# A MATRIX HANDBOOK FOR STATISTICIANS

**George A. F. Seber**
Department of Statistics
University of Auckland
Auckland, New Zealand

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

**WILEY-INTERSCIENCE**
**A John Wiley & Sons, Inc., Publication**

This Page Intentionally Left Blank

# A MATRIX HANDBOOK FOR STATISTICIANS

BICENTENNIAL
1807 WILEY 2007
BICENTENNIAL

## THE WILEY BICENTENNIAL–KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

William J. Pesce
WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

Peter Booth Wiley
PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

# A MATRIX HANDBOOK FOR STATISTICIANS

**George A. F. Seber**
Department of Statistics
University of Auckland
Auckland, New Zealand

**WILEY-INTERSCIENCE**
**A John Wiley & Sons, Inc., Publication**

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



Wiley Bicentennial Logo: Richard J. Pacifico

***Library of Congress Cataloging-in-Publication Data:***

Seber, G. A. F. (George Arthur Frederick), 1938–
A matrix handbook for statisticians / George A.F. Seber.
p.; cm.
Includes bibliographical references and index.
ISBN 978-0-471-74869-4 (cloth )
1. Matrices. 2. Statistics. I. Title.
QA188.S43 2007
512.9'434—dc22 2007024691

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

# CONTENTS

# PREFACE

This book has had a long gestation period; I began writing notes for it in 1984 as a partial distraction when my first wife was fighting a terminal illness. Although I continued to collect material on and off over the years, I turned my attention to writing in other fields instead. However, in my recent "retirement", I finally decided to bring the book to birth as I believe even more strongly now of the need for such a book. Vectors and matrices are used extensively throughout statistics, as evidenced by appendices in many books (including some of my own), in published research papers, and in the extensive bibliography of Puntanen et al. [1998]. In fact, C. R. Rao [1973a] devoted his first chapter to the topic in his pioneering book, which many of my generation have found to be a very useful source. In recent years, a number of helpful books relating matrices to statistics have appeared on the scene that generally assume no knowledge of matrices and build up the subject gradually. My aim was not to write such a how-to-do-it book, but simply to provide an extensive list of results that people could look up – very much like a dictionary or encyclopedia. I therefore assume that the reader already has a basic working knowledge of vectors and matrices. Alhough the book title suggests a statistical orientation, I hope that the book's wide scope will make it useful to people in other disciplines as well.

In writing this book, I faced a number of challenges. The first was what to include. It was a bit like writing a dictionary. When do you stop adding material; I guess when other things in life become more important! The temptation was to begin including almost every conceiveble matrix result I could find on the grounds that one day they might all be useful in statistical research! After all, the history of science tells us that mathematical theory usually precedes applications. However, this is not practical and my selection is therefore somewhat personal and reflects my own general knowledge, or lack of it! Also, my selection is tempered by my ability to access certain books and journals, so overall there is a fair dose of randomness in the selection process. To help me keep my feet on the ground and keep my focus on statistics, I have listed, where possible, some references to statistical applications of the theory. Clearly, readers will spot some gaps and I apologize in advance for leaving out any of your favorite results or topics. Please let me know about them (e-mail: seber@stat.auckland.ac.nz). A helpful source of matrix definitions is the free encyclopedia, wikipedia at http://en.wikipedia.org.

My second challenge was what to do about proofs. When I first started this project, I began deriving and collecting proofs but soon realized that the proofs would make the book too big, given that I wanted the book to be reasonably comprehensive. I therefore decided to give only references to proofs at the end of each section or subsection. Most of the time I have been able to refer to book sources, with the occasional journal article referenced, and I have tried to give more than one reference for a result when I could. Although there are many excellent matrix books that I could have used for proofs, I often found in consulting a book that a particular result that I wanted was missing or perhaps assigned to the exercises, which often didn't have outline solutions. To avoid casting my net too widely, I have therefore tended to quote from books that are more encyclopedic in nature. Occasionally, there are lesser known results that are simply quoted without proof in the source that I have used, and I then use the words "Quoted by ..."; the reader will need to consult that source for further references to proofs. Some of my references are to exercises, and I have endeavored to choose sources that have at least outline solutions (e.g., Rao and Bhimasankaram [2000] and Seber [1984]) or perhaps some hints (e.g., Horn and Johnson [1985, 1991]); several books have solutions manuals (e.g., Harville [2001] and Meyer [2000b]). Sometimes I haven't been able to locate the proof of a fairly of straightforward result, and I have found it quicker to give an outline proof that I hope is sufficient for the reader.

In relation to proofs, there is one other matter I needed to deal with. Initially, I wanted to give the original references to important results, but found this too difficult for several reasons. Firstly, there is the sheer volume of results, combined with my limited access to older documents. Secondly, there is often controversy about the original authors. However, I have included some names of original authors where they seem to be well established. We also need to bear in mind Stigler's maxim, simply stated, that "no scientific discovery is named after its original discoverer." (Stigler [1999: 277]). It should be noted that there are also statistical proofs of some matrix results (cf. Rao [2000]).

The third challenge I faced was choosing the order of the topics. Because this book is not meant to be a teach-yourself matrix book, I did not have to follow a "logical" order determined by the proofs. Instead, I was able to collect like results together for an easier look-up. In fact, many topics overlap, so that a logical order is not completely possible. A disadvantage of such an approach is that concepts are sometimes mentioned before they are defined. I don't believe this will cause any difficulties because the cross-referencing and the index will, hopefully, be sufficiently detailed for definitions to be readily located.

My fourth challenge was deciding what level of generality I should use. Some authors use a general field for elements of matrices, while others work in a framework of complex matrices, because most results for real matrices follow as a special case.

Most books with the word "statistics" in the title deal with real matrices only. Although the complex approach would seem the most logical, I am aware that I am writing mainly for the research statistician, many of whom are not involved with complex matrices. I have therefore used a mixed approach with the choice depending on the topic and the proofs available in the literature. Sometimes I append the words "real case" or "complex case" to a reference to inform the reader about the nature of the proof referenced. Frequently, proofs relating to real matrices can be readily extended with little change to those for the complex case.

In a book of this size, it has not been possible to check the correctness of all the results quoted. However, where a result appears in more than one reference, one would have confidence in its accuracy. My aim has been been to try and faithfully reproduce the results. As we know with data, there is always a percentage that is either wrong or incorrectly transcribed. This book won't be any different. If you do find a typo, I would be grateful if you could e-mail me so that I can compile a list of errata for distribution.

With regard to contents, after some notation in Chapter 1, Chapter 2 focuses on vector spaces and their properties, especially on orthogonal complements and column spaces of matrices. Inner products, orthogonal projections, metrics, and convexity then take up most of the balance of the chapter. Results relating to the rank of a matrix take up all of Chapter 3, while Chapter 4 deals with important matrix functions such as inverse, transpose, trace, determinant, and norm. As complex matrices are sometimes left out of books, I have devoted Chapter 5 to some properties of complex matrices and then considered Hermitian matrices and some of their close relatives.

Chapter 6 is devoted to eigenvalues and eigenvectors, singular values, and (briefly) antieigenvalues. Because of the increasing usefulness of generalized inverses, Chapter 7 deals with various types of generalized inverses and their properties. Chapter 8 is a bit of a potpourri; it is a collection of various kinds of special matrices, except for those specifically highlighted in later chapters such as non-negative matrices in Chapter 9 and positive and non-negative definite matrices in Chapter 10. Some special products and operators are considered in Chapter 11, including (a) the Kronecker, Hadamard, and Rao–Khatri products and (b) operators such as the vec, vech, and vec-permutation (commutation) operators. One could fill several books with inequalities so that in Chapter 12 I have included just a selection of results that might have some connection with statistics. The solution of linear equations is the topic of Chapter 13, while Chapters 14 and 15 deal with partitioned matrices and matrices with a pattern.

A wide variety of factorizations and decompositions of matrices are given in Chapter 16, and in Chapter 17 and 18 we have the related topics of differentiation and Jacobians. Following limits and sequences of matrices in Chapter 19, the next three chapters involve random variables - random vectors (Chapter 20), random matrices (Chapter 21), and probability inequalities (Chapter 22). A less familiar topic, namely majorization, is considered in Chapter 23, followed by aspects of optimization in the last chapter, Chapter 24.

I want to express my thanks to a number of people who have provided me with preprints, reprints, reference material and answered my queries. These include Harold Henderson, Nye John, Simo Puntanen, Jim Schott, George Styan, Gary Tee, Goëtz Trenkler, and Yongge Tian. I am sorry if I have forgotten anyone because of the length of time since I began this project. My thanks also go to several anonymous referees who provided helpful input on an earlier draft of the book, and to the Wiley team for their encouragement and support. Finally, special thanks go to my wife Jean for her patient support throughout this project.

GEORGE A. F. SEBER

*Auckland, New Zealand*
*Setember 2007*

This Page Intentionally Left Blank

# CHAPTER 1

# NOTATION

## 1.1 GENERAL DEFINITIONS

Vectors and matrices are denoted by boldface letters $\mathbf{a}$ and $\mathbf{A}$, respectively, and scalars are denoted by italics. Thus $\mathbf{a} = (a_i)$ is a vector with $i$th element $a_i$ and $\mathbf{A} = (a_{ij})$ is a matrix with $i,j$th elements $a_{ij}$. I maintain this notation even with random variables, because using uppercase for random variables and lowercase for their values can cause confusion with vectors and matrices. In Chapters 20 and 21, which focus on random variables, we endeavor to help the reader by using the latter half of the alphabet $u, v, \ldots, z$ for random variables and the rest of the alphabet for constants.

Let $\mathbf{A}$ be an $n_1 \times n_2$ matrix. Then any $m_1 \times m_2$ matrix $\mathbf{B}$ formed by deleting any $n_1 - m_1$ rows and $n_2 - m_2$ columns of $\mathbf{A}$ is called a *submatrix* of $\mathbf{A}$. It can also be regarded as the intersection of $m_1$ rows and $m_2$ columns of $\mathbf{A}$. I shall define $\mathbf{A}$ to be a submatrix of itself, and when this is not the case I refer to a submatrix that is not $\mathbf{A}$ as a *proper submatrix* of $\mathbf{A}$. When $m_1 = m_2 = m$, the square matrix $\mathbf{B}$ is called a *principal submatrix* and it is said to be of *order m*. Its determinant, $\det(\mathbf{B})$, is called an *mth-order minor* of $\mathbf{A}$. When $\mathbf{B}$ consists of the intersection of the same numbered rows and columns (e.g., the first, second, and fourth), the minor is called a *principal minor*. If $\mathbf{B}$ consists of the intersection of the first $m$ rows and the first $m$ columns of $\mathbf{A}$, then it is called a *leading principal submatrix* and its determinant is called a *leading principal m-th order* minor.

*A Matrix Handbook for Statisticians.* By George A. F. Seber

Many matrix results hold when the elements of the matrices belong to a general field $\mathcal{F}$ of scalars. For most practitioners, this means that the elements can be real or complex, so we shall use $\mathbb{F}$ to denote either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. The expression $\mathbb{F}^n$ will denote the $n$-dimensional counterpart.

If $\mathbf{A}$ is complex, it can be expressed in the form $\mathbf{A} = \mathbf{B} + i\mathbf{C}$, where $\mathbf{B}$ and $\mathbf{C}$ are real matrices, and its *complex conjugate* is $\overline{\mathbf{A}} = \mathbf{B} - i\mathbf{C}$. We call $\mathbf{A}' = (a_{ji})$ the *transpose* of $\mathbf{A}$ and define the *conjugate transpose* of $\mathbf{A}$ to be $\mathbf{A}^* = \overline{\mathbf{A}}'$. In practice, we can often transfer results from real to complex matrices, and vice versa, by simply interchanging $'$ and $*$.

When adding or multiplying matrices together, we will assume that the sizes of the matrices are such that these operations can be carried out. We make this assumption by saying that the matrices are *conformable*. If there is any ambiguity we shall denote an $m \times n$ matrix $\mathbf{A}$ by $\mathbf{A}_{m\times n}$. A matrix partitioned into blocks is called a block matrix.

If $x$ and $y$ are random variables, then the symbols $\mathrm{E}(y)$, $\mathrm{var}(y)$, $\mathrm{cov}(x, y)$, and $\mathrm{E}(x \mid y)$ represent expectation, variance, covariance, and conditional expectation, respectively.

Before we give a list of all the symbols used we mention some univariate statistical distributions.

## 1.2 SOME CONTINUOUS UNIVARIATE DISTRIBUTIONS

We assume that the reader is familiar with the normal, chi-square, $t$, $F$, gamma, and beta univariate distributions. Multivariate vector versions of the normal and $t$ distributions are given in Sections 20.5.1 and 20.8.1, respectively, and matrix versions of the gamma and beta are found in Section 21.9. As some noncentral distributions are referred to in the statistical chapters, we define two univariate distributions below.

**1.1.** (Noncentral Chi-square Distribution) The random variable $x$ with probability density function

$$f(x) = \frac{1}{2^{\nu/2}} e^{-x^2/2} x^{(\nu/2)-1} \sum_{i=1}^{\infty} e^{-\delta/2} \left(\frac{\delta}{4}\right)^i \frac{1}{i,\, \Gamma(\frac{1}{2}\nu + i)} x^i$$

is called the *noncentral chi-square distribution* with $\nu$ degrees of freedom and noncentrality parameter $\delta$, and we write $x \sim \chi^2_\nu(\delta)$.

(a) When $\delta = 0$, the above density reduces to the (central) chi-square distribution, which is denoted by $\chi^2_\nu$.

(b) The noncentral chi-square can be defined as the distribution of the sum of the squares of independent univariate normal variables $y_i$ $(i = 1, 2, \ldots, n)$ with variances 1 and respective means $\mu_i$. Thus if $\mathbf{y} \sim N_d(\boldsymbol{\mu}, \mathbf{I}_d)$, the multivariate normal distribution, then $x = \mathbf{y}'\mathbf{y} \sim \chi^2_d(\delta)$, where $\delta = \boldsymbol{\mu}'\boldsymbol{\mu}$ (Anderson [2003: 81–82]).

(c) $E(x) = \nu + \delta$.

Since $\delta > 0$, some authors set $\delta = \tau^2$, say. Others use $\delta/2$, which, because of (c), is not so memorable.

**1.2.** (Noncentral $F$-Distribution) If $x \sim \chi_m^2(\delta)$, $y \sim \chi_n^2$, and $x$ and $y$ are statistically independent, then $F = (x/m)/(y/n)$ is said to have a noncentral $F$-distribution with $m$ and $n$ degrees of freedom, and noncentrality parameter $\delta$. We write $F \sim F_{m,n}(\delta)$. For a derivation of this distribution see Anderson [2003: 185]. When $\delta = 0$, we use the usual notation $F_{m,n}$ for the $F$-distribution.

## 1.3 GLOSSARY OF NOTATION

### Scalars

| | |
|---|---|
| $\mathcal{F}$ | field of scalars |
| $\mathbb{R}$ | real numbers |
| $\mathbb{C}$ | complex numbers |
| $\mathbb{F}$ | $\mathbb{R}$ or $\mathbb{C}$ |
| $z = x + iy$ | a complex number |
| $\overline{z} = x - iy$ | complex conjugate of $z$ |
| $\|z\| = (x^2 + y^2)^{1/2}$ | modulus of $z$ |

### Vector Spaces

| | |
|---|---|
| $\mathbb{F}^n$ | $n$-dimensional coordinate space |
| $\mathbb{R}^n$ | $\mathbb{F}^n$ with $\mathbb{F} = \mathbb{R}$ |
| $\mathbb{C}^n$ | $\mathbb{F}^n$ with $\mathbb{F} = \mathbb{C}$ |
| $\mathcal{C}(\mathbf{A})$ | column space of $\mathbf{A}$, the space spanned by the columns of $\mathbf{A}$ |
| $\mathcal{C}(\mathbf{A}')$ | row space of $\mathbf{A}$ |
| $\mathcal{N}(\mathbf{A})$ | $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}$, null space (kernel) of $\mathbf{A}$ |
| $\mathcal{S}(A)$ | span of the set $A$, the vector space of all linear combinations of vectors in $A$ |
| $\dim \mathcal{V}$ | dimension of the vector space $\mathcal{V}$ |
| $\mathcal{V}^\perp$ | the orthogonal complement of $\mathcal{V}$ |
| $\mathbf{x} \in \mathcal{V}$ | $\mathbf{x}$ is an element of $\mathcal{V}$ |
| $\mathcal{V} \subseteq \mathcal{W}$ | $\mathcal{V}$ is a subset of $\mathcal{W}$ |
| $\mathcal{V} \subset \mathcal{W}$ | $\mathcal{V}$ is a proper subset of $\mathcal{W}$ ( i.e., $\mathcal{V} \neq \mathcal{W}$) |
| $\mathcal{V} \cap \mathcal{W}$ | intersection, $\{\mathbf{x} : \mathbf{x} \in \mathcal{V} \text{ and } \mathbf{x} \in \mathcal{W}\}$ |
| $\mathcal{V} \cup \mathcal{W}$ | union, $\{\mathbf{x} : \mathbf{x} \in \mathcal{V} \text{ and/or } \mathbf{x} \in \mathcal{W}\}$ |
| $\mathcal{V} + \mathcal{W}$ | sum, $\{\mathbf{x} + \mathbf{y} : \mathbf{x} \in \mathcal{V}, \mathbf{y} \in \mathcal{W}\}$ |
| $\mathcal{V} \oplus \mathcal{W}$ | direct sum, $\{\mathbf{x} + \mathbf{y} : \mathbf{x} \in \mathcal{V}, \mathbf{y} \in \mathcal{W}; \mathcal{V} \cap \mathcal{W} = \mathbf{0}\}$ |
| $\langle , \rangle$ | an inner product defined on a vector space |
| $\mathbf{x} \perp \mathbf{y}$ | $\mathbf{x}$ is perpendicular to $\mathbf{y}$ (i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = 0$) |

