done during the last 15 years and a number of unpublished results recently obtained by the authors. The applications of g-inverse are rapidly increasing; we have considered in some detail only a few of them. We hope that this full-length monograph on the subject will be of use to students and research workers in various fields. In the book we confine our attention to matrices only. Extension of the results on matrices to more general operators in abstract spaces offers a good scope for research work.

There is enough material in the book for one term course on g-inverse of matrices. In addition the book would be useful as supplementary material in a variety of courses such as Matrix Algebra, Network Theory, Mathematical Statistics, Optimization Problems and Numerical Analysis.

It gives us great pleasure to thank Mr. Arun Das and Mr. Mehar Lal, who have undertaken the heavy burden of typing the manuscript for the press at various stages of preparation, and to P. Bhimasankaram for reading the manuscript and making helpful comments.

Calcutta, India
April 1971

C. R. Rao
S. K. Mitra

Contents

CHAPTER 1

Notations and Preliminaries

In this chapter we introduce the notations and some of the preliminary results on matrices needed elsewhere in the text. The proofs of these results will be found in standard textbooks on matrix algebra and are therefore omitted here. See, for instance, books by Gantmacher (1959), Householder (1964), Pease (1965), and Perlis (1952). A fairly complete discussion of properties of matrices, with special reference to applications in mathematical statistics, is contained in Chapter 1 of Rao (1965).

Matrices are denoted by boldface capital letters such as $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma}$. Boldface lower-case letters $\mathbf{x}$, $\mathbf{y}$, ... denote column vectors, used synonymously for matrices with only one column. A null matrix or a null vector is denoted by $\mathbf{0}$. Unless otherwise stated, we shall consider only those matrices and vectors with elements defined over the field of complex numbers. In Chapters 7, 8, and 9, however, we consider only real matrices and vectors.

1.1 ROW AND COLUMN SPACES OF A MATRIX, SUBSPACES AND ORTHOGONAL COMPLEMENT, PROJECTION OPERATOR

Vector Spaces

For a matrix $\mathbf{A}$ of order $m \times n$ the linear space spanned by the columns of $\mathbf{A}$ is called the column space of $\mathbf{A}$ and denoted by the symbol $\mathscr{M}(\mathbf{A})$. Row space of $\mathbf{A}$, defined analogously, can therefore be denoted by $\mathscr{M}(\mathbf{A}')$; $\mathscr{E}^n$ and $\mathscr{R}^n$ denote the vector spaces of all n-tuples with complex and real coordinates, respectively.

Notice that $\mathscr{M}(\mathbf{A})$ consists of precisely those vectors in $\mathscr{E}^m$ which can be expressed as $\mathbf{Ax}$ for some $\mathbf{x}$ in $\mathscr{E}^n$. It is convenient to think of a matrix $\mathbf{A}$ as a linear transformation $\mathscr{E}^n \xrightarrow{\mathbf{A}} \mathscr{E}^m$, in which case $\mathscr{M}(\mathbf{A})$ is the range of the transformation $\mathbf{A}$. The null space of $\mathbf{A}$ is, on the other hand, the set of all vectors in $\mathscr{E}^n$ that are mapped into the null vector in $\mathscr{E}^m$ under this transformation.

Basis. Any set of linearly independent vectors spanning a given vector space (which may be a subspace) is called a basis of the vector space.

Dimension. The dimension of a vector space $\mathscr{V}$, denoted by $d[\mathscr{V}]$, is the number of vectors in a basis of $\mathscr{V}$.

Linear functional. A linear functional on a complex vector space $\mathscr{V}$ (i.e., a vector space on the field of complex numbers) is a complex-valued, additive homogeneous function ξ defined on $\mathscr{V}$; that is,

$$\xi(\mathbf{x} + \mathbf{y}) = \xi(\mathbf{x}) + \xi(\mathbf{y}) \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathscr{V},$$

$$\xi(\alpha\mathbf{x}) = \alpha\xi(\mathbf{x}) \quad \forall\, \mathbf{x} \in \mathscr{V} \quad \text{and arbitrary complex number } \alpha.$$

Bilinear Functional. A bilinear functional on a complex vector space $\mathscr{V}$ is a complex-valued function ϕ, defined on the cartesian product of $\mathscr{V}$ with itself such that, if

$$\xi_{\mathbf{y}}(\mathbf{x}) = \eta_{\mathbf{x}}(\mathbf{y}) = \phi(\mathbf{x}, \mathbf{y}),$$

then, for each $\mathbf{y} \in \mathscr{V}$, $\xi_{\mathbf{y}}$ is a linear functional on $\mathscr{V}$ and, for each $\mathbf{x} \in \mathscr{V}$, $\eta_{\mathbf{x}}$ is a conjugate linear functional.

Inner Product and Orthogonality. An inner product in a complex vector space $\mathscr{V}$ denoted by $(\mathbf{x}, \mathbf{y})$ is a symmetric, strictly positive, bilinear functional on $\mathscr{V}$, that is, a bilinear functional satisfying (i) $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$, and (ii) $(\mathbf{x}, \mathbf{x}) > 0\ \forall$ non-null vectors $\mathbf{x} \in \mathscr{V}$. An inner product space is a complex vector space with an agreed inner product definition. In an inner product space, vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be mutually orthogonal, if $(\mathbf{x}, \mathbf{y}) = 0$.

Reciprocal Bases (Dual Bases). Consider a vector space of finite dimension k and let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ be two alternative bases. One is called the reciprocal basis of the other if $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = \delta_{ij}$, where δ_{ij} is the Kronecker symbol.

Note that if $\mathbf{A}$ is a nonsingular matrix of order k, (a) the columns of $\mathbf{A}$ and the columns of $(\mathbf{A}^{-1})^*$ provide reciprocal bases of $\mathscr{E}^k$ if $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^*\mathbf{x}$, (b) the columns of $\mathbf{A}$ and the columns of $\mathbf{\Lambda}^{-1}(\mathbf{A}^{-1})^*$ provide reciprocal bases of $\mathscr{E}^k$ if $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^*\mathbf{\Lambda}\mathbf{x}$.

Intersection, Sum, and Direct Sum of Subspaces. If $\mathscr{S}$ and $\mathscr{T}$ are subspaces of a vector space $\mathscr{V}$, the set of vectors common to both $\mathscr{S}$ and $\mathscr{T}$ forms a subspace of $\mathscr{V}$. This subspace, called the intersection of $\mathscr{S}$ and $\mathscr{T}$, is denoted by the symbol $\mathscr{S} \cap \mathscr{T}$. Here, the set of all vectors in $\mathscr{V}$ which can be expressed as $\boldsymbol{\alpha} + \boldsymbol{\beta}$ with $\boldsymbol{\alpha} \in \mathscr{S}$ and $\boldsymbol{\beta} \in \mathscr{T}$ also forms a subspace of $\mathscr{V}$, called the sum of $\mathscr{S}$ and $\mathscr{T}$ and denoted by the symbol $\mathscr{S} + \mathscr{T}$. We have the dimensional relation

$$d(\mathscr{S} + \mathscr{T}) + d(\mathscr{S} \cap \mathscr{T}) = d(\mathscr{S}) + d(\mathscr{T}). \tag{1.1.1}$$

Since the null vector is a necessary constituent of every subspace, $\mathscr{S} \cap \mathscr{T}$ will always contain the null vector. If $\mathscr{S} \cap \mathscr{T}$ is a single vector set consisting only of the null vector, $\mathscr{S}$ and $\mathscr{T}$ are said to be virtually disjoint and the sum of $\mathscr{S}$ and $\mathscr{T}$ is called the direct sum and denoted by the symbol $\mathscr{S} \oplus \mathscr{T}$.