## Complex Matrix

| | |
|---|---|
| $\mathbf{A} = \mathbf{B} + i\mathbf{C}$ | complex matrix, with $\mathbf{B}$ and $\mathbf{C}$ real |
| $\overline{\mathbf{A}} = (\overline{a}_{ij}) = \mathbf{B} - i\mathbf{C}$ | complex conjugate of $\mathbf{A}$ |
| $\mathbf{A}^* = \overline{\mathbf{A}}' = (\overline{a}_{ji})$ | conjugate transpose of $\mathbf{A}$ |
| $\mathbf{A} = \mathbf{A}^*$ | $\mathbf{A}$ is a Hermitian matrix |
| $\mathbf{A} = -\mathbf{A}^*$ | $\mathbf{A}$ is a skew-Hermitian matrix |
| $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$ | $\mathbf{A}$ is a normal matrix |

## Special Symbols

| | |
|---|---|
| sup | supremum |
| inf | infemum |
| max | maximum |
| min | minimum |
| $\rightarrow$ | tends to |
| $\Rightarrow$ | implies |
| $\propto$ | proportional to |
| $\mathbf{1}_n$ | the $n \times 1$ vector with unit elements |
| $\mathbf{I}_n$ | the $n \times n$ identity matrix |
| $\mathbf{0}$ | a vector or matrix of zeros |
| $\text{diag}(\mathbf{d})$ | $n \times n$ matrix with diagonal elements $\mathbf{d}' = (d_1, \ldots, d_n)$, and zeros elsewhere |
| $\text{diag}(d_1, d_2, \ldots, d_n)$ | same as above |
| $\text{diag}\,\mathbf{A}$ | diagonal matrix ; same diagonal elements as $\mathbf{A}$ |
| $\mathbf{A} \geq \mathbf{0}$ | the elements of $\mathbf{A}$ are all non-negative |
| $\mathbf{A} > \mathbf{0}$ | the elements of $\mathbf{A}$ are all positive |
| $\mathbf{A} \succeq \mathbf{0}$, n.n.d | $\mathbf{A}$ is non-negative definite ($\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$) |
| $\mathbf{A} \succeq \mathbf{B}$, $\mathbf{B} \preceq \mathbf{A}$ | $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$ |
| $\mathbf{A} \succ \mathbf{0}$, p.d. | $\mathbf{A}$ is positive definite ($\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$) |
| $\mathbf{A} \succ \mathbf{B}$, $\mathbf{B} \prec \mathbf{A}$ | $\mathbf{A} - \mathbf{B} \succ \mathbf{0}$ |
| $\mathbf{x} \ll \mathbf{y}$ | $\mathbf{x}$ is (strongly) majorized by $\mathbf{y}$ |
| $\mathbf{x} \ll_w \mathbf{y}$ | $\mathbf{x}$ is weakly submajorized by $\mathbf{y}$ |
| $\mathbf{x} \ll^w \mathbf{y}$ | $\mathbf{x}$ is weakly supermajorized by $\mathbf{y}$ |
| $\mathbf{A}' = (a_{ji})$ | the transpose of $\mathbf{A}$ |
| $\mathbf{A}^{-1}$ | inverse of $\mathbf{A}$ when $\mathbf{A}$ is nonsingular |
| $\mathbf{A}^-$ | weak inverse of $\mathbf{A}$ satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ |
| $\mathbf{A}^+$ | Moore-Penrose inverse of $\mathbf{A}$ |
| $\text{trace}\,\mathbf{A}$ | sum of the diagonal elements of a square matrix $\mathbf{A}$ |
| $\det \mathbf{A}$ | determinant of a square matrix $\mathbf{A}$ |
| $\text{rank}\,\mathbf{A}$ | rank of $\mathbf{A}$ |
| $\text{per}\ \mathbf{A}$ | permanent of a square matrix $\mathbf{A}$ |
| $\text{mod}(\mathbf{A})$ | modulus of $\mathbf{A} = (a_{ij})$, given by $(\lvert a_{ij})\rvert)$ |
| $\text{Pf}(\mathbf{A})$ | pfaffian of $\mathbf{A}$ |
| $\rho(\mathbf{A})$ | spectral radius of a square matrix $\mathbf{A}$ |
| $\kappa_v(\mathbf{A})$ | condition number of an $m \times n$ matrix, $v = 1, 2, \infty$ |

| | |
|---|---|
| $\langle \mathbf{x}, \mathbf{y} \rangle$ | inner product of $\mathbf{x}$ and $\mathbf{y}$ |
| $\|\mathbf{x}\|$ | a norm of vector $\mathbf{x}$ $(= \langle \mathbf{x}, \mathbf{x} \rangle^{1/2})$ |
| $\|\mathbf{x}\|_2$ | length of $\mathbf{x}$ $(= (\mathbf{x}^*\mathbf{x})^{1/2})$ |
| $\|\mathbf{x}\|_p$ | $L_p$ vector norm of $\mathbf{x}$ $(= \sum_{i=1}^n \|x_i\|^p)^{1/p})$ |
| $\|\mathbf{x}\|_\infty$ | $L_\infty$ vector norm of $\mathbf{x}$ $(= \max_i \|x_i\|)$ |
| $\|\mathbf{A}\|_p$ | a generalized matrix norm of $m \times n$ $\mathbf{A}$ $(= \sum_{i=1}^m \sum_{j=1}^n \|a_{ij}\|^p)^{1/p}, p \geq 1)$ |
| $\|\mathbf{A}\|_F$ | Frobenius norm of matrix $\mathbf{A}$ $(= (\sum_i \sum_j \|a_{ij}\|^2)^{1/2})$ |
| $\|\mathbf{A}\|_{v,in}$ | generalized matrix norm for $m \times n$ matrix $\mathbf{A}$ induced by a vector norm $\|\cdot\|_v$ |
| $\|\mathbf{A}\|_{ui}$ | unitarily invariant norm of $m \times n$ matrix $\mathbf{A}$ |
| $\|\mathbf{A}\|_{oi}$ | orthogonally invariant norm of $m \times n$ matrix $\mathbf{A}$ |
| $\|\|\|\mathbf{A}\|\|\|$ | matrix norm of square matrix $\mathbf{A}$ |
| $\|\|\|\mathbf{A}\|\|\|_{v,in}$ | matrix norm for a square matrix $\mathbf{A}$ induced by a vector norm $\|\cdot\|_v$ |
| $\mathbf{A}_{m\times n}$ | $m \times n$ matrix |
| $(\mathbf{A}, \mathbf{B})$ | matrix partitioned by two matrices $\mathbf{A}$ and $\mathbf{B}$ |
| $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ | matrix partitioned by column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ |
| $\mathbf{A} \otimes \mathbf{B}$ | Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ |
| $\mathbf{A} \circ \mathbf{B}$ | Hadamard (Schur) product of $\mathbf{A}$ and $\mathbf{B}$ |
| $\mathbf{A} \odot \mathbf{B}$ | Rao–Khatri product of $\mathbf{A}$ and $\mathbf{B}$ |
| $\mathrm{vec}\, \mathbf{A}_{m\times n}$ | $mn \times 1$ vector formed by writing the columns of $\mathbf{A}$ one below the other |
| $\mathrm{vech}\, \mathbf{A}_{m\times m}$ | $\frac{1}{2}m(m+1) \times 1$ vector formed by writing the columns of the lower triangle of $\mathbf{A}$ (including the diagonal elements) one below the other |
| $\mathbf{I}_{(m,n)}$ or $\mathbf{K}_{nm}$ | vec-permutation (commutation) matrix |
| $\mathbf{G}_n$ or $\mathbf{D}_n$ | duplication matrix |
| $\mathbf{P}_n$ or $\mathbf{N}_n$ | symmetrizer matrix |
| $\lambda(\mathbf{A})$ | eigenvalue of a square matrix $\mathbf{A}$ |
| $\sigma(\mathbf{B})$ | singular value of any matrix $\mathbf{B}$ |

This Page Intentionally Left Blank

# CHAPTER 2

# VECTORS, VECTOR SPACES, AND CONVEXITY

Vector spaces and subspaces play an important role in statistics, the key ones being orthogonal complements as well as the column and row spaces of matrices. Projections onto vector subspaces occur in topics like least squares, where orthogonality is defined in terms of an inner product. Convex sets and functions arise in the development of inequalities and optimization. Other topics such as metric spaces and coordinate geometry are also included in this chapter. A helpful reference for vector spaces and their properties is Kollo and von Rosen [2005: section 1.2].

## 2.1 VECTOR SPACES

### 2.1.1 Definitions

***Definition* 2.1.** If $S$ and $T$ are subsets of some space $V$, then $S \cap T$ is called the *intersection* of $S$ and $T$ and is the set of all vectors in $V$ common to both $S$ and $T$. The *sum* of $S$ and $T$, written $S + T$, is the set of all vectors in $V$ that are a sum of a vector in $S$ and a vector in $T$. Thus

$$W = S + T = \{\mathbf{w} : \mathbf{w} = \mathbf{s} + \mathbf{t}, \mathbf{s} \in S \text{ and } \mathbf{t} \in T\}.$$

(In most applications $S$ and $T$ are vector subspaces, defined below.)

***Definition* 2.2.** A *vector space* $\mathcal{U}$ over a field $\mathcal{F}$ is a set of elements $\{\mathbf{u}\}$ called vectors and a set $\mathcal{F}$ of elements called scalars with four binary operations ($+$, $\cdot$, $\star$, and $\circ$) that satisfy the following axioms.

(1) $\mathcal{F}$ is a field with regard to the operations $+$ and $\cdot$.

(2) For all $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{U}$ we have the following:

(i) $\mathbf{u} \star \mathbf{v} \in \mathcal{U}$.

(ii) $\mathbf{u} \star \mathbf{v} = \mathbf{v} \star \mathbf{u}$.

(iii) $(\mathbf{u} \star \mathbf{v}) \star \mathbf{w} = \mathbf{u} \star (\mathbf{v} \star \mathbf{w})$ for all $\mathbf{w} \in \mathcal{U}$.

(iv) There is a vector $\mathbf{0} \in \mathcal{U}$, called the *zero vector*, such that $\mathbf{u} \star \mathbf{0} = \mathbf{u}$ for all $\mathbf{u} \in \mathcal{U}$.

(v) For each $\mathbf{u} \in \mathcal{U}$ there exists a vector $-\mathbf{u} \in \mathcal{U}$ such that $\mathbf{u} \star -\mathbf{u} = \mathbf{0}$.

(3) For all $\alpha$ and $\beta$ in $\mathcal{F}$ and all $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{U}$ we have:

(i) $\alpha \circ \mathbf{u} \in V$.

(ii) There exists an element in $\mathcal{F}$ called the *unit element* such that $1 \circ \mathbf{u} = \mathbf{u}$.

(iii) $(\alpha + \beta) \circ \mathbf{u} = (\alpha \circ \mathbf{u}) \star (\beta \circ \mathbf{u})$.

(iv) $\alpha \circ (\mathbf{u} \star \mathbf{v}) = (\alpha \circ \mathbf{u}) \star (\alpha \circ \mathbf{v})$.

(v) $(\alpha \cdot \beta) \circ \mathbf{u} = \alpha \circ (\beta \circ \mathbf{u})$.

We note from (2) that $\mathcal{U}$ is an abelian group under "$\star$". Also, we can replace "$\star$" by "$+$" and remove "$\cdot$" and "$\circ$" wihout any ambiguity. Thus (iv) and (v) of (3) above can be written as $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ and $(\alpha\beta)\mathbf{u} = \alpha(\beta\mathbf{u})$, which we shall do in what follows.

Normally $\mathcal{F} = \mathbb{F}$, where $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$. However, one field that has been useful in the construction of experimental designs such as orthogonal Latin squares, for example, is a finite field consisting of a finite number of elements. A finite field is known as a *Galois field.* The number of elements in any Galois field is $p^m$, where $p$ is a prime number and $m$ is a positive integer. For a brief discussion see Rao and Rao [1998: 6–10].

If $\mathcal{F}$ is a finite field, then a vector space $\mathcal{U}$ over $\mathcal{F}$ can be used to obtain a finite projective geometry with a finite set of elements or "points" $S$ and a collection of subsets of $S$ or "lines." By identifying a block with a "line" and a treatment with a "point," one can use the projective geometry to construct balanced incomplete block designs—as, for example, described by Rao and Rao [1998: 48–49].

For general, less abstract, references on this topic see Friedberg et al. [2003], Lay [2003], and Rao and Bhimasankaram [2000].

***Definition* 2.3.** A subset $\mathcal{V}$ of a vector space $\mathcal{U}$ that is also a vector space is called a *subspace* of $\mathcal{U}$.

**2.1.** $\mathcal{V}$ is a vector subspace if and only if $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathcal{V}$ for all $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{V}$ and all $\alpha$ and $\beta$ in $\mathcal{F}$. Setting $\alpha = \beta = 0$, we see that $\mathbf{0}$, the zero vector in $\mathcal{U}$, must belong to every vector subspace.

**2.2.** The set $\mathcal{V}$ of all $m \times n$ matrices over $\mathcal{F}$, along with the usual operations of addition and scalar multiplication, is a vector space. If $m = n$, the subset $\mathcal{A}$ of all symmetric matrices is a vector subspace of $\mathcal{V}$.

**Proofs.** Section 2.1.1.

2.1. Rao and Bhimasankaram [2000: 23].

2.2. Harville [1997: chapters 3 and 4].

### 2.1.2 Quadratic Subspaces

Quadratic subspaces arise in certain inferential problems such as the estimation of variance components (Rao and Rao [1998: chapter 13]). They also arise in testing multivariate linear hypotheses when the variance–covariance matrix has a certain structure or pattern (Rogers and Young [1978: 204] and Seeley [1971]). Klein [2004] considers their use in the design of mixture experiments.

***Definition* 2.4.** Suppose $\mathcal{B}$ is a subspace of $\mathcal{A}$, where $\mathcal{A}$ is the set of all $n \times n$ real symmetric matrices. If $\mathbf{B} \in \mathcal{B}$ implies that $\mathbf{B}^2 \in \mathcal{B}$, then $\mathcal{B}$ is called a *quadratic subspace* of $\mathcal{A}$.

**2.3.** If $\mathbf{A}_1$ and $\mathbf{A}_2$ are real symmetric idempotent matrices (i.e., $\mathbf{A}_i^2 = \mathbf{A}_i$) with $\mathbf{A}_1\mathbf{A}_2 = \mathbf{0}$, and $\mathcal{A}$ is the set of all real symmetric $n \times n$ matrices, then

$$\mathcal{B} = \{\alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2 : \alpha_1 \text{ and } \alpha_2 \text{ real}\},$$

is a quadratic subspace of $\mathcal{A}$.

**2.4.** If $\mathcal{B}$ is a quadratic subspace of $\mathcal{A}$, then the following hold.

(a) If $\mathbf{A} \in \mathcal{B}$, then the Moore–Penrose inverse $\mathbf{A}^+ \in \mathcal{B}$.

(b) If $\mathbf{A} \in \mathcal{B}$, then $\mathbf{A}\mathbf{A}^+ \in \mathcal{B}$.

(c) There exists a basis of $\mathcal{B}$ consisting of idempotent matrices.

**2.5.** The following statements are equivalent.

(1) $\mathcal{B}$ is a quadratic subspace of $\mathcal{A}$.

(2) If $\mathbf{A}, \mathbf{B} \in \mathcal{B}$, then $(\mathbf{A} + \mathbf{B})^2 \in \mathcal{B}$.

(3) If $\mathbf{A}, \mathbf{B} \in \mathcal{B}$, then $\mathbf{AB} + \mathbf{BA} \in \mathcal{B}$.

(4) If $\mathbf{A} \in \mathcal{B}$, then $\mathbf{A}^k \in \mathcal{B}$ for $k = 1, 2, \ldots$.

**2.6.** Let $\mathcal{B}$ be a quadratic subspace of $\mathcal{A}$. Then:

(a) If $\mathbf{A}, \mathbf{B} \in \mathcal{B}$, then $\mathbf{ABA} \in \mathcal{B}$.

(b) Let $\mathbf{A} \in \mathcal{B}$ be fixed and let $\mathcal{C} = \{\mathbf{ABA} : \mathbf{B} \in \mathcal{B}\}$. Then $\mathcal{C}$ is a quadratic subspace of $\mathcal{B}$.

(c) If $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{B}$, then $\mathbf{ABC} + \mathbf{CBA} \in \mathcal{B}$.

**Proofs.** Section 2.1.2.

2.3. This follows from the definition and noting that $\mathbf{A}_2\mathbf{A}_1 = \mathbf{0}$.

2.3 to 2.6. Rao and Rao [1998: 434–436, 440].

### 2.1.3 Sums and Intersections of Subspaces

***Definition* 2.5.** Let $\mathcal{V}$ and $\mathcal{W}$ be vector subspaces of a vector space $\mathcal{U}$. As with sets, we define $\mathcal{V} + \mathcal{W}$ to be the *sum* of the two vector subspaces. If $\mathcal{V} \cap \mathcal{W} = \mathbf{0}$ (some authors use $\{\mathbf{0}\}$), we say that $\mathcal{V}$ and $\mathcal{W}$ are *disjoint vector subspaces* (Harville [1997] uses the term "essentially disjoint"). Note that this differs from the notion of disjoint sets, namely $\mathcal{V} \cap \mathcal{W} = \phi$, which we will not need. When $\mathcal{V}$ and $\mathcal{W}$ are disjoint, we refer to the sum as a *direct sum* and write $\mathcal{V} \oplus \mathcal{W}$. Also $\mathcal{V} \cap \mathcal{W}$ is called the *intersection* of $\mathcal{V}$ and $\mathcal{W}$.

The ordered pair $(\cap, \subseteq)$ forms a lattice of subspaces so that lattice theory can be used to determine properties relating to the sum and intersection of subspaces. Kollo and von Rosen [2006: section 1.2] give detailed lists of such properties, and some of these are given below.