Orthogonal Subspace. Let $\mathscr{V}$ be a vectorspace with a proper inner product defined for all pairs of vectors in $\mathscr{V}$. If $\mathscr{S}$ is a subspace of $\mathscr{V}$, the set of all vectors in $\mathscr{V}$ that are orthogonal to every vector in $\mathscr{S}$ forms a subspace $\mathscr{S}^{\perp}$ called the orthogonal complement of $\mathscr{S}$ (in $\mathscr{V}$). We have the dimensional equality

$$d(\mathscr{S}) + d(\mathscr{S}^{\perp}) = d(\mathscr{V}). \tag{1.1.2}$$

The orthogonal complement of $\mathscr{M}(\mathbf{A})$ in $\mathscr{E}^m$ is denoted by $\mathscr{O}(\mathbf{A})$. $\mathbf{A}^{\perp}$ denotes a matrix such that

$$\mathscr{M}(\mathbf{A}^{\perp}) = \mathscr{O}(\mathbf{A}). \tag{1.1.3}$$

Unless it is otherwise clear from the context, the columns of $\mathbf{A}^{\perp}$ are assumed to be linearly independent.

Projection Operator

A general treatment of projection operators is given in Chapter 5. However, a special class of projection operators needed in the discussion of generalized inverse of a matrix considered in Chapter 3 is described here. Let $\mathbf{A}$ be $m \times n$ matrix. We shall call a matrix $\mathbf{P_A}$ a projection operator onto $\mathscr{M}(\mathbf{A})$ with respect to a n.n.d. matrix $\mathbf{M}$ iff

$$\begin{gathered}\mathbf{P_A x} \in \mathscr{M}(\mathbf{A}), \qquad \forall\, \mathbf{x} \in \mathscr{E}^m \\ (\mathbf{x} - \mathbf{P_A x})^*\mathbf{M}(\mathbf{x} - \mathbf{P_A x}) \le (\mathbf{x} - \mathbf{Ay})^*\mathbf{M}(\mathbf{x} - \mathbf{Ay}), \qquad \forall\, \mathbf{x} \in \mathscr{E}^m, \mathbf{y} \in \mathscr{E}^n.\end{gathered} \tag{1.1.4}$$

It is easy to see that the conditions in (1.1.4) are equivalent to

$$\mathbf{P_A^* M P_A} = \mathbf{MP_A}, \qquad \mathbf{MP_A A} = \mathbf{MA}, \quad \text{and} \quad R(\mathbf{P_A}) = R(\mathbf{A}), \tag{1.1.5}$$

where $R(\mathbf{X})$ indicates the rank of the matrix $\mathbf{X}$. If $\mathbf{M} = \mathbf{I}$, the conditions (1.1.5) reduce to

$$\begin{gathered}\mathbf{P_A^2} = \mathbf{P_A}, \qquad \mathbf{P_A^*} = \mathbf{P_A}, \\ \mathbf{P_A A} = \mathbf{A}, \quad R(\mathbf{P_A}) = R(\mathbf{A}).\end{gathered} \tag{1.1.6}$$

If $\mathbf{M}$ is p.d., the conditions (1.1.5) reduce to

$$\mathbf{P}_\mathbf{A}^2 = \mathbf{P}_\mathbf{A}, \qquad (\mathbf{MP}_\mathbf{A})^* = \mathbf{MP}_\mathbf{A},$$
$$\mathbf{P}_\mathbf{A}\mathbf{A} = \mathbf{A}, \qquad R(\mathbf{P}_\mathbf{A}) = R(\mathbf{A}). \tag{1.1.7}$$

Adjoint Matrix (Transformation)

Let $\mathbf{A}$ be a $m \times n$ matrix or a transformation mapping $\mathscr{E}^n$ into $\mathscr{E}^m$. If $\mathbf{x} \in \mathscr{E}^n$, then the transformation is written $\mathbf{y} = \mathbf{Ax}$, $\mathbf{y} \in \mathscr{E}^m$. Let $(\cdot,\cdot)_m$ and $(\cdot,\cdot)_n$ denote inner products in $\mathscr{E}^m$ and $\mathscr{E}^n$, respectively. The adjoint matrix of $\mathbf{A}$, denoted by $\mathbf{A}^\#$, is defined by the relation

$$(\mathbf{Ax}, \mathbf{z})_m = (\mathbf{x}, \mathbf{A}^\#\mathbf{z})_n \text{ for all } \mathbf{x} \in \mathscr{E}^n,\ \mathbf{z} \in \mathscr{E}^m. \tag{1.1.8}$$

By definition, if $\mathbf{A}$ is a $m \times n$ matrix, then $\mathbf{A}^\#$ is a $n \times m$ matrix.