**2.7.** Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be vector subspaces of $\mathcal{U}$.

(a) $\mathcal{A} \cap \mathcal{B}$ and $\mathcal{A} + \mathcal{B}$ are vector subspaces. However, $\mathcal{A} \cup \mathcal{B}$ need not be a vector space. Here $\mathcal{A} \cap \mathcal{B}$ is the smallest subspace containing $\mathcal{A}$ and $\mathcal{B}$, and $\mathcal{A} + \mathcal{B}$ is the largest. Also $\mathcal{A} + \mathcal{B}$ is the smallest subspace containing $\mathcal{A} \cup \mathcal{B}$. By smallest subspace we mean one with the smallest dimension.

(b) If $\mathcal{U} = \mathcal{A} \oplus \mathcal{B}$, then every $\mathbf{u} \in \mathcal{U}$ can be expressed uniquely in the form $\mathbf{u} = \mathbf{a} + \mathbf{b}$, where $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$.

(c) $\mathcal{A} \cap (\mathcal{A} + \mathcal{B}) = \mathcal{A} + (\mathcal{A} \cap \mathcal{B}) = \mathcal{A}$.

(d) (Distributive)

(i) $\mathcal{A} \cap (\mathcal{B} + \mathcal{C}) \supseteq (\mathcal{A} \cap \mathcal{B}) + (\mathcal{A} \cap \mathcal{C})$.

(ii) $\mathcal{A} + (\mathcal{B} \cap \mathcal{C}) \subseteq (\mathcal{A} + \mathcal{B}) \cap (\mathcal{A} + \mathcal{C})$.

(e) In the following results we can interchange $+$ and $\cap$.

(i) $[\mathcal{A} \cap (\mathcal{B} + \mathcal{C})] + \mathcal{B} = [(\mathcal{A} + \mathcal{B}) \cap \mathcal{C}] + \mathcal{B}$.

(ii) $\mathcal{A} \cap [\mathcal{B} + (\mathcal{A} \cap \mathcal{C})] = (\mathcal{A} \cap \mathcal{B}) + (\mathcal{A} \cap \mathcal{C})$.

(iii) $\mathcal{A} \cap (\mathcal{B} + \mathcal{C}) = \mathcal{A} \cap [\mathcal{B} \cap (\mathcal{A} + \mathcal{C})] + \mathcal{C}$.

(iv) $(\mathcal{A} \cap \mathcal{B}) + (\mathcal{A} \cap \mathcal{C}) + (\mathcal{B} \cap \mathcal{C}) = [\mathcal{A} + (\mathcal{B} \cap \mathcal{C})] \cap [\mathcal{B} + (\mathcal{A} \cap \mathcal{C})]$.

(v) $\mathcal{A} \cap \mathcal{B} = [(\mathcal{A} \cap \mathcal{B}) + (\mathcal{A} \cap \mathcal{C})] \cap [(\mathcal{A} \cap \mathcal{B}) + (\mathcal{B} \cap \mathcal{C})]$.

**Proofs.** Section 2.1.3.

2.7a. Schott [2005: 68].

2.7b. Assume $\mathbf{u} = \mathbf{a}_1 + \mathbf{b}_1$ so that $\mathbf{a} - \mathbf{a}_1 = -(\mathbf{b} - \mathbf{b}_1)$, with the two vectors being in disjoint subspaces; hence $\mathbf{a} = \mathbf{a}_1$ and $\mathbf{b} = \mathbf{b}_1$.

2.7c–e. Kollo and von Rosen [2006: section 1.2].

2.7d. Harville [2001: 163, exercise 4].

### 2.1.4 Span and Basis

***Definition* 2.6.** We can always construct a vector space $\mathcal{U}$ from $\mathcal{F}$, called an *n-tuple space*, by defining $\mathbf{u} = (u_1, u_2, \ldots, u_n)'$, where each $u_i \in \mathcal{F}$.

In practice, $\mathcal{F}$ is usually $\mathbb{F}$ and $\mathcal{U}$ is $\mathbb{F}^n$. This will generally be the case in this book, unless indicated otherwise. However, one useful exception is the vector space consisting of all $m \times n$ matrices with elements in $\mathcal{F}$.

***Definition* 2.7.** Given a subset $A$ of a vector space $\mathcal{V}$, we define the *span* of $A$, denoted by $\mathcal{S}(A)$, to be the set of all vectors obtained by taking all linear combinations of vectors in $A$. We say that $A$ is a *generating set* of $\mathcal{S}(A)$.

**2.8.** Let $A$ and $B$ be subsets of a vector space. Then:

(a) $\mathcal{S}(A)$ is a vector space (even though $A$ may not be).

(b) $A \subseteq \mathcal{S}(A)$. Also $\mathcal{S}(A)$ is the smallest subspace of $\mathcal{V}$ containing $A$ in the sense that every subspace of $\mathcal{V}$ containing $A$ also contains $\mathcal{S}(A)$.

(c) $A$ is a vector space if and only if $A = \mathcal{S}(A)$.

(d) $\mathcal{S}[\mathcal{S}(A)] = \mathcal{S}(A)$.

(e) If $A \subseteq B$, then $\mathcal{S}(A) \subseteq \mathcal{S}(B)$.

(f) $\mathcal{S}(A) \cup \mathcal{S}(B) \subseteq \mathcal{S}(A \cup B)$.

(g) $\mathcal{S}(A \cap B) \subseteq \mathcal{S}(A) \cap \mathcal{S}(B)$.

***Definition* 2.8.** A set of vectors $\mathbf{v}_i$ $(i = 1, 2, \ldots, r)$ in a vector space are *linearly independent* if $\sum_{i=1}^r a_i \mathbf{v}_i = \mathbf{0}$ implies that $a_1 = a_2 = \cdots = a_r = 0$. A set of vectors that are not linearly independent are said to be *linearly dependent.* For further properties of linearly independent sets see Rao and Bhimasankaram [2000: chapter 1].

The term "vector" here and in the following definitions is quite general and simply refers to an element of a vector space. For example, it could be an $m \times n$ matrix in the vector space of all such matrices; Harville [1997: chapters 3 and 4] takes this approach.

***Definition* 2.9.** A set of vectors $\mathbf{v}_i$ $(i = 1, 2, \ldots, r)$ *span* a vector space $\mathcal{V}$ if the elements of $\mathcal{V}$ consist of all linear combinations of the vectors (i.e., if $\mathbf{v} \in \mathcal{V}$, then $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r$). The set of vectors is called a *generating set* of $\mathcal{V}$. If the vectors are also linearly independent, then the $\mathbf{v}_i$ form a *basis* for $\mathcal{V}$.

**2.9.** Every vector space has a basis. (This follows from Zorn's lemma, which can be used to prove the existence of a maximal linearly independent set of vectors, i.e., a basis.)

***Definition* 2.10.** All bases contain the same number of vectors so that this number is defined to be the dimension of $\mathcal{V}$.

**2.10.** Let $\mathcal{V}$ be a subspace of $\mathcal{U}$. Then:

(a) Every linearly independent set of vectors in $\mathcal{V}$ can be extended to a basis of $\mathcal{U}$.

(b) Every generating set of $\mathcal{V}$ contains a basis of $\mathcal{V}$.

**2.11.** If $\mathcal{V}$ and $\mathcal{W}$ are vector subspaces of $\mathcal{U}$, then:

(a) If $\mathcal{V} \subseteq \mathcal{W}$ and $\dim \mathcal{V} = \dim \mathcal{W}$, then $\mathcal{V} = \mathcal{W}$.

(b) If $\mathcal{V} \subseteq \mathcal{W}$ and $\mathcal{W} \subseteq \mathcal{V}$, then $\mathcal{V} = \mathcal{W}$. This is the usual method for proving the equality of two vector subspaces.

(c) $\dim(\mathcal{V} + \mathcal{W}) = \dim(\mathcal{V}) + \dim(\mathcal{W}) - \dim(\mathcal{V} \cap \mathcal{W})$.

**2.12.** If the columns of $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_r)$ and the columns of $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_r)$ both form a basis for a vector subspace of $\mathbb{F}^n$, then $\mathbf{A} = \mathbf{BR}$, where $\mathbf{R} = (r_{ij})$ is $r \times r$ and nonsingular.

**Proofs.** Section 2.1.4.

2.8. Rao and Bhimasankaram [2000: 25–28].

2.9. Halmos [1958].

2.10. Rao and Bhimasankaram [2000: 39].

2.11a–b. Proofs are straightforward.

2.11c. Meyer [2000a: 205] and Rao and Bhimasankaram [2000: 48].

2.12. Firstly, $\mathbf{a}_j = \sum_i \mathbf{b}_i r_{ij}$ so that $\mathbf{A} = \mathbf{BR}$. Now assume $\operatorname{rank} \mathbf{R} < r$; then $\operatorname{rank} \mathbf{A} \leq \min\{\operatorname{rank} \mathbf{B}, \operatorname{rank} \mathbf{R}\} < r$ by (3.12), which is a contradiction.

### 2.1.5 Isomorphism

***Definition* 2.11.** Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be two vector spaces over the same field $\mathcal{F}$. Then a map (function) $\phi$ from $\mathcal{V}_1$ to $\mathcal{V}_2$ is said to be an *isomorphism* if the following hold.

(1) $\phi$ is a bijection (i.e., $\phi$ is one-to-one and onto).

(2) $\phi(\mathbf{u} + \mathbf{v}) = \phi(\mathbf{u}) + \phi(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{V}_1$.

(3) $\phi(\alpha \mathbf{u}) = \alpha \phi(\mathbf{u})$ for all $\alpha \in \mathcal{F}$ and $\mathbf{u} \in \mathcal{V}_1$.

$\mathcal{V}_1$ is said to be *isomorphic* to $\mathcal{V}_2$ if there is an isomorphism from $\mathcal{V}_1$ to $\mathcal{V}_2$.

**2.13.** Two vector spaces over a field $\mathcal{F}$ are isomorphic if and only if they have the same dimension.

**Proofs.** Section 2.1.5.

2.13. Rao and Bhimasankaram [2000: 59].

## 2.2 INNER PRODUCTS

### 2.2.1 Definition and Properties

The concept of an inner product is an important one in statistics as it leads to ideas of length, angle, and distance between two points.

***Definition* 2.12.** Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ (i.e., $\mathbb{R}$ or $\mathbb{C}$), and let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be any vectors in $\mathcal{V}$. An inner product $\langle\cdot,\cdot\rangle$ defined on $\mathcal{V}$ is a function $\langle\mathbf{x},\mathbf{y}\rangle$ of two vectors $\mathbf{x},\mathbf{y} \in \mathcal{V}$ satisfying the following conditions:

(1) $\langle\mathbf{x},\mathbf{y}\rangle = \overline{\langle\mathbf{y},\mathbf{x}\rangle}$, the complex conjugate of $\langle\mathbf{y},\mathbf{x}\rangle$.

(2) $\langle\mathbf{x},\mathbf{x}\rangle \geq \mathbf{0}$; $\langle\mathbf{x},\mathbf{x}\rangle = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$.

(3) $\langle\alpha\mathbf{x},\mathbf{y}\rangle = \alpha\langle\mathbf{x},\mathbf{y}\rangle$, where $\alpha$ is a scalar in $\mathbb{F}$.

(4) $\langle\mathbf{x}+\mathbf{y},\mathbf{z}\rangle = \langle\mathbf{x},\mathbf{z}\rangle + \langle\mathbf{y},\mathbf{z}\rangle$.

When $\mathcal{V}$ is over $\mathbb{R}$, (1) becomes $\langle\mathbf{x},\mathbf{y}\rangle = \langle\mathbf{y},\mathbf{x}\rangle$, a symmetry condition. Inner products can also be defined on infinite-dimensional spaces such as a Hilbert space. A vector space together with an inner product is called an *inner product space.* A complex inner product space is also called a *unitary space*, and a real inner product space is called a *Euclidean space.*

The *norm* or *length* of $\mathbf{x}$, denoted by $\|\mathbf{x}\|$, is defined to be the positive square root of $\langle\mathbf{x},\mathbf{x}\rangle$. We say that $\mathbf{x}$ has *unit length* if $\|\mathbf{x}\| = 1$. More general norms, which are not associated with an inner product, are discussed in Section 4.6.

We can define the angle $\theta$ between $\mathbf{x}$ and $\mathbf{y}$ by

$$\cos\theta = \langle\mathbf{x},\mathbf{y}\rangle/(\|\mathbf{x}\|\|\mathbf{y}\|).$$

The *distance* between $\mathbf{x}$ and $\mathbf{y}$ is defined to be $d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$ and has the properties of a metric (Section 2.4). Usually, $\mathcal{V} = \mathbb{R}^n$ and $\langle\mathbf{x},\mathbf{y}\rangle = \mathbf{x}'\mathbf{y}$ in defining angle and distance.

Suppose (2) above is replaced by the weaker condition

(2$'$) $\langle\mathbf{x},\mathbf{x}\rangle \geq 0$. (It is now possible that $\langle\mathbf{x},\mathbf{x}\rangle = 0$, but $\mathbf{x} \neq \mathbf{0}$.)

We then have what is called a *semi-inner product* (quasi-inner product) and a corresponding *seminorm.* We write $\langle\mathbf{x},\mathbf{y}\rangle_s$ for a semi-inner product.

**2.14.** For any inner product the following hold:

(a) $\langle\mathbf{x},\alpha\mathbf{y}+\beta\mathbf{z}\rangle = \bar{\alpha}\langle\mathbf{x},\mathbf{y}\rangle + \bar{\beta}\langle\mathbf{x},\mathbf{z}\rangle$.

(b) $\langle\mathbf{x},\mathbf{0}\rangle = \langle\mathbf{0},\mathbf{x}\rangle = 0$.

(c) $\langle\alpha\mathbf{x},\beta\mathbf{y}\rangle = \alpha\langle\mathbf{x},\beta\mathbf{y}\rangle = \alpha\bar{\beta}\langle\mathbf{x},\mathbf{y}\rangle$.

**2.15.** The following hold for any norm associated with an inner product.

(a) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality).

(b) $\|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}\| \geq \|\mathbf{x}\|$.

(c) $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$ (parallelogram law).

(d) $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ (Pythagoras theorem).

(e) $\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle \leq 2\|\mathbf{x}\| \cdot \|\mathbf{y}\|$.

**2.16.** (Semi-Inner Product) The following hold for any semi-inner product $\langle \cdot , \cdot \rangle_s$ on a vector space $\mathcal{V}$.

(a) $\langle \mathbf{0}, \mathbf{0} \rangle_s = 0$

(b) $\|\mathbf{x}+\mathbf{y}\|_s \leq \|\mathbf{x}\|_s + \|\mathbf{y}\|_s$.

(c) $\mathcal{N} = \{\mathbf{x} \in \mathcal{V} : \|\mathbf{x}\|_s = 0\}$ is a subspace of $\mathcal{V}$.

**2.17.** (Schwarz Inequality) Given an inner product space, we have for all $\mathbf{x}$ and $\mathbf{y}$

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle,$$

or

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|,$$

with equality if either $\mathbf{x}$ or $\mathbf{y}$ is zero or $\mathbf{x} = k\mathbf{y}$ for some scalar $k$. We can obtain various inequalities from the above by changing the inner product space (cf. Section 12.1).

**2.18.** Given an inner product space and *unit* vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, then

$$\sqrt{1 - |\langle \mathbf{u}, \mathbf{v} \rangle|^2} \leq \sqrt{1 - |\langle \mathbf{u}, \mathbf{w} \rangle|^2} + \sqrt{1 - |\langle \mathbf{w}, \mathbf{v} \rangle|^2}.$$

Equality holds if and only if $\mathbf{w}$ is a multiple of $\mathbf{u}$ or of $\mathbf{v}$.

**2.19.** Some inner products are as follows.

(a) If $\mathcal{V} = \mathbb{R}^n$, then common inner products are:

(1) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}'\mathbf{x} = \sum_{i=1}^n x_i y_i$ $(= \mathbf{x}'\mathbf{y})$. If $\mathbf{x} = \mathbf{y}$, we denote the norm by $\|\mathbf{x}\|_2$, the so-called *Euclidean norm.*

The *minimal angle between two vector subspaces* $\mathcal{V}$ and $\mathcal{W}$ in $\mathbb{R}^n$ is given by

$$\cos\theta_{\min} = \max_{\mathbf{x}\in\mathcal{V}, \mathbf{y}\in\mathcal{W}} \frac{(\mathbf{x}'\mathbf{y})^2}{\|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2}.$$

For some properties see Meyer [2000a: section 5.15].

(2) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}'\mathbf{A}\mathbf{x}$ $(= \mathbf{x}'\mathbf{A}\mathbf{y})$, where $\mathbf{A}$ is a positive definite matrix.

(b) If $\mathcal{V} = \mathbb{C}^n$, then we can use $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^*\mathbf{x} = \sum_{i=1}^n x_i \overline{y}_i$.

(c) Every inner product defined on $\mathbb{C}^n$ can be expressed in the form $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^*\mathbf{A}\mathbf{x} = \sum_i \sum_j a_{ij} x_i \overline{y}_j$, where $\mathbf{A} = (a_{ij})$ is a Hermitian positive definite matrix. This follows by setting $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = a_{ij}$ for all $i, j$, where $\mathbf{e}_i$ is the $i$th column of $\mathbf{I}_n$. If we have a semi-inner product, then $\mathbf{A}$ is Hermitian non-negative definite. (This result is proved in Drygas [1970: 29], where symmetric means Hermitian.)

**2.20.** Let $\mathcal{V}$ be the set of all $m \times n$ real matrices, and in scalar multiplication all scalars belong to $\mathbb{R}$. Then:

(a) $\mathcal{V}$ is vector space.

(b) If we define $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}'\mathbf{B})$, then $\langle \, , \rangle$ is an inner product.