Note that, if $(\mathbf{y}_1, \mathbf{y}_2)_m = \mathbf{y}_2^*\mathbf{M}\mathbf{y}_1$, $(\mathbf{x}_1, \mathbf{x}_2)_n = \mathbf{x}_2^*\mathbf{N}\mathbf{x}_1$, where $\mathbf{M}$ and $\mathbf{N}$ are p.d. matrices, then $\mathbf{A}^*\mathbf{M} = \mathbf{NA}^\#$ or $\mathbf{A}^\# = \mathbf{N}^{-1}\mathbf{A}^*\mathbf{M}$. (If $\mathbf{M}$ and $\mathbf{N}$ are identity matrices of order m and n, respectively, then $\mathbf{A}^\# = \mathbf{A}^*$.)

From (1.1.8)

$$(\mathbf{A}^\#\mathbf{z}, \mathbf{x})_n = (\mathbf{z}, \mathbf{Ax})_m \quad \text{so that } (\mathbf{A}^\#)^\# = \mathbf{A}. \tag{1.1.9}$$

1.2 CANONICAL FORMS OF MATRICES

In this section we consider canonical reduction of matrices into simpler forms by post- and premultiplication. For proofs of most of the results mentioned reference may be made to Section 1b.2 of Rao (1965).

Hermite Canonical Form

A square matrix $\mathbf{H}$ is said to be in Hermite canonical form if its principal diagonal consists of only zeros and unities and all subdiagonal elements are zero such that when a diagonal element is zero the entire row is zero, and when a diagonal element is unity the rest of the elements in the column are zero. Alternatively $\mathbf{H}$ is in the Hermite canonical form if there exists a permutation matrix $\mathbf{P}$ such that

$$\mathbf{PHP}' = \begin{pmatrix} \mathbf{I}_r & \mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $r = R(\mathbf{H})$ and $\mathbf{B}$ could be arbitrary. A matrix $\mathbf{H}$ in Hermite canonical form is necessarily idempotent (i.e., $\mathbf{H}^2 = \mathbf{H}$).

Let $\mathbf{A}$ be a square matrix of order m. Then there exists a nonsingular matrix $\mathbf{C}$ of order m such that

$$\mathbf{CA} = \mathbf{H}, \tag{1.2.1}$$

where $\mathbf{H}$ is in Hermite canonical form.

Diagonal Reduction

Let $\mathbf{A}$ be $m \times n$ matrix of rank r. Then there exist nonsingular square matrices $\mathbf{B}$ and $\mathbf{C}$ such that

$$\mathbf{BAC} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{1.2.2}$$

which give the representations

$$\mathbf{A} = \mathbf{B}^{-1}\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{C}^{-1}$$

$$\mathbf{A} = \mathbf{DE} = \partial_1\boldsymbol{\varepsilon}_1' + \cdots + \partial_r\boldsymbol{\varepsilon}_r', \tag{1.2.3}$$

where $\mathbf{D}$ is $m \times r$ matrix of rank r consisting of the first r column vectors $\partial_1, \ldots, \partial_r$ of $\mathbf{B}^{-1}$ and $\mathbf{E}$ is $r \times n$ matrix of rank r consisting of the first r row vectors $\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_r'$ of $\mathbf{C}^{-1}$. The representation (1.2.3) is called the *rank factorization* of $\mathbf{A}$.

Householder's Transformation

Triangular Reduction. Let $\mathbf{A}$ be a $m \times n$ matrix and $m \geq n$. Then there exists a unitary matrix $\mathbf{B}$ of order m such that

$$\mathbf{BA} = \begin{pmatrix} \mathbf{T} \\ \mathbf{0} \end{pmatrix}, \tag{1.2.4}$$

where $\mathbf{T}$ is an upper triangular matrix of order n and $\mathbf{0}$ is a null matrix of order $(m - n) \times n$.

Bidiagonalization. Let $\mathbf{A}$ be $m \times n$ matrix and $m \geq n$. Then there exist unitary matrices $\mathbf{B}$ and $\mathbf{C}$ such that $\mathbf{BAC}$ is in bidiagonal form, that is, all the elements of $\mathbf{BAC}$ are zero except possibly those in the main diagonal and the one above (or below) it.