(c) The corresponding norm is $(\langle \mathbf{A}, \mathbf{A} \rangle)^{1/2} = (\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2)^{1/2}$. This is the so-called *Frobenius norm* $\|\mathbf{A}\|_F$ (cf. Definition 4.16 below (4.7)).

**Proofs.** Section 2.2.1.

2.14. Rao and Bhimasankaram [2000: 251–252].

2.15. We begin with the Schwarz inequality $|\langle \mathbf{x}, \mathbf{y} \rangle| = |\langle \mathbf{y}, \mathbf{x} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ of (2.17). Then, since $\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle$ is real,

$$\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle \leq |\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle| \leq |\langle \mathbf{x}, \mathbf{y} \rangle| + |\langle \mathbf{y}, \mathbf{x} \rangle| \leq 2\|\mathbf{x}\| \cdot \mathbf{y}\|,$$

which proves (e). We obtain (a) by writing $\|\mathbf{x} + \mathbf{y}\|^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle$ and using (e); the rest are straightforward. See also Rao and Rao [1998: 54].

2.16. Rao and Rao [1998: 77].

2.17. There are a variety of proofs (e.g., Schott [2005: 36] and Ben-Israel and Greville [2003: 7]). The inequality also holds for quasi-inner (semi-inner) products (Harville [1997: 255]).

2.18. Zhang [1999: 155].

2.20. Harville [1997: chapter 4] uses this approach.

### 2.2.2 Functionals

***Definition* 2.13.** A function $f$ defined on a vector space $\mathcal{V}$ over a field $\mathbb{F}$ and taking values in $\mathbb{F}$ is said to be a *linear functional* if

$$f(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) = \alpha_1 f(\mathbf{x}_1) + \alpha_2 f(\mathbf{x}_2)$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{V}$ and every $\alpha_1, \alpha_2 \in \mathbb{F}$. For a discussion of linear functionals and the related concept of a *dual space* see Rao and Rao [1998: section 1.7].

**2.21.** (Riesz) Let $\mathcal{V}$ be an an inner product space with inner product $\langle , \rangle$, and let $f$ be a linear functional on $\mathcal{V}$.

(a) There exists a unique vector $\mathbf{z} \in \mathcal{V}$ such that

$$f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{z} \rangle \text{ for every } \mathbf{x} \in \mathcal{V}.$$

(b) Here $\mathbf{z}$ is given by $\mathbf{z} = \overline{f(\mathbf{u})}\,\mathbf{u}$, where $\mathbf{u}$ is any vector of unit length in $\mathcal{V}^{\perp}$.

**Proofs.** Section 2.2.2.

2.21. Rao and Rao [1998: 71].

### 2.2.3 Orthogonality

***Definition* 2.14.** Let $\mathcal{U}$ be a vector space over $\mathbb{F}$ with an inner product $\langle\,,\rangle$, so that we have an inner product space. We say that $\mathbf{x}$ is *perpendicular* to $\mathbf{y}$, and we write $\mathbf{x} \perp \mathbf{y}$, if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

**2.22.** A set of vectors that are mutually orthogonal—that is, are pairwise orthogonal for every pair—are linearly independent.

***Definition* 2.15.** A basis whose vectors are mutually orthogonal with unit length is called an *orthonormal basis.* An orthonormal basis of an inner product space always exists and it can be constructed from any basis by the Gram–Schmidt orthogonalization process of (2.30).

**2.23.** Let $\mathcal{V}$ and $\mathcal{W}$ be vector subspaces of a vector space $\mathcal{U}$ such that $\mathcal{V} \subseteq \mathcal{W}$. Any orthonormal basis for $\mathcal{V}$ can be enlarged to form an orthonormal basis for $\mathcal{W}$.

***Definition* 2.16.** Let $\mathcal{U}$ be a vector space over $\mathbb{F}$ with an inner product $\langle\,,\rangle$, and let $\mathcal{V}$ be a subset or subspace of $\mathcal{U}$. Then the *orthogonal complement* of $\mathcal{V}$ with respect to $\mathcal{U}$ is defined to be

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

If $\mathcal{V}$ and $\mathcal{W}$ are two vector subspaces, we say that $\mathcal{V} \perp \mathcal{W}$ if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for all $\mathbf{x} \in \mathcal{V}$ and $\mathbf{y} \in \mathcal{W}$.

**2.24.** Suppose $\dim \mathcal{U} = n$ and $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n$ is an orthonormal basis of $\mathcal{U}$. If $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ $(r < n)$ is an orthonormal basis for a vector subspace $\mathcal{V}$ of $\mathcal{U}$, then $\boldsymbol{\alpha}_{r+1}, \ldots, \boldsymbol{\alpha}_n$ is an orthornormal basis for $\mathcal{V}^\perp$.

**2.25.** If $S$ and $T$ are subsets or subspaces of $\mathcal{U}$, then we have the following results:

(a) $S^\perp$ is a vector space.

(b) $S \subseteq (S^\perp)^\perp$ with equality if and only if $S$ is a vector space.

(c) If $S$ and $T$ both contain $\mathbf{0}$, then $(S + T)^\perp = S^\perp \cap T^\perp$.

**2.26.** If $\mathcal{V}$ is a vector subspace of $\mathcal{U}$, a vector space over $\mathbb{F}$, then:

(a) $\mathcal{V}^\perp$ is a vector subspace of $\mathcal{U}$, by (2.25a) above.

(b) $(\mathcal{V}^\perp)^\perp = \mathcal{V}$.

(c) $\mathcal{V} \oplus \mathcal{V}^\perp = \mathcal{U}$. In fact every $\mathbf{u} \in \mathcal{U}$ can be expressed uniquely in the form $\mathbf{u} = \mathbf{x} + \mathbf{y}$, where $\mathbf{x} \in \mathcal{V}$ and $\mathbf{y} \in \mathcal{V}^\perp$.

(d) $\dim(\mathcal{V}) + \dim(\mathcal{V}^\perp) = \dim(\mathcal{U})$.

**2.27.** If $\mathcal{V}$ and $\mathcal{W}$ are vector subspaces of $\mathcal{U}$, then:

(a) $\mathcal{V} \subseteq \mathcal{W}$ if and only if $\mathcal{V} \perp \mathcal{W}^\perp$.

(b) $\mathcal{V} \subseteq \mathcal{W}$ if and only if $\mathcal{W}^\perp \subseteq \mathcal{V}^\perp$.

(c) $(\mathcal{V} \cap \mathcal{W})^\perp = \mathcal{V}^\perp + \mathcal{W}^\perp$ and $(\mathcal{V} + \mathcal{W})^\perp = \mathcal{V}^\perp \cap \mathcal{W}^\perp$.

For more general results see Kollo and von Rosen [2005: section 1.2].

***Definition* 2.17.** Let $\mathcal{V}$ and $\mathcal{W}$ be vector subspaces of $\mathcal{U}$, a vector space over $\mathbb{F}$, and suppose that $\mathcal{V} \subseteq \mathcal{W}$. Then the set of all vectors in $\mathcal{W}$ that are perpendicular to $\mathcal{V}$ form a vector space called the *orthogonal complement* of $\mathcal{V}$ with respect to $\mathcal{W}$, and is denoted by $\mathcal{V}^\perp \cap \mathcal{W}$. Thus

$$\mathcal{V}^\perp \cap \mathcal{W} = \{\mathbf{w} : \mathbf{w} \in \mathcal{W}, \langle \mathbf{w}, \mathbf{v} \rangle = \mathbf{0} \text{ for every } \mathbf{v} \in \mathcal{V}\}.$$

**2.28.** Let $\mathcal{V} \subseteq \mathcal{W}$. Then

(a) (i) $\dim(\mathcal{V}^\perp \cap \mathcal{W}) = \dim(\mathcal{W}) - \dim(\mathcal{V})$.
(ii) $\mathcal{W} = \mathcal{V} \oplus (\mathcal{V}^\perp \cap \mathcal{W})$.

(b) From (a)(ii) we have $\mathcal{U} = \mathcal{W} \oplus \mathcal{W}^\perp = \mathcal{V} \oplus (\mathcal{V}^\perp \cap \mathcal{W}) \oplus \mathcal{W}^\perp$.
The above can be regarded as an orthogonal decomposition of $\mathcal{U}$ into three orthogonal subspaces. Using this, vectors can be added to any orthonormal basis of $\mathcal{V}$ to form an orthonormal basis of $\mathcal{W}$, which can then be extended to form an orthonormal basis of $\mathcal{U}$.

**2.29.** Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be vector subspaces of $\mathcal{U}$. If $\mathcal{B} \perp \mathcal{C}$ and $\mathcal{A} \perp \mathcal{C}$, then $\mathcal{A} \cap (\mathcal{B} \oplus \mathcal{C}) = \mathcal{A} \cap \mathcal{B}$.

**2.30.** (Classical Gram-Schmidt Algorithm) Given a basis $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ of an inner product space, there exists an orthonormal basis $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ given by $\mathbf{q}_1 = \mathbf{x}_1/\|\mathbf{x}_1\|$, $\mathbf{q}_j = \mathbf{w}_j/\|\mathbf{w}_j\|$ $(j = 2, \ldots, n)$, where

$$\mathbf{w}_j = \mathbf{x}_j - \langle \mathbf{x}_j, \mathbf{q}_1 \rangle \mathbf{q}_1 - \langle \mathbf{x}_j, \mathbf{q}_2 \rangle \mathbf{q}_2 - \cdots - \langle \mathbf{x}_j, \mathbf{q}_{j-1} \rangle \mathbf{q}_{j-1}.$$

This expression gives the algorithm for computing the basis. If we require an orthogonal basis only without the square roots involved with the normalizing, we can use $\mathbf{w}_1 = \mathbf{x}_1$ and, for $j = 2, 3, \ldots, n$,

$$\mathbf{w}_j = \mathbf{x}_j - \frac{\langle \mathbf{x}_j, \mathbf{w}_1 \rangle \mathbf{w}_1}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} - \cdots - \frac{\langle \mathbf{x}_j, \mathbf{w}_{j-1} \rangle \mathbf{w}_{j-1}}{\langle \mathbf{w}_{j-1}, \mathbf{w}_{j-1} \rangle}.$$

Also the vectors can be replaced by matrices using a suitable inner product such as $\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{trace}(\mathbf{A}'\mathbf{B})$.

**2.31.** Since, from (2.9), every vector space has a basis, it follows from the above algorithm that every inner product space has an orthonormal basis.

**2.32.** Let $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n\}$ be an orthonormal basis of $\mathcal{V}$, and let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ be any vectors. Then, for an inner product space:

(a) $\mathbf{x} = \langle \mathbf{x}, \boldsymbol{\alpha}_1 \rangle \boldsymbol{\alpha}_1 + \langle \mathbf{x}, \boldsymbol{\alpha}_2 \rangle \boldsymbol{\alpha}_2 + \cdots + \langle \mathbf{x}, \boldsymbol{\alpha}_n \rangle \boldsymbol{\alpha}_n$.

(b) (Parseval's identity) $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n \langle \mathbf{x}, \boldsymbol{\alpha}_i \rangle \langle \boldsymbol{\alpha}_i, \mathbf{y} \rangle$.

Conversely, if this equation holds for any $\mathbf{x}$ and $\mathbf{y}$, then $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ is an orthonormal basis for $\mathcal{V}$.

(c) Setting $\mathbf{x} = \mathbf{y}$ in (b) we have

$$\|\mathbf{x}\|^2 = |\langle \mathbf{x}, \boldsymbol{\alpha}_1 \rangle|^2 + |\langle \mathbf{x}, \boldsymbol{\alpha}_2 \rangle|^2 + \cdots + |\langle \mathbf{x}, \boldsymbol{\alpha}_n \rangle|^2.$$

(d) (Bessel's inequality) $\sum_{i=1}^{k}\langle \mathbf{x}, \boldsymbol{\alpha}_i\rangle \leq \|\mathbf{x}\|^2$ for each $k \leq n$.
Equality occurs if and only if $\mathbf{x}$ belongs to the space spanned by the $\boldsymbol{\alpha}_i$.

**Proofs.** Section 2.2.3.

2.24. Schott [2005: 54].

2.25a. If $\mathbf{x}_i, \mathbf{x}_2 \in S^{\perp}$, then $\langle \mathbf{x}_i, \mathbf{y}\rangle = 0$ for all $\mathbf{y} \in S$ and $\langle \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2, \mathbf{y}\rangle = \alpha_1\langle \mathbf{x}_1, \mathbf{y}\rangle + \alpha_2\langle \mathbf{x}_2, \mathbf{y}\rangle = 0$, i.e., $\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 \in S^{\perp}$.

2.25b. If $\mathbf{x} \in S$, then $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ for all $\mathbf{y} \in S^{\perp}$ and $\mathbf{x} \in (S^{\perp})^{\perp}$. By (a), $(S^{\perp})^{\perp}$ is a vector space even if $S$ is not; then use (2.26b).

2.25c. If $\mathbf{x}$ belongs to the left-hand side ($LHS$), then $\langle \mathbf{x}, \mathbf{s} + \mathbf{t}\rangle = \langle \mathbf{x}, \mathbf{s}\rangle + \langle \mathbf{x}, \mathbf{t}\rangle = 0$ for all $\mathbf{s} \in S$ and all $\mathbf{t} \in T$. Setting $\mathbf{s} = \mathbf{0}$, then $\langle \mathbf{x}, \mathbf{t}\rangle = 0$; similarly, $\langle \mathbf{x}, \mathbf{s}\rangle = 0$ and $LHS \subseteq RHS$. The argument reverses.

2.26. Rao and Rao [1998: 62–63].

2.27a–b. Harville [1997: 172].

2.27c. Harville [2001: 162, exercise 3] and Rao and Bhimasankaram [2000: 267].

2.28a(i). Follows from (2.26d) with $\mathcal{U} = \mathcal{W}$.

2.28a(ii). If $\mathbf{x} \in RHS$, then $\mathbf{x} = \mathbf{y} + \mathbf{z}$ where $\mathbf{y} \in \mathcal{V} \subseteq \mathcal{W}$ and $\mathbf{z} \in \mathcal{W}$ so that $\mathbf{x} \in \mathcal{W}$ and $RHS \subseteq LHS$. Then use (i) to show $\dim(RHS) = \dim(LHS)$.

2.29. Kollo and von Rosen [2005: 29].

2.30. Rao and Bhimasankaram [2000: 262] and Seber and Lee [2003: 338–339]. For matrices see Harville [1997: 63–64].

2.32a–c. Rao and Rao [1998: 59–61].

2.32d. Rao [1973a: 10].

### 2.2.4 Column and Null Spaces

***Definition* 2.18.** If $\mathbf{A}$ is a matrix (real or complex), then the space spanned by the columns of $\mathbf{A}$ is called the *column space* of $\mathbf{A}$, and is denoted by $\mathcal{C}(\mathbf{A})$. (Some authors, including myself in the past, call this the *range space* of $\mathbf{A}$ and write $\mathcal{R}(\mathbf{A})$.) The corresponding *row space* of $\mathbf{A}$ is $\mathcal{C}(\mathbf{A}')$, which some authors write as $\mathcal{R}(\mathbf{A})$; hence my choice of notation to avoid this confusion. The *null space* or *kernel*, $\mathcal{N}(\mathbf{A})$ of $\mathbf{A}$, is defined as follows:

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

The following results are all expressed in terms of complex matrices, but they clearly hold for real matrices as well.

**2.33.** From the definition of a vector subspace we find that $\mathcal{C}(\mathbf{A})$ and $\mathcal{N}(\mathbf{A})$ are both vector subspaces.

**2.34.** Let $\mathbf{A}$ and $\mathbf{B}$ both have $n$ columns. If any one of the following conditions holds, then all three hold:

(1) $\mathcal{C}(\mathbf{A}') \subseteq \mathcal{C}(\mathbf{B}')$.

(2) $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A})$.

(3) $\mathbf{A}(\mathbf{I}_n - \mathbf{B}^-\mathbf{B}) = \mathbf{0}$.

If (3) holds for a particular weak inverse $\mathbf{B}^-$, then (3) holds for any weak inverse $\mathbf{B}^-$.

**2.35.** Let $\mathbf{A}$ be any complex matrix.

(a) $\mathcal{N}(\mathbf{A}^*\mathbf{A}) = \mathcal{N}(\mathbf{A})$.

(b) $\mathcal{C}(\mathbf{A}\mathbf{A}^*) = \mathcal{C}(\mathbf{A})$.

(c) Two more results follow from (a) and (b) by interchanging $\mathbf{A}$ and $\mathbf{A}^*$.

In most applications $\mathbf{A}$ is real so that $\mathbf{A}^* = \mathbf{A}'$.

**2.36.** $\mathcal{N}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{I} - \mathbf{A})$ and $\mathcal{N}(\mathbf{I} - \mathbf{A}) \subseteq \mathcal{C}(\mathbf{A})$.

**2.37.** If $\mathbf{x} \perp \mathbf{y}$ when $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^*\mathbf{y} = 0$, and $\mathbf{A}$ is an $m \times n$ complex matrix, then $\mathcal{N}(\mathbf{A}) = \{\mathcal{C}(\mathbf{A}^*)\}^{\perp}$. We therefore have an orthogonal decomposition

$$\mathcal{N}(\mathbf{A}) \oplus \mathcal{C}(\mathbf{A}^*) = \mathbb{F}^n \quad \text{and} \quad \dim \mathcal{N}(\mathbf{A}) + \dim \mathcal{C}(\mathbf{A}^*) = n.$$

We get a further result by interchanging the roles of $\mathbf{A}$ and $\mathbf{A}^*$. Note that $\dim[\mathcal{C}(\mathbf{A}^*)] = \operatorname{rank} \mathbf{A}^* = \operatorname{rank} \mathbf{A}$, by (3.3c).