Spectral Decomposition

Hermitian Matrix. Let $\mathbf{A}$ be $k \times k$ hermitian matrix (i.e., $\mathbf{A} = \mathbf{A}^*$). Then there exists a unitary matrix $\mathbf{U}$ such that

$$\mathbf{A} = \mathbf{U}^*\boldsymbol{\Delta}\mathbf{U}, \tag{1.2.5}$$

where $\boldsymbol{\Delta}$ is a real diagonal matrix. If $\delta_1, \ldots, \delta_k$ are diagonal elements of $\boldsymbol{\Delta}$, (1.2.5) can also be written as

$$\mathbf{A} = \delta_1\mathbf{P}_1 + \cdots + \delta_k\mathbf{P}_k, \tag{1.2.6}$$

where $\mathbf{P}_i^2 = \mathbf{P}_i$, $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$ for $i \neq j$ and $\mathbf{P}_1 + \cdots + \mathbf{P}_k = \mathbf{I}$.

Normal Matrix. Let $\mathbf{A}$ be $n \times n$ normal matrix (i.e., $\mathbf{AA}^* = \mathbf{A}^*\mathbf{A}$). Then there exists a unitary matrix $\mathbf{U}$ such that

$$\mathbf{A} = \mathbf{U}^*\mathbf{\Delta U}, \tag{1.2.7}$$

where $\mathbf{\Delta}$ is a diagonal matrix. If $\delta_1, \ldots, \delta_k$ are distinct diagonal elements of $\mathbf{\Delta}$, then (1.2.7) can also be written as

$$\mathbf{A} = \delta_1\mathbf{P}_1 + \cdots + \delta_k\mathbf{P}_k, \tag{1.2.8}$$

where $\mathbf{P}_i^2 = \mathbf{P}_i, \mathbf{P}_i\mathbf{P}_j = \mathbf{0}$ for $i \neq j$ and $\mathbf{P}_1 + \cdots + \mathbf{P}_k = \mathbf{I}$. Thus a normal matrix is unitary congruent to a diagonal matrix.

Commuting Hermitian Matrices. Let $\mathbf{A}_1$ and $\mathbf{A}_2$ be two hermitian matrices such that $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1$. Then there exists a unitary matrix $\mathbf{U}$ such that

$$\mathbf{A}_1 = \mathbf{U}^*\mathbf{\Delta}_1\mathbf{U} \qquad \mathbf{A}_2 = \mathbf{U}^*\mathbf{\Delta}_2\mathbf{U}, \tag{1.2.9}$$

where $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ are diagonal matrices.

Singular Value Decomposition. Let $\mathbf{A}$ be a $m \times n$ matrix of rank r. Then there exist two unitary matrices $\mathbf{U}$ of order m and $\mathbf{V}$ of order n such that

$$\mathbf{U}^*\mathbf{AV} = \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{1.2.10}$$

where $\mathbf{\Delta}$ is a diagonal matrix of rank r with real elements, all of which are positive.

From (1.2.10) we have

$$\mathbf{A} = \mathbf{E\Delta F}^*, \tag{1.2.11}$$

where $\mathbf{E}^*\mathbf{E} = \mathbf{I}_r = \mathbf{F}^*\mathbf{F}$. If $\delta_1, \ldots, \delta_r$ (not all of which need be distinct) are the diagonal elements of $\mathbf{\Delta}$, $\mathbf{e}_1, \ldots, \mathbf{e}_r$ are the columns of $\mathbf{E}$, and $\mathbf{f}_1, \ldots, \mathbf{f}_r$ are the columns of $\mathbf{F}$, then (1.2.11) can be written as

$$\mathbf{A} = \delta_1\mathbf{e}_1\mathbf{f}_1^* + \cdots + \delta_r\mathbf{e}_r\mathbf{f}_r^*. \tag{1.2.12}$$

We note that $\delta_1^2, \ldots, \delta_r^2$ are the common positive eigenvalues of $\mathbf{AA}^*$ and $\mathbf{A}^*\mathbf{A}$, $\mathbf{e}_i$ is the eigenvector of $\mathbf{AA}^*$ corresponding to δ_i^2, and $\mathbf{f}_i$ is the eigenvector of $\mathbf{A}^*\mathbf{A}$ corresponding to δ_i^2. The vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$ are orthonormal and so are $\mathbf{f}_1, \ldots, \mathbf{f}_r$.