**2.38.** If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $m \times p$ , then $\mathcal{C}(\mathbf{B}) \subseteq \mathcal{C}(\mathbf{A})$ if and only if there exists an $n \times p$ matrix $\mathbf{R}$ such that $\mathbf{AR} = \mathbf{B}$. Furthermore, if $p = n$, $\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{B})$ if and only if there exists such a nonsingular $\mathbf{R}$. Similar results are available for row spaces by simply taking transposes. Thus if $\mathbf{C}$ is $q \times n$, then $\mathcal{C}(\mathbf{C}') \subseteq \mathcal{C}(\mathbf{A}')$ if and only if there exists a $q \times m$ matrix $\mathbf{S}$ such that $\mathbf{SA} = \mathbf{C}$.

**2.39.** The following hold for conformable matrices:

(a) If $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{B})$, then $\mathcal{C}(\mathbf{A}'\mathbf{B}) = \mathcal{C}(\mathbf{A}')$.

(b) $\mathcal{C}(\mathbf{B}_1) \subseteq \mathcal{C}(\mathbf{B}_2)$ implies that $\mathcal{C}(\mathbf{A}'\mathbf{B}_1) \subseteq \mathcal{C}(\mathbf{A}'\mathbf{B}_2)$.

(c) $\mathcal{C}(\mathbf{B}_1) = \mathcal{C}(\mathbf{B}_2)$ implies that $\mathcal{C}(\mathbf{A}'\mathbf{B}_1) = \mathcal{C}(\mathbf{A}'\mathbf{B}_2)$.

(d) If $\mathcal{C}(\mathbf{A} + \mathbf{BE}) \subseteq \mathcal{C}(\mathbf{B})$ for some conformable $\mathbf{E}$, then $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{B})$.

(e) If $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{B})$, then $\mathcal{C}(\mathbf{A} + \mathbf{BE}) \subseteq \mathcal{C}(\mathbf{B})$ for any conformable $\mathbf{E}$.

**Proofs.** Section 2.2.4.

2.34. Scott and Styan [1985: 210].

2.35. Meyer [2000a: 212–213].

2.36. Note that $\mathbf{Bx} = \mathbf{0}$ if and only if $\mathbf{x} = (\mathbf{I} - \mathbf{B})\mathbf{x}$. Set $\mathbf{B} = \mathbf{A}$ and $\mathbf{B} = \mathbf{I} - \mathbf{A}$.

2.37. Ben-Israel and Greville [2003: 12], Rao and Bhimasankaram [2000: 269], and Seber and Lee [2003: 477, real case].

2.38. Graybill [1983: 90] and Harville [1997: 30].

2.39. Quoted by Kollo and von Rosen [2005: 49]. For (a) we first have $\mathcal{C}(\mathbf{A}'\mathbf{B}) \subseteq \mathcal{C}(\mathbf{A}')$. Then, from (2.35), $\mathbf{A}'\mathbf{x} = \mathbf{A}'\mathbf{A}\mathbf{y} = \mathbf{A}'\mathbf{B}\mathbf{R}\mathbf{y} \in \mathcal{C}(\mathbf{A}'\mathbf{B})$, by (2.38), i.e., $\mathcal{C}(\mathbf{A}') \subseteq \mathcal{C}(\mathbf{A}'\mathbf{B})$. The rest are straightforward.

## 2.3 PROJECTIONS

***Definition* 2.19.** A square matrix $\mathbf{P}$ such that $\mathbf{P}^2 = \mathbf{P}$ is said to be *idempotent*. In this section we focus on the geometrical properties of such matrices, which are used extensively in statistics. Algebraic properties are considered in Section 8.6.

### 2.3.1 General Projections

***Definition* 2.20.** Let the vector space $\mathcal{U}$ be the direct sum of two vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ so that $\mathcal{U} = \mathcal{V}_1 \oplus \mathcal{V}_2$ (i.e., $\mathcal{V}_1 \cap \mathcal{V}_2 = \mathbf{0}$). Then every vector $\mathbf{v} \in \mathcal{V}$ has a unique decomposition $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_i \in \mathcal{V}_i$ $(i = 1, 2)$. The transformation $\mathbf{v} \to \mathbf{v}_1$ is called the *projection of* $\mathbf{v}$ *on* $\mathcal{V}_1$ *along* $\mathcal{V}_2$. Here uniqueness follows by assuming another decomposition $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$ so that $\mathbf{v}_1 - \mathbf{w}_1 = -(\mathbf{v}_2 - \mathbf{w}_2)$, which implies $\mathbf{v}_i = \mathbf{w}_i$ for $i = 1, 2$, otherwise $\mathcal{V}_1 \cap \mathcal{V}_2 \neq \mathbf{0}$. Usually $\mathcal{U} = \mathbb{F}^n$, and the following hold if $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.

**2.40.** The above projection on $\mathcal{V}_1$ along $\mathcal{V}_2$ can be represented by an $n \times n$ matrix $\mathbf{P}$ called a *projector* or *projection matrix* so that $\mathbf{P}\mathbf{v} = \mathbf{v}_1$. Also $\mathbf{P}$ is unique and idempotent.

**2.41.** Using the above notation, $\mathbf{v} = \mathbf{P}\mathbf{v} + (\mathbf{I}_n - \mathbf{P})\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, so that $\mathbf{v}_2 = (\mathbf{I}_n - \mathbf{P})\mathbf{v}$ is the projection of $\mathbf{v}$ on $\mathcal{V}_2$ along $\mathcal{V}_1$. Here $\mathbf{P}$ and $\mathbf{I}_n - \mathbf{P}$ are unique and idempotent, and

$$\mathbf{P}(\mathbf{I}_n - \mathbf{P}) = \mathbf{0}.$$

**2.42.** Using the above notation, we can identify $\mathcal{V}_1$ and $\mathcal{V}_2$ as follows:

(a) $\mathcal{C}(\mathbf{P}) = \mathcal{V}_1$.

(b) $\mathcal{C}(\mathbf{I}_n - \mathbf{P}) = \mathcal{V}_2$.

(c) If $\mathbf{P}$ is idempotent, then from (8.61) we obtain

$$\mathcal{C}(\mathbf{P}) \oplus \mathcal{N}(\mathbf{P}) = \mathcal{V}_1 \oplus \mathcal{V}_2.$$

**2.43.** Using the notation of (2.42), suppose that $\mathcal{V}_1 = \mathcal{C}(\mathbf{A})$, where $\mathbf{A}$ is $n \times n$ of rank $r$. Let $\mathbf{A} = \mathbf{R}_{n \times r}\mathbf{C}_{r \times n}$ be a full-rank factorization of $\mathbf{A}$ (cf. 3.5). Then

$$\mathbf{P} = \mathbf{R}(\mathbf{C}\mathbf{R})^{-1}\mathbf{C}$$

is the projection onto $\mathcal{V}_1$ along $\mathcal{V}_2$.

**Proofs.** Section 2.3.1.

2.40. Assume two projectors $\mathbf{P}_i$ $(i = 1, 2)$, then $(\mathbf{P}_1 - \mathbf{P}_2)\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_1 = \mathbf{0}$ for all $\mathbf{v}$ so that $\mathbf{P}_1 = \mathbf{P}_2$. Now $\mathbf{v}_1 = \mathbf{v}_1 + \mathbf{0}$ is the unique decomposition of $\mathbf{v}_1$ so that $\mathbf{P}^2\mathbf{v} = \mathbf{P}(\mathbf{P}\mathbf{v}) = \mathbf{P}\mathbf{v}_1 = \mathbf{v}_1 = \mathbf{P}\mathbf{v}$ for all $\mathbf{v}$ so that $\mathbf{P}^2 = \mathbf{P}$.

2.41. Rao and Rao [1998: 240–241]. Multiply the first equation by $\mathbf{P}$ to prove $\mathbf{P}(\mathbf{I}_n - \mathbf{P}) = \mathbf{0}$.

2.42a. $\mathcal{C}(\mathbf{P}) \subseteq \mathcal{V}_1$ as $\mathbf{P}$ projects onto $\mathcal{V}_1$. Conversely, if $\mathbf{v}_1 \in \mathcal{V}_1$, then $\mathbf{P}\mathbf{v}_1 = \mathbf{v}_1$, and $\mathcal{V}_1 \subseteq \mathcal{C}(\mathbf{P})$; (b) is similar.

2.43. Meyer [2000a: 634].

### 2.3.2 Orthogonal Projections

***Definition* 2.21.** Suppose $\mathcal{U}$ has an inner product $\langle , \rangle$, and let $\mathcal{V}$ be a vector subspace with orthogonal complement $\mathcal{V}^\perp$, namely

$$\mathcal{V}^\perp = \{\mathbf{x} : \langle \mathbf{x}, \mathbf{y} \rangle = 0, \text{ for every } \mathbf{y} \in \mathcal{V}\}.$$

Then $\mathcal{U} = \mathcal{V} \oplus \mathcal{V}^\perp$ so that every $\mathbf{v} \in \mathcal{U}$ can be expressed uniquely in the form $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1 \in \mathcal{V}$ and $\mathbf{v}_2 \in \mathcal{V}^\perp$. The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are called the *orthogonal projections* of $\mathbf{v}$ onto $\mathcal{V}$ and $\mathcal{V}^\perp$, respectively (we shall omit the words "along $\mathcal{V}^\perp$" and "along $\mathcal{V}$", respectively). Orthogonal projections will, of course, share the same properties as general projections. If $\mathcal{V} = \mathcal{C}(\mathbf{A})$, we shall denote the orthogonal projection $\mathbf{P}_\mathcal{V}$ onto $\mathcal{V}$ by $\mathbf{P}_\mathbf{A}$. In what follows we assume that $\mathcal{U} = \mathbb{F}^n$.

**2.44.** Using the above notation, $\mathbf{v}_1 = \mathbf{P}_\mathcal{V}\mathbf{v}$ and $\mathbf{v}_2 = (\mathbf{I}_n - \mathbf{P}_\mathcal{V})\mathbf{v}$, where $\mathbf{P}_\mathcal{V}$ and $\mathbf{I}_n - \mathbf{P}_\mathcal{V}$ are unique idempotent matrices. The matrix $\mathbf{P}_\mathcal{V}$ is said to be the *orthogonal projector* or *orthogonal projection matrix* of $\mathbb{F}^n$ onto $\mathcal{V}$, while $\mathbf{P}_{\mathcal{V}^\perp} = \mathbf{I}_n - \mathbf{P}_\mathcal{V}$ is the orthogonal projector of $\mathbb{F}^n$ onto $\mathcal{V}^\perp$. As we shall see below, the definition of orthogonality depends on the definition of $\langle \mathbf{x}, \mathbf{y} \rangle$.

**2.45.** If $\mathbb{F}^n = \mathbb{R}^n$ and $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$, then from the orthogonality we have

$$\mathbf{P}'_\mathcal{V}(\mathbf{I} - \mathbf{P}_\mathcal{V}) = \mathbf{0},$$

and $\mathbf{P}_\mathcal{V}$ is symmetric as well as being idempotent.

**2.46.** Let $\mathbb{F}^n = \mathbb{C}^n$ and define $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^*\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a Hermitian positive definite matrix. Note that $\mathbf{x} \perp \mathbf{y}$ if $\mathbf{y}^*\mathbf{A}\mathbf{x} = 0$ (cf. 2.19c).

(a) Let $\mathbf{P}_\mathcal{V}$ be the orthogonal projection matrix that projects onto $\mathcal{V}$. Then $\mathbf{P}_\mathcal{V}^2 = \mathbf{P}_\mathcal{V}$ and $\mathbf{A}\mathbf{P}_\mathcal{V}$ is Hermitian, that is,

$$\mathbf{A}\mathbf{P}_\mathcal{V} = \mathbf{P}_\mathcal{V}^*\mathbf{A}.$$

(Note that $\mathbf{P}_\mathcal{V}$ is generally not Hermitian. However, if $\mathbf{A} = \mathbf{I}_n$, then $\mathbf{P}_\mathcal{V}$ is Hermitian.)

(b) $\mathcal{C}(\mathbf{P}_\mathcal{V}) = \mathcal{V}$ and $\mathcal{C}(\mathbf{I}_n - \mathbf{P}_\mathcal{V}) = \mathcal{V}^\perp$ (from 2.42). Also

$$\mathbf{P}_\mathcal{V}^*\mathbf{A}(\mathbf{I}_n - \mathbf{P}_\mathcal{V}) = \mathbf{A}\mathbf{P}_\mathcal{V}(\mathbf{I}_n - \mathbf{P}_\mathcal{V}) = \mathbf{0}.$$

(c) Let $\mathcal{V} = \mathcal{C}(\mathbf{X})$. Then

$$\mathbf{P}_{\mathcal{V}} = \mathbf{X}(\mathbf{X}^*\mathbf{A}\mathbf{X})^-\mathbf{X}^*\mathbf{A},$$

which is unique for any weak inverse $(\mathbf{X}^*\mathbf{A}\mathbf{X})^-$ and therefore invariant. Also $\mathbf{P}_{\mathcal{V}^\perp} = \mathbf{I}_n - \mathbf{P}_{\mathcal{V}}$.

(d) If $\mathcal{V} = \mathcal{C}(\mathbf{X})$, then $\mathbf{P}_{\mathcal{V}}\mathbf{X} = \mathbf{X}$.

**2.47.** Of particular interest is a special case of (2.46) above, namely $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}'\mathbf{V}^{-1}\mathbf{y}$, where $\mathbf{V}$ is positive definite and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Because of its statistical importance in a variety of nonlinear models including nonlinear regression (e.g., generalized or weighted least squares) and multinomial models, $\langle \mathbf{x}, \mathbf{y}\rangle$ has been called the *weighted inner product space* (Wei [1997]). We now list some special cases of the previous general theory. Let $\mathbf{X}$ be $n \times p$ of rank $p$ and $\mathcal{V} = \mathcal{C}(\mathbf{X})$. Then:

(a) $\mathbf{P}_{\mathcal{V}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}$, which implies $\mathbf{P}_{\mathcal{V}}^2 = \mathbf{P}_{\mathcal{V}}$ and $\mathbf{P}_{\mathcal{V}}'\mathbf{V}^{-1} = \mathbf{V}^{-1}\mathbf{P}_{\mathcal{V}}$. Here $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-$ is any weak inverse of $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. Further properties of $\mathbf{P}_{\mathcal{V}}$ (with $\mathbf{V}^{-1}$ replaced by $\mathbf{V}$) are given by Harville [2001: 106–112].

(b) If the columns of $\mathbf{Q}$ and $\mathbf{N}$ are respectively orthonormal bases of $\mathcal{V}$ and $\mathcal{V}^\perp$, then $\mathbf{P}_{\mathcal{V}} = \mathbf{Q}\mathbf{Q}'\mathbf{V}^{-1}$ and $\mathbf{P}_{\mathcal{V}^\perp} = \mathbf{N}\mathbf{N}'\mathbf{V}^{-1}$, where $\mathbf{P}_{\mathcal{V}} + \mathbf{P}_{\mathcal{V}^\perp} = \mathbf{I}_n$.

(c) From (b), $\mathbf{Q}'\mathbf{V}^{-1}\mathbf{N} = \mathbf{0}$.

We can set $\mathbf{V} = \mathbf{I}$ is the above to get the unweighted case.

**2.48.** Let $\mathbf{V}$ be an $n \times n$ positive definite matrix, $\mathbf{G}$ an $n \times g$ matrix of rank $g$ $(g \le n)$, and $\mathbf{F}$ an $n \times f$ matrix $(f = n - g)$ of rank $f$ such that $\mathbf{G}'\mathbf{F} = \mathbf{0}$. Then

$$\mathbf{V}\mathbf{F}(\mathbf{F}'\mathbf{V}\mathbf{F})^{-1}\mathbf{F}' + \mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1} = \mathbf{I}_n.$$

**2.49.** Let $\mathbb{F}^n = \mathbb{C}^n$, $\mathbf{v} \in \mathbb{C}^n$, and define $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^*\mathbf{y}$, i.e., $\mathbf{A} = \mathbf{I}_n$ in (2.46). Then:

(a) $\mathbf{P}_{\mathcal{V}}$ is an orthogonal projection matrix on some vector space if and only if $\mathbf{P}_{\mathcal{V}}$ is idempotent and Hermitian.

(b) From (2.42) we have $\mathcal{V} = \mathcal{C}(\mathbf{P}_{\mathcal{V}})$.

(c) Let $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p)$, where the columns $\mathbf{t}_i$ of $\mathbf{T}$ form an orthonormal basis for $\mathcal{V}$. Then $\mathbf{P}_{\mathcal{V}} = \mathbf{T}\mathbf{T}^*$, and the projection of $\mathbf{v}$ onto $\mathcal{V}$ is $\mathbf{v}_1 = \mathbf{T}\mathbf{T}^*\mathbf{v} = \sum_{i=1}^r (\mathbf{t}_i^*\mathbf{v})\mathbf{t}_i$.

(d) If $\mathcal{V} = \mathcal{C}(\mathbf{X})$, then $\mathbf{P}_{\mathcal{V}} = \mathbf{X}(\mathbf{X}^*\mathbf{X})^-\mathbf{X}^* = \mathbf{X}\mathbf{X}^+$, where $(\mathbf{X}^*\mathbf{X})^-$ is a weak inverse of $\mathbf{X}^*\mathbf{X}$ and $\mathbf{X}^+$ is the Moore–Penrose inverse of $\mathbf{X}$. When the columns of $\mathbf{X}$ are linearly independent, $\mathbf{P}_{\mathcal{V}} = \mathbf{X}(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*$.

(e) Let $\mathcal{V} = \mathcal{N}(\mathbf{A})$, the null space of $\mathbf{A}$. Then, since $\mathcal{V}^\perp = \mathcal{C}(\mathbf{A}^*)$ (by 2.37), $\mathbf{P}_{\mathcal{V}} = \mathbf{I}_n - \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^-\mathbf{A}$.

(f) If $\mathbb{F}^n = \mathbb{R}^n$, then the previous results hold by replacing $^*$ by $'$ and replacing Hermitian by real symmetric. For example, if $\mathcal{V} = \mathcal{C}(\mathbf{A})$, then $\mathbf{P}_{\mathcal{V}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$. Furthermore, $\mathbf{x}'\mathbf{P}_{\mathcal{V}}\mathbf{x} = \mathbf{x}\mathbf{P}_{\mathcal{V}}'\mathbf{P}_{\mathcal{V}}\mathbf{x} = \mathbf{y}'\mathbf{y} \ge 0$, so that $\mathbf{P}_{\mathcal{V}}$ is non-negative definite. This result is used frequently in this book.

**2.50.** Let $\mathbf{A}$ be an $n \times m$ real matrix and $\mathbf{B}$ an $n \times p$ real matrix. Assuming that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$, let $\mathbf{P_D}$ denote the orthogonal projection onto $\mathcal{C}(\mathbf{D})$ for any matrix $\mathbf{D}$.

(a) $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}) = \mathcal{C}[\mathbf{A}(\mathbf{I}_m - \mathbf{P}_\mathcal{V})]$, where $\mathcal{V} = \mathcal{C}[\mathbf{A}'(\mathbf{I} - \mathbf{P_B})]$.

(b) $\mathcal{C}(\mathbf{A}, \mathbf{B}) = \mathcal{C}(\mathbf{A}) \oplus \mathcal{C}[(\mathbf{I} - \mathbf{P_A})\mathbf{B}]$.

(c) From (b) we have $\mathbf{P}_{(\mathbf{A},\mathbf{B})} = \mathbf{P_A} + \mathbf{P}_{(\mathbf{I}-\mathbf{P_A})\mathbf{B}}$.

(d) $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{B})$ if and only if $\mathbf{P_B} - \mathbf{P_A}$ is non-negative definite, and $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$ if and only if $\mathbf{P_B} - \mathbf{P_A}$ is positive definite.

The above results are particularly useful in partitioned linear models.

**2.51.** (Some Subspace Properties) Let $\omega$, $\Omega$, and $\mathcal{V}$ be vector subspaces in $\mathbb{R}^n$ with $\omega \subset \Omega$, and let $\mathbf{P}_\omega$ and $\mathbf{P}_\Omega$ be the respective orthogonal projectors onto $\omega$ and $\Omega$ with respect to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$ defined on $\mathbb{R}^n$. Thus $\mathbf{P}_\omega$ and $\mathbf{P}_\Omega$ are symmetric and idempotent. The following results hold (see also (2.53c)).

(a) $\mathbf{P}_\Omega \mathbf{P}_\omega = \mathbf{P}_\omega \mathbf{P}_\Omega = \mathbf{P}_\omega$.

(b) $\mathbf{P}_{\omega^\perp \cap \Omega} = \mathbf{P}_\Omega - \mathbf{P}_\omega$.

(c) $\mathbf{A}\mathbf{P}_\Omega\mathbf{A}'$ is nonsingular if and only if the rows of $\mathbf{A}$ are linearly independent and $\mathcal{C}(\mathbf{A}') \cap \Omega^\perp = \mathbf{0}$.

(d) If $\omega = \Omega \cap \mathcal{N}(\mathbf{A})$, where $\mathcal{N}(\mathbf{A})$ is the null space of $\mathbf{A}$, then:

  (i) $\omega^\perp \cap \Omega = \mathcal{C}(\mathbf{P}_\Omega \mathbf{A}')$.

  (ii) $\mathbf{P}_{\omega^\perp \cap \Omega} = \mathbf{P}_\Omega \mathbf{A}'(\mathbf{A}\mathbf{P}_\Omega\mathbf{A}')^- \mathbf{A}\mathbf{P}_\Omega$, where $(\mathbf{A}\mathbf{P}_\Omega\mathbf{A}')^-$ is any weak inverse of $\mathbf{A}\mathbf{P}_\Omega\mathbf{A}'$.

(e) Let $\Omega = \mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{X}_1, \mathbf{X}_2)$, where the columns of $n \times p$ $\mathbf{X}$ are linearly independent, and let $\omega = \mathcal{C}(\mathbf{X}_1)$, where $\dim(\omega) = r$.

  (i) We have from (c), with $\mathcal{V} = \omega^\perp$ and $\mathbf{P}_\omega = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$ ($= \mathbf{P}_1$, say), that $\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2$ is nonsingular.

  (ii) $\omega = \Omega \cap \mathcal{N}[\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1)]$.

  (iii) It follows from (b) and (d)(ii)) that

$$\mathbf{P}_\Omega - \mathbf{P}_\omega = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2[\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2]^{-1}\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1).$$

By interchanging the subscripts 1 and 2, a further result can be obtained.

Note that (a)–(d) are used in testing a linear hypothesis for a linear regression model (e.g., Seber [1977: sections 3.9.3 and 4.5] and Seber and Lee [2003: theorems 4.1 and 4.3]); (e) is related to subset regression (see Seber and Wild [1989: Appendix D] for a summary).

**2.52.** If $\Omega$ and $\omega_i$ $(i = 1, 2, \ldots, k)$ are vector subspaces of $\mathbb{R}^n$ satisfying $\omega_i \subset \Omega$, with inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$, then the following results are equivalent:

(1) $\mathbf{P}_{\omega_1 \cap \omega_2 \cap \cdots \cap \omega_i} - \mathbf{P}_{\omega_1 \cap \omega_2 \cap \cdots \cap \omega_{i-1}} = \mathbf{P}_\Omega - \mathbf{P}_{\omega_i}$ for $i = 1, 2, \ldots, k$.

(2) $\omega_i^\perp \cap \Omega \perp \omega_j^\perp \cap \Omega$ for all $i, j = 1, 2, \ldots, k; \quad i \neq j$.

(3) $\omega_i^\perp \cap \Omega \subset \omega_j$ for all $i, j = 1, 2, \ldots, k; \quad i \neq j$.

The above results are useful in testing a sequence of nested hypotheses in an analysis of variance, when there are equal numbers of observations per cell (balanced designs) leading to an underlying orthogonal structure (cf. Darroch and Silvey [1963], Seber [1980: section 6.2], and Seber and Lee [2003: 203]).

**2.53.** Let $\omega_1$ and $\omega_2$ be vector subspaces of $\mathbb{R}^n$ with inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$.

(a) $\mathbf{P} = \mathbf{P}_{\omega_1} + \mathbf{P}_{\omega_2}$ is an orthogonal projector if and only if $\omega_1 \perp \omega_2$, in which case $\mathbf{P}_{\omega_1} + \mathbf{P}_{\omega_2} = \mathbf{P}_\omega$, where $\omega = \omega_1 \oplus \omega_2$.

(b) If $\omega_1 = \mathcal{C}(\mathbf{A})$ and $\omega_2 = \mathcal{C}(\mathbf{B})$ in (a), then $\omega_1 \oplus \omega_2 = \mathcal{C}(\mathbf{A}, \mathbf{B})$.

(c) The following statements are equivalent:

(1) $\mathbf{P}_{\omega_1} - \mathbf{P}_{\omega_2}$ is an orthogonal projection matrix.

(2) $\|\mathbf{P}_{\omega_1}\mathbf{x}\|_2 \geq \|\mathbf{P}_{\omega_2}\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathbb{R}^n$.

(3) $\mathbf{P}_{\omega_1}\mathbf{P}_{\omega_2} = \mathbf{P}_{\omega_2}$.

(4) $\mathbf{P}_{\omega_2}\mathbf{P}_{\omega_1} = \mathbf{P}_{\omega_2}$.

(5) $\omega_2 \subset \omega_1$.

(d) $\mathbf{P}_{\omega_1 \cap \omega_2} = 2\mathbf{P}_{\omega_1}(\mathbf{P}_{\omega_1} + \mathbf{P}_{\omega_2})^+\mathbf{P}_{\omega_2} = 2\mathbf{P}_{\omega_2}(\mathbf{P}_{\omega_1} + \mathbf{P}_{\omega_2})^+\mathbf{P}_{\omega_1}$. Here $\mathbf{B}^+$ denotes the Moore–Penrose inverse of $\mathbf{B}$.

The above results hold for $\mathbb{C}^n$ if $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^*\mathbf{x}$ and $'$ is replaced by $*$.

***Definition* 2.22.** (Centering) Let $\mathbf{a} = (a_i)$ be an $n \times 1$ real vector, and let $\overline{a} = \sum_{i=1}^n a_i/n$. We say that the $\mathbf{a}$ is *centered* when we transform $a_i$ to $b_i = a_i - \overline{a}$.

If we have the $n \times p$ matrix $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_n)' = (\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(p)})$ and $\overline{\mathbf{a}} = n^{-1}\sum_{i=1}^n \mathbf{a}_i$, then we say that $\mathbf{A}$ is *row centered* if we transform it to the matrix $\mathbf{B} = (\mathbf{a}_1 - \overline{\mathbf{a}}, \mathbf{a}_2 - \overline{\mathbf{a}}, \ldots, \mathbf{a}_n - \overline{\mathbf{a}})'$.

If $\overline{\mathbf{a}}^{(col)} = \sum_{j=1}^p \mathbf{a}^{(j)}/p$, then we say that $\mathbf{A}$ is *column centered* if we form the matrix $\mathbf{C} = (\mathbf{a}^{(1)} - \overline{\mathbf{a}}^{(col)}, \mathbf{a}^{(2)} - \overline{\mathbf{a}}^{(col)}, \ldots, \mathbf{a}^{(p)} - \overline{\mathbf{a}}^{(col)})$.

We say that $\mathbf{A}$ is *double-centered* if we apply both row and column centering.

**2.54.** Using the above notation, we have the following results:

(a) We can write $\overline{a} = \mathbf{1}_n'\mathbf{a}/n$ so that $(\overline{a}) = n^{-1}\mathbf{1}_n\mathbf{1}_n'\mathbf{a} = \mathbf{P}_{\mathbf{1}_n}\mathbf{a}$, where $\mathbf{P}_{\mathbf{1}_n} = n^{-1}\mathbf{1}_n\mathbf{1}_n'$ represents the orthogonal projection of $\mathbb{R}^n$ onto $\mathbf{1}_n$. Furthermore, $\mathbf{b} = \mathbf{a} - (\overline{a}) = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{a}$, where $\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}$ represents an orthogonal projection perpendicular to $\mathbf{1}_n$; this projection matrix is called a *centering matrix*.

(b) $\overline{\mathbf{a}} = \mathbf{A}'\mathbf{1}_n/n$ and $\mathbf{B} = \mathbf{A} - \mathbf{1}_n\overline{\mathbf{a}}' = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{A}$.

(c) $\overline{\mathbf{a}}^{(col)} = \mathbf{A}\mathbf{1}_p/p$ and $\mathbf{C} = \mathbf{A}(\mathbf{I}_p - \mathbf{P}_{\mathbf{1}_p})$.

(d) When $\mathbf{A}$ is double centered we obtain $\mathbf{D} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{A}(\mathbf{I}_p - \mathbf{P}_{\mathbf{1}_p})$, where $d_{ij} = a_{ij} - \overline{a}_{i\cdot} - \overline{a}_{\cdot j} - \overline{a}_{\cdot\cdot}$, $\overline{a}_{i\cdot} = \sum_j a_{ij}/p$, $\overline{a}_{\cdot j} = \sum_i a_{ij}/n$, and $\overline{a}_{\cdot\cdot} = \sum_i \sum_j a_{ij}/(np)$.

Centering is used extensively in statistics, for example linear regression (Seber and Lee [2003: section 3.11.1 and section 11.7 for computing algorithms]) and principal component analysis, and double centering is used in classical metric scaling, in principal component analysis (Jolliffe [1992: section 14.2.3]), and in the singular-spectrum analysis (SAS) of times series, where it is applied to trajectory matrices (Golyandina et al. [2001: section 4.4, 272]).

**Proofs.** Section 2.3.2.

2.46. Rao [1973a: 47].

2.47. Wei [1997: 185–187].

2.48. Seber [1984: 536].

2.49. Seber and Lee [2003: Appendices B1 and B2, real case].

2.50a. Quoted by Rao and Mitra [1971: 118, exercise 7a].

2.50b–d. Sengupta and Jammalamadaka [2003: 39, 47]; (c) uses (2.44).

2.51a–d(i). Seber and Lee [2003: Appendix B3, 477-478, real case] and Seber [1984: Appendix B3, 535, real case].

2.51d(ii). If $\mathbf{x} \in \mathcal{C}(\mathbf{X}_1) = \omega$, then $\mathbf{P}_1\mathbf{x} = \mathbf{x}$, $\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1)\mathbf{x} = \mathbf{0}$, and $\mathbf{x} \in \mathcal{N}(\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1))$. Conversely, if $\mathbf{x} = \mathbf{X}_1\boldsymbol{\alpha}_1 + \mathbf{X}_2\boldsymbol{\alpha}_2 \in \Omega$ and $\mathbf{0} = \mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1)\mathbf{x} = \mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2\boldsymbol{\alpha}_2$ (since $\mathbf{P}_1\mathbf{X}_1 = \mathbf{X}_1$), then $\boldsymbol{\alpha}_2 = \mathbf{0}$ (by (i)) and $\mathbf{x} \in \mathcal{C}(\mathbf{X}_1)$.

2.52. Seber [1980: section 6.2].

2.53a. $\mathbf{P}$ is clearly symmetric and idempotent if and only $\mathbf{P}_{\omega_1}\mathbf{P}_{\omega_2} = -\mathbf{P}_{\omega_2}\mathbf{P}_{\omega_1}$. Multiplying on the left by $\mathbf{P}_{\omega_2}$ shows that $\mathbf{P}_{\omega_1}\mathbf{P}_{\omega_2}$ is symmetric and therefore $\mathbf{P}_{\omega_1}\mathbf{P}_{\omega_2} = \mathbf{0}$. Furthermore, since $\mathbf{P}_{\omega_i}$ is idempotent, we have from (2.35)

$$\mathcal{C}(\mathbf{P}_{\omega_1} + \mathbf{P}_{\omega_2}) = \mathcal{C}\left[(\mathbf{P}_{\omega_1}, \mathbf{P}_{\omega_2})\begin{pmatrix}\mathbf{P}_{\omega_1} \\ \mathbf{P}_{\omega_2}\end{pmatrix}\right] = \mathcal{C}(\mathbf{P}_{\omega_1}, \mathbf{P}_{\omega_2}) = \omega_1 \oplus \omega_2.$$

2.53b. $\mathbf{A}'\mathbf{B} = \mathbf{0}$ implies that $\mathbf{P}_{\mathbf{A}}\mathbf{P}_{\mathbf{B}} = \mathbf{0}$.

2.53c. Quoted, less generally, by Isotalo et al. [2005a: 61]. The proofs are straightforward. For (2), note that for a symmetric idempotent matrix, $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|_2^2$.

2.53d. Anderson and Duffin [1969] and Meyer [2000a: 441].

## 2.4 METRIC SPACES

***Definition* 2.23.** Let $S$ be a subset of $\mathbb{R}^n$. By a *metric* for $S$ we mean a real-valued function $d(\cdot,\cdot)$ on $S \times S$ such that:

(a) $d(\mathbf{x}, \mathbf{y}) \geq 0$ for all $\mathbf{x}, \mathbf{y} \in S$ with equality if and only if $\mathbf{x} = \mathbf{y}$ ($d$ is positive definite).

(b) $d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x})$ for all $\mathbf{x},\mathbf{y} \in S$ ($d$ is symmetric).

(c) $d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{x},\mathbf{z}) + d(\mathbf{y},\mathbf{z})$ for all $\mathbf{x},\mathbf{y},\mathbf{z} \in S$ (triangle inequality).

If we replace (c) by the stronger condition

(c′) $d(\mathbf{x},\mathbf{y}) \leq \max[d(\mathbf{x},\mathbf{z}), d(\mathbf{y},\mathbf{z})]$,

$d$ is called an *ultrametric.* Note that (c′) implies (c).

***Definition* 2.24.** A *metric space* is a pair $(S,d)$ consisting of a set $S$ and a metric $d$ for $S$.

**2.55.** If $d$ is a metric, then so are $d_1$, $d_2$, and $d_3$, where

$$\begin{aligned} d_1(\mathbf{x},\mathbf{y}) &= d(\mathbf{x},\mathbf{y})/(1+d(\mathbf{x},\mathbf{y})), \\ d_2(\mathbf{x},\mathbf{y}) &= \sqrt{d(\mathbf{x},\mathbf{y})}, \\ d_3(\mathbf{x},\mathbf{y}) &= kd(\mathbf{x},\mathbf{y}) \qquad (k>0). \end{aligned}$$

**2.56.** If $d$ is a metric, then $D(\mathbf{x},\mathbf{y}) = [d(\mathbf{x},\mathbf{y})]^2$ is not necessarily a metric.

**2.57.** (Canberra metric) If $\mathbf{x}$ and $\mathbf{y}$ have positive elements, then the function

$$d(\mathbf{x},\mathbf{y}) = \frac{1}{n}\sum_{j=1}^{n} \frac{|x_j - y_j|}{x_j + y_j}$$

is a metric.

**2.58.** (Minkowski Metrics) The function $\Delta_p$ is a metric, where

$$\Delta_p(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}, \quad p > 0.$$

The most common ones are $p = 1$ (the *city block metric*) and $p = 2$ (the *Euclidean metric*). Various scaled versions of $\Delta_1$ have also been used.

**2.59.** $\Delta_\infty(\mathbf{x},\mathbf{y}) = \sup_{1\leq i\leq n} |x_i - y_i|$, for all $\mathbf{x}$ and $\mathbf{y}$, is a metric.