Simultaneous Singular Value Decomposition. If $\mathbf{A}_1$ and $\mathbf{A}_2$ are two $m \times n$ matrices, there exist two unitary matrices $\mathbf{U}$ and $\mathbf{V}$ such that

$$\mathbf{U}^*\mathbf{A}_1\mathbf{V} = \begin{pmatrix} \mathbf{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \qquad \mathbf{U}^*\mathbf{A}_2\mathbf{V} = \begin{pmatrix} \mathbf{\Delta}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{1.2.13}$$

where Δ_1 and Δ_2 are both diagonal matrices with real elements; Δ_1 has no negative elements if and only if $\mathbf{A}_1\mathbf{A}_2^*$ and $\mathbf{A}_1^*\mathbf{A}_2$ are both hermitian matrices.

***Singular Value Decomposition with Respect to* M *and* N.** Let **A** be a $m \times n$ matrix of rank r and **M** and **N** be p.d. matrices of orders m and n, respectively. Then **A** can be expressed in the form

$$\mathbf{MAN} = \mu_1 \boldsymbol{\xi}_1 \boldsymbol{\eta}_1^* + \cdots + \mu_r \boldsymbol{\xi}_r \boldsymbol{\eta}_r^*, \tag{1.2.14}$$

where $\boldsymbol{\xi}_i^* \mathbf{M}^{-1} \boldsymbol{\xi}_j = 0$ for $i \neq j$ and $= 1$ for $i = j$, and $\boldsymbol{\eta}_i^* \mathbf{N}^{-1} \boldsymbol{\eta}_j = 0$ for $i \neq j$ and $= 1$ for $i = j$.

In (1.2.14) $\mu_1^2, \ldots, \mu_r^2$ are the nonzero eigenvalues of $\mathbf{A}^*\mathbf{MA}$ with respect to $\mathbf{N}^{-1}$ or of $\mathbf{ANA}^*$ with respect to $\mathbf{M}^{-1}$, $\boldsymbol{\xi}_i$ is the eigenvector of $\mathbf{ANA}^*$ with respect to $\mathbf{M}^{-1}$ corresponding to the eigenvalue μ_i^2, and $\boldsymbol{\eta}_i$ is the eigenvector of $\mathbf{A}^*\mathbf{MA}$ with respect to $\mathbf{N}^{-1}$ corresponding to the eigenvalue μ_i^2.

Polar Reduction. Let **A** be a square matrix. Then there exists a n.n.d. matrix **G** such that

$$\mathbf{A} = \mathbf{HG}, \qquad \mathbf{H} \text{ unitary}. \tag{1.2.15}$$

In fact, **G** is the hermitian square root of $\mathbf{A}^*\mathbf{A}$. **H** is unique if $|\mathbf{A}| \neq 0$.

Similarly, $\mathbf{A} = \mathbf{FH}$, where **H** is unitary and **F** is the hermitian square root of $\mathbf{AA}^*$.

Polar Representations. A complex orthogonal matrix **M** can always be represented in the form

$$\mathbf{M} = \mathbf{R}e^{i\mathbf{K}}, \tag{1.2.16}$$

where **R** is real orthogonal and **K** is real antisymmetric. A unitary matrix **U** can always be represented as

$$\mathbf{U} = \mathbf{R}e^{i\mathbf{S}}, \tag{1.2.17}$$

where **R** is real orthogonal and **S** is real symmetric.

1.3 CHARACTERISTIC FUNCTION, MINIMUM POLYNOMIALS

With every square matrix **A** of order n is associated the matric polynomial

$$\lambda\mathbf{I} - \mathbf{A}, \tag{1.3.1}$$

which is called the characteristic matrix of **A**. Its determinant is a polynomial in λ:

$$f(\lambda) = |\lambda\mathbf{I} - \mathbf{A}| = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0 \tag{1.3.2}$$

called the characteristic function of **A**. The equation $f(\lambda) = 0$ is called the