***Definition* 2.25.** The *Mahalanobis distance* is defined to be

$$d(\mathbf{x},\mathbf{y}) = \{(\mathbf{x}-\mathbf{y})'\mathbf{A}(\mathbf{x}-\mathbf{y})\}^{1/2},$$

where $\mathbf{A}$ is positive definite. Here $d$ is a metric. The *Mahalanobis angle* $\theta$ between $\mathbf{x}$ and $\mathbf{y}$ subtended at the origin is defined by

$$\cos\theta = \frac{\mathbf{x}'\mathbf{A}\mathbf{y}}{(\mathbf{x}'\mathbf{A}\mathbf{x})^{1/2}(\mathbf{y}'\mathbf{A}\mathbf{y})^{1/2}}.$$

***Definition* 2.26.** A sequence of points $\{\mathbf{x}_i\}$ in $S$ for a metric space $(S,d)$ is called a *Cauchy sequence* if, for every $\epsilon > 0$, there exists a positive integer $N$ such the $d(\mathbf{x}_i,\mathbf{x}_j) < \epsilon$ for all $i,j > N$.

A sequence $\{\mathbf{x}_i\}$*converges* to a point $\mathbf{x}$ if, for every $\epsilon > 0$, there exists a positive integer $N$ such that $d(\mathbf{x},\mathbf{x}_i) < \epsilon$ for all $i > N$.

A metric space is said to be *complete* if every Cauchy sequence converges to a point in $S$.

***Definition* 2.27.** Let $f$ be a mapping of a metric space $(S, d)$ into itself. We call $f$ a *contraction* if there exists a constant $c$ with $0 < c \leq 1$ such that

$$d(f(\mathbf{x}), f(\mathbf{y})) \leq cd(\mathbf{x}, \mathbf{y}), \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in S.$$

If $0 < c < 1$, we say that $f$ is a *strict contraction.* If $f(\mathbf{x}) = \mathbf{x}$, then $\mathbf{x}$ is referred to as a *fixed point* of $f$.

**2.60.** (Contraction Mapping Theorem) Let $f$ be a strict contraction of a complete metric space into itself. Then $f$ has one and only one fixed point and, for any point $\mathbf{y} \in S$, the sequence

$$\mathbf{y},\ f(\mathbf{y}),\ f^2(\mathbf{y}),\ f^3(\mathbf{y}), \ldots,$$

where $f^r(\mathbf{y}) = f(f^{r-1}(\mathbf{y}))$, converges to the fixed point.

**2.61.** Let $(S, d)$ be a metric space with $S = \mathbb{C}^n$ and $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$. A matrix $\mathbf{A}$ is a contraction, that is

$$\|\mathbf{Ax} - \mathbf{Ay}\|_2 \leq c\|\mathbf{x} - \mathbf{y}\|_2 \quad \text{for} \quad 0 < c \leq 1,$$

if and only if $\sigma_{\max}(\mathbf{A}) \leq 1$, where $\sigma_{\max}(\mathbf{A})$ is the maximum singular value of $\mathbf{A}$. Further necessary and sufficient conditions for a matrix to be a contraction are given by Zhang [1999: section 5.4].

**Proofs.** Section 2.4.

2.55–2.57. Seber [1984: : 392, exercises 7.4–7.6, see the solutions].

2.58. Seber [1984: 352]. Use Minkowski's inequalities (12.17b) and $x_i - z_i = x_i - y_i + y_i - z_i$ to prove the triangle inequality.

2.59. Use the properties of sup.

2.60–2.61. Zhang [1999: 143–144].

## 2.5 CONVEX SETS AND FUNCTIONS

***Definition* 2.28.** A subset $C$ of $\mathbb{R}^n$ is called *convex* if, for any two points $\mathbf{x}_1, \mathbf{x}_2 \in C$, the line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$ is contained in $C$, that is,

$$\alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2 \in C \quad \text{for} \quad 0 \leq \alpha \leq 1.$$

We will list some properties of convex sets below. For a more comprehensive discussion see Berkovitz [2002], Kelly and Weiss [1979], Lay [1982], and Rockafellar [1970].

**2.62.** If $C_1$ and $C_2$ are convex sets in $\mathbb{R}^n$, then:

(a) $C_1 \cap C_2$ is convex.

(b) $C_1 + C_2$ is convex.

(c) $C_1 \cup C_2$ need not be convex.

These results clearly hold for any finite number of convex sets. The result (a) also holds for a countably infinite number of convex sets.

**2.63.** Given any set $A \in \mathbb{R}^n$, the set $C_A$ of points generated by taking the *convex combination* of every finite set of points $\mathbf{x}_i$ in $A$, namely

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_k\mathbf{x}_k \quad (\text{each } \alpha_i \geq 0 \text{ and } \sum_i \alpha_i = 1)$$

is a convex set containing $A$. The set $C_A$ is the smallest convex set containing $A$ and is called the *convex hull* of $A$. It is also the intersection of all convex sets containing $A$.

***Definition* 2.29.** Given $A$ a set in $\mathbb{R}^n$, we define $\mathbf{x}$ to be an *inner (interior) point* of $A$ if there is an *open sphere* with center $\mathbf{x}$ that is a subset of $A$; that is, there exists $\delta > 0$ such that

$$S_\delta = \{\mathbf{y} : \mathbf{y} \in \mathbb{R}^n, (\mathbf{y} - \mathbf{x})'(\mathbf{y} - \mathbf{x}) < \delta\} \subseteq A.$$

A *boundary point* $\mathbf{x}$ of $A$ (not necessarily belonging to $A$) is such that every open sphere with center $\mathbf{x}$ contains points both in $A$ and in $A^c$, the complement of $A$ with respect to $\mathbb{R}^n$.

A point $\mathbf{x}$ is a *limit (accumulation) point* if, for every $\delta > 0$, $S_\delta$ contains at least one point of $S$ distinct from $\mathbf{x}$.

The *closure* of set $A$ is obtained by adding to it all its boundary points not already in it, and is denoted by $\overline{A}$. It can also be obtained by adding to $S$ all its limit points.

The set $A$ is *closed* if $A = \overline{A}$, while the set is *open* if $A^c$, the complement of $A$, is closed. For any set $A$, $\overline{A}$ is the smallest closed set containing $A$.

An *exterior* point of $A$ is a point in $\overline{A}^c$. A point $\mathbf{x} \in A$ is an *extreme point* of $A$ if there are no distinct points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $A$ such that $\mathbf{x} = \alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$ for some $\alpha$ $(0 < \alpha < 1)$.

A set $A$ is *bounded* if it is contained in an open sphere $S_\delta$ for some $\delta > 0$.

A set which is closed and bounded is said to be *compact*. For some properties of open and closed sets see Magnus and Neudecker [1999: 66–69].

The above results generalize to more general spaces using a more general distance metric other than $\|\mathbf{x} - \mathbf{y}\|_2$.

**2.64.** Let $C$ be a convex set.

(a) The closure $\overline{C}$ is convex.

(b) $C$ and $\overline{C}$ have the same inner, boundary, and exterior points.

(c) Let $\mathbf{x}$ be an inner point and $\mathbf{y}$ a boundary point of $C$. Then the points $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}$ are inner points of $C$ for $0 < \alpha \leq 1$ and exterior points of $C$ for $\alpha > 1$

(d) If $T$ is an open subset of $\mathbb{R}^n$ and $T \subseteq \overline{C}$, then $T \subseteq C$.

**2.65.** (Separation theorems)

(a) Let $C$ be a closed convex subset and suppose $\mathbf{0} \notin C$. Then there exists a vector $\mathbf{a}$ such that $\mathbf{a}'\mathbf{x} > 0$ for all $\mathbf{x} \in C$.

(b) Let $C$ be a convex set and $\mathbf{y}$ an exterior point. Then there exists a unit vector $\mathbf{u}$ (i.e., $\|\mathbf{u}\|_2 = 1$) such that

$$\inf_{\mathbf{x} \in C} \mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}.$$

(c) Let $C$ be a convex set and $\mathbf{y}$ a point not in $C$, or a boundary point if in $C$. Then there exists a supporting plane through $\mathbf{y}$; that is, there exists a nonzero vector $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{a}'\mathbf{x} \geq \mathbf{a}'\mathbf{y}$ for all $\mathbf{x} \in C$, or equivalently $\inf_{\mathbf{x} \in C} \mathbf{a}'\mathbf{x} = \mathbf{a}'\mathbf{y}$, if $\mathbf{y}$ is a boundary point.

(d) Let $C_1$ and $C_2$ be convex sets with no inner point in common. Then there exists a hyperplane $\mathbf{a}'\mathbf{x} = b$ separating the two sets; that is, there exists a vector $\mathbf{a}$ and a scalar $b$ such that $\mathbf{a}'\mathbf{x} \geq b$ for all $\mathbf{x} \in C_1$ and $\mathbf{a}'\mathbf{y} \leq b$ for all $\mathbf{y} \in C_2$. This also implies that $\mathbf{a}'\mathbf{x}_1 \geq \mathbf{a}'\mathbf{x}_2$ for all $\mathbf{x}_1 \in C_1$ and all $\mathbf{x}_2 \in C_2$.

If $C_1$ and $C_2$ are also closed, we have strict separation so that there exist $\mathbf{a}$ and $b$ such that $\mathbf{a}'\mathbf{x} > b$ for $\mathbf{x} \in C_1$ and $\mathbf{a}'\mathbf{y} < b$ for $\mathbf{y} \in C_2$.

(e) Let $C$ be a convex subset, symmetric about $\mathbf{0}$, so that if $\mathbf{x} \in C$, then $-\mathbf{x} \in C$ also. Let $f(\mathbf{x}) \geq 0$ be a function for which (i) $f(\mathbf{x}) = f(-\mathbf{x})$, (ii) $C_\alpha = \{\mathbf{x} : f(\mathbf{x}) \geq \alpha\}$ is convex for any positive $\alpha$, and (iii) $\int_C f(\mathbf{x})\, d\mathbf{x} < \infty$. Then

$$\int_C f(\mathbf{x} + c\mathbf{y})\, d\mathbf{x} \geq \int_C f(\mathbf{x} + \mathbf{y})\, d\mathbf{x},$$

for all $0 \leq c \leq 1$ and $\mathbf{y} \in \mathbb{R}^n$.

**2.66.** (Convex Hull) If $C_A$ is the convex hull of a subset $A \in \mathbb{R}^n$, then every point of $A$ can be expressed as a convex combination of at most $n+1$ points in $A$.

**2.67.** (Extreme Points) If $C$ is a closed bounded convex set, it is spanned by its extreme points; that is, every point in $C$ can be expressed as a linear combination of its extreme points. Also $C$ has extreme points in every supporting hyperplane.

***Definition* 2.30.** A real valued function $f$ is *convex* in an interval $I$ of $\mathbb{R}$ if

$$f[\alpha x + (1-\alpha)y] \leq \alpha f(x) + (1-\alpha) f(y), \quad \text{all } \alpha \text{ such that } \quad 0 < \alpha < 1,$$

for all $x, y \in I$ $(x \neq y)$. The function $f$ is said to be *strictly convex* if $\leq$ is replaced by $<$ above.

We say that $f$ is (strictly) *concave* if $-f$ is (strictly) convex. A linear function is both convex and concave. A similar definition applies if $x$ is replaced by a vector or matrix.

A vector convex function is defined along the same lines. We say that $\mathbf{f}$ is *convex* if

$$\mathbf{f}(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha\mathbf{f}(\mathbf{x}) + (1-\alpha)\mathbf{f}(\mathbf{y})$$

for every $\alpha$ such that $0 \leq \alpha \leq 1$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$; $\mathbf{f}$ is *concave* if $-\mathbf{f}$ is convex. Here $\mathbf{a} \leq \mathbf{b}$ means $a_i \leq b_i$ for all $i$.

**2.68.** The following functions are convex.

(a) $-\log x$ $(x > 0)$.

(b) $x^p$, $p > 1\,(x > 0)$.

They can be used to establish a number of well-known inequalities (e.g., Horn and Johnson [1985: 535–536]).

**2.69.** The function

$$f(\mathbf{A}) = \log\det\mathbf{A}$$

is a strictly concave function on the convex set of Hermitian positive definite matrices.

**2.70.** Every convex and every concave function is continuous on its interior. However, a convex function may have a discontinuity at a boundary point and may not be differentiable at an interior point.

**2.71.** Every increasing convex (respectively concave) function of a convex (respectively concave) function is convex (respectively concave). Every strictly increasing convex (respectively concave) function of a strictly convex (respectively concave) function is strictly convex (respectively concave).

**2.72.** (Weirstrass's Theorem) Let $S$ be a compact subset of a real or complex vector space. If $f : S \to \mathbb{R}$ is a continuous function, then there exist points $\mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}} \in S$ such that

$$f(\mathbf{x}_{\text{min}}) \leq f(\mathbf{x}) \leq f(\mathbf{x}_{\text{max}}) \quad \text{for all } \mathbf{x} \in S.$$

***Definition* 2.31.** The *numerical range (field of values)* of an $n \times n$ complex matrix $\mathbf{A}$ is

$$\{\mathbf{x}^*\mathbf{A}\mathbf{x} : \|\mathbf{x}\| = 1, \mathbf{x} \in \mathbb{C}^n\}.$$

**2.73.** (Toeplitz–Hausdorff) The numerical range of an $n \times n$ complex matrix is a convex compact subset of $\mathbb{C}^n$. For further properties of a field of values see Gustafson and Rao [1997] and Horn and Johnson [1991].

**Proofs.** Section 2.5.

2.62. Schott [2005: 71].

2.64a–c. Quoted by Rao [1973a: 51].

2.64d. Schott [2005: 72].

2.65a. Schott [2005: 71].

2.65b. Rao [1973a: 51].

2.65c–d. Rao [1973a: 52] and Schott [2005: 73].

2.65e. Anderson [1955], and quoted by Schott [2005: 74].

2.66–2.67. Quoted by Rao [1973a: 53].

2.69. Horn and Johnson [1985: 466–467].

2.70–2.71. Magnus and Neudecker [1999: 76].

2.73. Horn and Johnson [1991: 8] and Zhang [1999: 88–89].

## 2.6 COORDINATE GEOMETRY

Occasionally one may need some results from coordinate geometry. Some of these are listed below for easy reference.

### 2.6.1 Hyperplanes and Lines

**2.74.** The equation of a hyperplane passing through the points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in $\mathbb{R}^n$ can be expressed in the form

$$\det \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ \mathbf{x} & \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{pmatrix} = 0.$$

**2.75.** Given the points $\mathbf{x}_1 = (a_1, b_1, c_1)'$ and $\mathbf{x}_2 = (a_2, b_2, c_2)'$ in $\mathbb{R}^3$, then the equation of the line through the points is

$$\frac{x - a_1}{a_1 - a_2} = \frac{y - b_1}{b_1 - b_2} = \frac{z - c_1}{c_1 - c_2}.$$

If the two points are A and B, then $a_1 - a_2 = AB\cos\theta_1$, and so on, so that we can replace the denominators of the above line by the direction cosines $\cos\theta_i$ of the line with respect to each axis. Then $\cos\theta_1^2 + \cos\theta_2^2 + \cos\theta_3^2 = 1$. This result clearly generalizes to two points in $\mathbb{R}^n$.

**2.76.** Given the plane $ax + by + cz + d = 0$ in $\mathbb{R}^3$, a normal vector to the plane is given by $(a, b, c)'$, and the perpendicular distance of the point $\mathbf{x}_1 = (x_1, y_1, z_1)'$ from the plane is

$$\frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

This result clearly generalises to $\mathbb{R}^n$. Given the plane $\mathbf{a}'\mathbf{x} + d = 0$, the distance of $\mathbf{x}_1$ from the plane is $(|\mathbf{a}'\mathbf{x}_1 + d|)/\|\mathbf{a}\|_2$.

**2.77.** Given $0 < \alpha < 1$, then $\mathbf{z} = (1 - \alpha)\mathbf{x} + \alpha\mathbf{y}$ divides the line segment joining $\mathbf{x}$ and $\mathbf{y}$ in the proportion $\alpha : (1 - \alpha)$.

**Proofs.** Section 2.6.1.

2.77. Abadir and Magnus [2005: 6].

### 2.6.2 Quadratics

**2.78.** If $\mathbf{A}$ is an $n \times n$ symmetric indefinite matrix (i.e., has both positive and negative eigenvalues), then $(\mathbf{x} - \mathbf{a})'\mathbf{A}(\mathbf{x} - \mathbf{a}) \leq c$ with $c > 0$ is a *hyperboloid* with center $\mathbf{a}$.

**2.79.** If $\mathbf{A}$ is an $n \times n$ positive definite matrix, then $(\mathbf{x} - \mathbf{a})'\mathbf{A}(\mathbf{x} - \mathbf{a}) \leq c$ with $c > 0$ is an *ellipsoid* with center $\mathbf{a}$. By shifting the origin to $\mathbf{a}$ and rotating the ellipsoid, the latter can be expressed in a standard form $\sum_{i=1}^{n} \lambda_i z_i^2 \leq c$ with $\lambda_i > 0$ $(i = 1, 2, \ldots, n)$, where the $\lambda_i$ are the eigenvalues of $\mathbf{A}$. Setting all the $z_i$s equal to zero except $z_j$, we see that the lengths of the semi-major axes are $b_j = \sqrt{c/\lambda_j}$ for $j = 1, 2, \ldots, n$, and the volume of the ellipsoid is

$$\begin{aligned} v &= \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \prod_{j=1}^{n} b_j \\ &= \frac{\pi^{n/2} c^{n/2}}{\Gamma(\frac{n}{2}+1)(\det \mathbf{A})^{1/2}}, \end{aligned}$$

by (6.17c). Such a volume arises in finding the constant associated with various elliptical multivariate distributions such as the multivariate normal and the multivariate $t$-distributions (cf. Chapter 20).

**2.80.** (Quadrics) If $\mathbf{x} \in \mathbb{R}^n$, then a general quadric is $Q \equiv 0$, where $Q = \mathbf{x}'\mathbf{A}\mathbf{x} + 2\mathbf{b}'\mathbf{x} + c$ and $\mathbf{A}$ is an $n \times n$ symmetric matrix. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two points in $\mathbb{R}^n$ that we denote by $P_1$ and $P_2$, respectively. From (2.77), the coordinates of the point $P$ dividing the line $P_1P_2$ in the ratio $\mu : 1$ is given by $(1+\mu)^{-1}(\mathbf{x}_1 + \mu\mathbf{x}_2)$. Let $Q_{ij} = \mathbf{x}_i'\mathbf{A}\mathbf{x}_j + \mathbf{b}'\mathbf{x}_i + \mathbf{b}'\mathbf{x}_j + c$.

(a) Substituting for $P$ we find that $P$ lies on the quadric if

$$\mu^2 Q_{22} + 2\mu Q_{12} + Q_{11} = 0.$$

This is a quadratic in $\mu$ so that an arbitrary line meets a quadric in two points.

(b) (Tangent Plane) If $P_1$ lies on $Q = 0$, then $Q_{11} = 0$ and one root $\mu$ is zero. If $P_1P_2$ is a tangent, then the other root must also be zero; that is, the sum of the roots is zero and $Q_{12} = 0$. As $P_2$ varies subject to $Q_{12} = 0$, $P_2$ lies on $Q_1 = 0$, so that

$$\mathbf{x}_1'\mathbf{A}\mathbf{x} + \mathbf{b}'(\mathbf{x}_1 + \mathbf{x}) + c = 0,$$

is the tangent plane at $\mathbf{x}_1$.

(c) (Tangent Cone) Suppose $P_1$ and $P_2$ are not on $Q = 0$, but $P_1P_2$ touches the quadric so that the equation in $\mu$ has equal roots, i.e., $Q_{11}Q_{22} = Q_{12}^2$. Therefore as $P_2$ varies subject to this condition, we trace out the tangent cone from $P_1$, namely,

$$Q_{11}Q = Q_1^2.$$

(d) (Envelope) Suppose $Q = \mathbf{x}'\mathbf{A}\mathbf{x} - 1 \equiv 0$, where $\mathbf{A}$ is nonsingular, is a central quadric (i.e., $\mathbf{b} = \mathbf{0}$). Then using (c), $\mathbf{a}'\mathbf{x} = 1$ touches the quadric if $\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} = 1$. As $\mathbf{a}$ varies, $\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} = 1$ is the envelope equation.

### 2.6.3 Areas and Volumes

**2.81.** In two dimensions the area of a triangle with vertices $(x_i, y_i)'$, $i = 1, 2, 3$ is $\frac{1}{2}|\Delta|$, where

$$\Delta = \det \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{pmatrix}.$$

The three points are collinear if and only if $\Delta = 0$.

**2.82.** If $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$, where the $\mathbf{v}_i$ are vectors in $\mathbb{R}^n$, then the square of the two-dimensional volume of the parallelotope with $\mathbf{v}_1, \ldots, \mathbf{v}_p$ as principal edges is $\det(\mathbf{V}'\mathbf{V})$. A 2-dimensional parallelotope is a parallelogram; in this case we get the square of the area. When $p = 3$ we have the conventional parallelopiped. For statistical applications see Anderson [2003: section 7.5].

**2.83.** From (2.74), the four points $(x_i, y_i, z_i)'$, $i = 1, 2, 3, 4$, in three dimensions are coplanar if and only if

$$\det \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{pmatrix} = 0.$$

**Proofs.** Section 2.6.3.

2.81. Cullen [1997: 121].

2.82. Anderson [2003: 266]. For the area of a parallelogram see Basilevsky [1983: 64].

This Page Intentionally Left Blank

# CHAPTER 3

# RANK

The concept of rank undergirds much of matrix theory. In statistics it is frequently linked to the concept of degrees of freedom. Both equalities and inequalities are considered in this chapter, and partitioned matrices play an important role.

## 3.1 SOME GENERAL PROPERTIES

All the matrices in this section are defined over a general field $\mathcal{F}$, unless otherwise stated.

***Definition* 3.1.** The *rank*, denoted by $\operatorname{rank}\mathbf{A}$ $(= r$, say$)$, of a matrix $\mathbf{A}$ is $\dim \mathcal{C}(\mathbf{A})$, the dimension of the column space of $\mathbf{A}$. Here $r$ is also called the *column rank* of $\mathbf{A}$. The *row rank* is $\dim \mathcal{C}(\mathbf{A}')$. If $\mathbf{A}$ is $m \times n$ of rank $m$ (respectively $n$), then $\mathbf{A}$ is said to have *full row (respectively column) rank*. An $n \times n$ matrix $\mathbf{A}$ is said to be *nonsingular* if $\operatorname{rank}\mathbf{A} = n$.

As noted in Section 2.2.4, an associated vector space of $\mathcal{C}(\mathbf{A})$ is the null space $\mathcal{N}(\mathbf{A})$, and its dimension is called the *nullity*.

**3.1.** $\operatorname{rank}\mathbf{A}' = \operatorname{rank}\mathbf{A} = r$ so that the row rank equals the column rank.

**3.2.** Let $\mathbf{A}$ be an $m \times n$ matrix of rank $r$ $(r \leq \min\{m, n\})$.

(a) $\mathbf{A}$ has $r$ linearly independent columns and $r$ linearly independent rows.

(b) There exists an $r \times r$ nonzero principal minor. When $r < \min\{m, n\}$, all principal minors of larger order than $r$ are zero.

(c) If $\mathbf{B}$ is $m \times p$ and $\mathcal{C}(\mathbf{B}) \subseteq \mathcal{C}(\mathbf{A})$, then $\operatorname{rank} \mathbf{B} \leq \operatorname{rank} \mathbf{A}$.

**3.3.** Let $\mathbf{A}$ be an $m \times n$ matrix over $\mathbb{F}$.

(a) $\operatorname{rank} \mathbf{A} + \operatorname{nullity} \mathbf{A} =$ number of columns of $\mathbf{A}$.

(b) Suppose $\mathbf{A}$ is real, then

$$\operatorname{rank}(\mathbf{A}'\mathbf{A}) = \operatorname{rank}(\mathbf{A}\mathbf{A}') = \operatorname{rank} \mathbf{A}.$$

(c) Suppose $\mathbf{A}$ is complex, then:

(i) $\operatorname{rank} \mathbf{A} = \operatorname{rank} \overline{\mathbf{A}}$.

(ii) Since $\operatorname{rank} \overline{\mathbf{A}} = \operatorname{rank} \overline{\mathbf{A}}'$ by (3.1), we have $\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{A}^*$.

(iii) $\operatorname{rank} \mathbf{A} = \operatorname{rank}(\mathbf{A}\mathbf{A}^*) = \operatorname{rank}(\mathbf{A}^*\mathbf{A})$.

Thus, combining the above,

$$\operatorname{rank} \mathbf{A} = \operatorname{rank} \overline{\mathbf{A}} = \operatorname{rank} \mathbf{A}^* = \operatorname{rank}(\mathbf{A}\mathbf{A}^*) = \operatorname{rank}(\mathbf{A}^*\mathbf{A}).$$

(d) If $\mathbf{A}$ is complex, it is not necessarily true that $\operatorname{rank} \mathbf{A}'\mathbf{A} = \operatorname{rank} \mathbf{A}$.

**3.4.** We consider two special cases of rank.

(a) If $\operatorname{rank} \mathbf{A} = 0$, then $\mathbf{A} = \mathbf{0}$. This is a simple but key result that can be used to prove the equality of two matrices.

(b) If $\operatorname{rank} \mathbf{A} = 1$, then there exist nonzero $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{A} = \mathbf{a}\mathbf{b}'$.

**3.5.** (Full-Rank Factorization) Any $m \times n$ real or complex matrix $\mathbf{A}$ of rank $r$ ($r > 0$) can be expressed in the form $\mathbf{A}_{m \times n} = \mathbf{C}_{m \times r}\mathbf{R}_{r \times n}$, where $\mathbf{C}$ and $\mathbf{R}$ have (full) rank $r$. We call this a *full-rank factorization*. The columns of $\mathbf{C}$ may be an arbitrary basis of $\mathcal{C}(\mathbf{A})$, and then $\mathbf{R}$ is uniquely determined, or else the rows of $\mathbf{R}$ may be an arbitrary basis of $\mathcal{C}(\mathbf{A}')$, and then $\mathbf{C}$ is uniquely determined. Note that $\mathbf{C}$ has a left inverse, namely $(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$, and $\mathbf{R}$ has a right inverse, $\mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}$. Two full-rank factorizations can be obtained from the singular value decomposition of $\mathbf{A}$ (cf. 16.34e).

**3.6.** If $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices, then $\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{B}$ if and only if there exist a nonsingular $m \times m$ matrix $\mathbf{C}$ and an $n \times n$ nonsingular matrix $\mathbf{D}$ such that $\mathbf{A} = \mathbf{CBD}$.

**3.7.** If $\mathcal{C}(\mathbf{B}) = \mathcal{C}(\mathbf{C})$, then $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}(\mathbf{AC})$ for all $\mathbf{A}$.

**3.8.** If $\mathbf{V}$ is Hermitian non-negative definite, then $\mathbf{V} = \mathbf{R}\mathbf{R}^*$ (by 10.10) and $\operatorname{rank}(\mathbf{AV}) = \operatorname{rank}(\mathbf{AR})$ for all $\mathbf{A}$.

**Proofs.** Section 3.1.

3.1. Abadir and Magnus [2005: 77–78].

3.2. (a) and (c) follow from the definition; for (b) see Meyer [2000a: 215].

3.3a. Follows from (2.37) and (c)(ii) below. See also Seber and Lee [2003: 458].

3.3b. Abadir and Magnus [2005: 81] and Meyer [2000a: 212].

3.3c(i). Rao and Bhimasankaram [2000: 145].

3.3c(iii). Ben-Israel and Greville [2003: 46] and Meyer [2000a: 212].

3.3d. For a counter example consider $\mathbf{A} = (1, i)'(1, 1)$.

3.4b. Abadir and Magnus [2005: 80].

3.5. Ben-Israel and Greville [2003: 26], Marsaglia and Styan [1974a: theorem 1], and Searle [1982: 175].

3.6. If $\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{B} = r$, then by (16.33a) $\mathbf{A}$ and $\mathbf{B}$ are equivalent to the same diagonal matrix. The converse follows from (3.14a).

3.7. Follows from $\mathcal{C}(\mathbf{AB}) = \mathcal{C}(\mathbf{AC})$.

3.8. By (10.10), $\mathbf{V} = \mathbf{RR}^*$ and from (2.35) we have $\mathcal{C}(\mathbf{V}) = \mathcal{C}(\mathbf{R})$. The result follows from (3.7).

## 3.2 MATRIX PRODUCTS

All the matrices in this section are real or complex.

**3.9.** Given conformable matrices $\mathbf{A}$ and $\mathbf{B}$, we have the following.

(a) $\operatorname{rank}(\mathbf{BA}) = \operatorname{rank} \mathbf{A}$ if $\mathbf{B}$ has full row rank.

(b) $\operatorname{rank}(\mathbf{AC}) = \operatorname{rank} \mathbf{A}$ if $\mathbf{C}$ has full column rank.

(c) $\operatorname{rank}(\mathbf{A}'\mathbf{AB}) = \operatorname{rank}(\mathbf{AB}) = \operatorname{rank}(\mathbf{ABB}')$.

**3.10.** Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ and $n \times p$ matrices, respectively. Then:

(a) $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank} \mathbf{B} - \dim\{[\mathcal{N}(\mathbf{A})]^{\perp} \cap \mathcal{C}(\mathbf{B})\}$.

(b) $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank} \mathbf{A} - \dim\{\mathcal{C}(\mathbf{A}') \cap [\mathcal{N}(\mathbf{B})]^{\perp}\}$.

The above results immediately give us conditions for $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank} \mathbf{A}$ and $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank} \mathbf{B}$. Other conditions are given in (3.13c) and (3.13d) below.

**3.11.** Let $\mathbf{A}$ be a square matrix. If $\operatorname{rank}(\mathbf{A}^m) = \operatorname{rank}(\mathbf{A}^{m+1})$, then $\operatorname{rank}(\mathbf{A}^m) = \operatorname{rank}(\mathbf{A}^n)$ for all $n \geq m$.

**3.12.** $\operatorname{rank}(\mathbf{AB}) \leq \min\{\operatorname{rank} \mathbf{A}, \operatorname{rank} \mathbf{B}\}$.

**3.13.** Let $\mathbf{A}$ have $n$ columns and $\mathbf{B}$ have $n$ rows. Let $\mathbf{A}^-$ and $\mathbf{B}^-$ be any weak inverses of $\mathbf{A}$ and $\mathbf{B}$, respectively. Then:

$$
\begin{aligned}
\operatorname{rank}\begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{B} & \mathbf{I}_n \end{pmatrix} &= \operatorname{rank} \mathbf{A} + \operatorname{rank}(\mathbf{B}, \mathbf{I}_n - \mathbf{A}^-\mathbf{A}) \\
&= \operatorname{rank}\begin{pmatrix} \mathbf{A} \\ \mathbf{I}_n - \mathbf{BB}^- \end{pmatrix} + \operatorname{rank} \mathbf{B} \\
&= \operatorname{rank} \mathbf{A} + \operatorname{rank} \mathbf{B} + \operatorname{rank}[(\mathbf{I}_n - \mathbf{BB}^-)(\mathbf{I}_n - \mathbf{A}^-\mathbf{A})] \\
&= n + \operatorname{rank}(\mathbf{AB}).
\end{aligned}
$$

We can deduce the following.

(a) $\operatorname{rank}(\mathbf{B}, \mathbf{I}_n - \mathbf{A}^-\mathbf{A}) = \operatorname{rank}\mathbf{B} + \operatorname{rank}[(\mathbf{I}_n - \mathbf{B}\mathbf{B}^-)(\mathbf{I}_n - \mathbf{A}^-\mathbf{A})]$.

(b) $\operatorname{rank}\begin{pmatrix} \mathbf{A} \\ \mathbf{I}_n - \mathbf{B}\mathbf{B}^- \end{pmatrix} = \operatorname{rank}\mathbf{A} + \operatorname{rank}[(\mathbf{I}_n - \mathbf{B}\mathbf{B}^-)(\mathbf{I}_n - \mathbf{A}^-\mathbf{A})]$.

(c) $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}\mathbf{A}$ if and only if $(\mathbf{B}, \mathbf{I}_n - \mathbf{A}^-\mathbf{A})$ has full row rank $n$.

(d) $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}\mathbf{B}$ if and only if $\left(\begin{smallmatrix} \mathbf{A} \\ \mathbf{I}_n - \mathbf{B}\mathbf{B}^- \end{smallmatrix}\right)$ has full column rank $n$.

(e) (Sylvester)
$$\operatorname{rank}(\mathbf{AB}) \geq \operatorname{rank}\mathbf{A} + \operatorname{rank}\mathbf{B} - n,$$
with equality if and only if $(\mathbf{I}_n - \mathbf{B}\mathbf{B}^-)(\mathbf{I}_n - \mathbf{A}^-\mathbf{A}) = \mathbf{0}$. This result also follows from the Frobenius inequality (3.18b) by setting $\mathbf{B} = \mathbf{I}_n$.
If $\mathbf{AB} = \mathbf{0}$, $\operatorname{rank}\mathbf{A} + \operatorname{rank}\mathbf{B} \leq n$.

**3.14.** Let $\mathbf{A}$ be any matrix.

(a) If $\mathbf{P}$ and $\mathbf{Q}$ are any conformable nonsingular matrices,
$$\operatorname{rank}(\mathbf{PAQ}) = \operatorname{rank}\mathbf{A}.$$

(b) If $\mathbf{C}$ has full column rank and $\mathbf{R}$ has full row rank, then
$$\operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{CA}) = \operatorname{rank}(\mathbf{AR}).$$

**3.15.** If $\mathbf{A}$ is $p \times q$ of rank $q$ and $\mathbf{B}$ is $q \times r$ of rank $r$, then $\mathbf{AB}$ is $p \times r$ of rank $r$.

**3.16.** If $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}\mathbf{A}$, then $\mathcal{C}(\mathbf{AB}) = \mathcal{C}(\mathbf{A})$.

**3.17.** Suppose that the following products of matrices exist. Then:

(a) $\operatorname{rank}(\mathbf{XA}) = \operatorname{rank}\mathbf{A}$ implies $\operatorname{rank}(\mathbf{XAF}) = \operatorname{rank}(\mathbf{AF})$ for every $\mathbf{F}$.

(b) $\operatorname{rank}(\mathbf{AY}) = \operatorname{rank}\mathbf{A}$ implies $\operatorname{rank}(\mathbf{KAY}) = \operatorname{rank}(\mathbf{KA})$ for every $\mathbf{K}$.

**3.18.** Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be conformable matrices, and let $(\mathbf{AB})^-$ and $(\mathbf{BC})^-$ be any weak inverses. Then:

(a)
$$\begin{aligned} \operatorname{rank}\begin{pmatrix} \mathbf{0} & \mathbf{AB} \\ \mathbf{BC} & \mathbf{B} \end{pmatrix} &= \operatorname{rank}(\mathbf{AB}) + \operatorname{rank}(\mathbf{BC}) + \operatorname{rank}\mathbf{L} \\ &= \operatorname{rank}\mathbf{B} + \operatorname{rank}(\mathbf{ABC}), \end{aligned}$$
where $\mathbf{L} = [\mathbf{I} - \mathbf{BC}(\mathbf{BC})^-]\mathbf{B}[\mathbf{I} - (\mathbf{AB})^-(\mathbf{AB})]$.

(b) (Frobenius Inequality) From (a) we have
$$\operatorname{rank}(\mathbf{ABC}) \geq \operatorname{rank}(\mathbf{AB}) + \operatorname{rank}(\mathbf{BC}) - \operatorname{rank}\mathbf{B},$$
with equality if and only if $\mathbf{L} = \mathbf{0}$.