

**SHAYLE R. SEARLE**

**MATRIX ALGEBRA USEFUL  
FOR STATISTICS**

**WILEY SERIES IN PROBABILITY  
AND MATHEMATICAL STATISTICS**



# Matrix Algebra Useful for Statistics

SHAYLE R. SEARLE

*Professor of Biological Statistics*

*Biometrics Unit*

*New York State College of Agriculture and Life Sciences*

*Cornell University,*

*Ithaca, N.Y.*



**JOHN WILEY & SONS**

New York • Chichester • Brisbane • Toronto • Singapore

Copyright © 1982 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

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

Searle, S. R. (Shayle R.), 1928-  
Matrix algebra useful for statistics.

(Wiley series in probability and mathematical statistics. Applied probability and statistics,

ISSN 0271-6356)

Bibliography: p.

Includes index.

I. Matrices. 2. Statistics. I. Title. II. Series.

QA188.S4 512.9'434 82-4862

ISBN 0-471-86681-4 AACR2

Printed in the United States of America

20 19 18 17 16 15

To my wife Helen

## Preface

Algebra is a mathematical shorthand for language, and matrices are a shorthand for algebra. Consequently, a special value of matrices is that they enable many mathematical operations, especially those arising in statistics and the quantitative sciences, to be expressed concisely and with clarity. The algebra of matrices is, of course, in no way new, but its presentation is often so surrounded by the trappings of mathematical generality that assimilation can be difficult for readers who have only limited ability or training in mathematics. Yet many such people nowadays find a knowledge of matrix algebra necessary for their work, especially where statistics and/or computers are involved. It is to these people that I address this book; and for them, I have attempted to keep the mathematical presentation as informal as possible.

The pursuit of knowledge frequently involves collecting data; and those responsible for the collecting must appreciate the need for analyzing their data to recover and interpret the information contained therein. Such people must therefore understand some of the mathematical tools necessary for this analysis, to an extent either that they can carry out their own analysis, or that they can converse with statisticians and mathematicians whose help will otherwise be needed. One of the necessary tools is matrix algebra. It is becoming as necessary to science today as elementary calculus has been for generations. Matrices originated in mathematics more than a century ago, but their broad adaptation to science is relatively recent, prompted by the widespread acceptance of statistical analysis of data, and of computers to do that analysis; both statistics and computing rely heavily on matrix algebra. The purpose of this book is therefore that of bringing to a broad spectrum of readers a knowledge of matrix algebra that is useful in the statistical analysis of data and in statistics generally.

The basic prerequisite for using the book is high school algebra. Differential calculus is used on only a few pages, which can easily be omitted;

nothing will be lost insofar as a general understanding of matrix algebra is concerned. Proofs and demonstrations of most of the theory are given, for without them the presentation would be lifeless. But in every chapter the theoretical development is profusely illustrated with elementary numerical examples and with illustrations taken from a variety of applied sciences. And the last three chapters are devoted solely to uses of matrix algebra in statistics, with Chapters 14 and 15 outlining two of the most widely used statistical techniques: regression and linear models.

The mainstream of the book is its first eleven chapters, beginning with one on introductory concepts that includes a discussion of subscript and summation notation. This is followed by four chapters dealing with basic arithmetic, special matrices, determinants and inverses. Chapters 6 and 7 are on rank and canonical forms, 8 and 9 deal with generalized inverses and solving linear equations, 10 is a collection of results on partitioned matrices and 11 describes eigenvalues and eigenvectors. Background theory for Chapter 11 is collected in an appendix, Chapter 11A, some summaries and miscellaneous topics make up Chapter 12, statistical illustrations constitute Chapter 13, and Chapters 14 and 15 describe regression and linear models. All chapters except the last two end with exercises.

Occasional sections and paragraphs can be omitted at a first reading, especially by those whose experience in mathematics is somewhat limited. These portions of the book are printed in small type and, generally speaking, contain material subsidiary to the main flow of the text—material that may be a little more advanced in mathematical presentation than the general level otherwise maintained.

Chapters, and sections within chapters, are numbered with Arabic numerals 1, 2, 3, . . . . Within-chapter references to sections are by section number, but references across chapters use the decimal system; e.g., Section 1.3 is Section 3 of Chapter 1. These numbers are also shown in the running head of each page; e.g., [1.3] is found on page 4. Numbered equations are (1), (2), . . . , within each chapter. Those of one chapter are seldom referred to in another, but when they are, the chapter reference is explicit; otherwise “equation (3),” or more simply “(3),” means the equation numbered (3) in the chapter concerned. Exercises are in un-numbered sections and are referenced by their chapter number; e.g., Exercise 6.2 is Exercise 2 at the end of Chapter 6.

I am greatly indebted to George P. H. Styan for his exquisitely thorough readings of two drafts of the manuscript and his extensive and very helpful array of comments. Harold V. Henderson's numerous suggestions for the final manuscript were equally as helpful. Readers of *Matrix Algebra for the*

*Biological Sciences* (Wiley, 1966), and students in fifteen years of my matrix algebra course at Cornell have also contributed many useful ideas. Particular thanks go to Mrs. Helen Seamon for her superb accuracy on the typewriter, her patience and fantastic attention to detail; such attributes are greatly appreciated.

*Ithaca, New York*  
*May 1982*

SHAYLE R. SEARLE

## List of Chapters

1.	Introduction	1
2.	Basic Operations	22
3.	Special Matrices	60
4.	Determinants	84
5.	Inverse Matrices	119
6.	Rank	155
7.	Canonical Forms	205
8.	Generalized Inverses	212
9.	Solving Linear Equations	227
10.	Partitioned Matrices	257
11.	Eigenvalues and Eigenvectors	272
11A.	Appendix to Chapter 11	305
12.	Miscellanea	320
13.	Applications in Statistics	346
14.	The Matrix Algebra of Regression Analysis	363
15.	An Introduction to Linear Statistical Models	392
	References	429
	Index	433

## Contents

<b>1. Introduction</b>	<b>1</b>
1. The scope of matrix algebra	1
2. General description of a matrix	3
3. Subscript notation	4
4. Summation notation	6
5. Dot notation	11
6. Definition of a matrix	12
7. Vectors and scalars	15
8. General notation	16
9. Illustrative examples	16
Exercises	17
<b>2. Basic Operations</b>	<b>22</b>
1. The transpose of a matrix	22
a. <i>A reflexive operation, 23</i>	
b. <i>Vectors, 24</i>	
2. Partitioned matrices	24
a. <i>Example, 24</i>	
b. <i>General specification, 26</i>	
c. <i>Transposing a partitioned matrix, 26</i>	
d. <i>Partitioning into vectors, 26</i>	
3. The trace of a matrix	27
4. Addition	28
5. Scalar multiplication	29
6. Subtraction	30
7. Equality and the null matrix	31

<b>2. Basic Operations (Continued)</b>	
8. Multiplication	32
<i>a. The inner product of two vectors, 32</i>	
<i>b. A matrix-vector product, 33</i>	
<i>c. A product of two matrices, 36</i>	
<i>d. Existence of matrix products, 38</i>	
<i>e. Products with vectors, 39</i>	
<i>f. Products with scalars, 42</i>	
<i>g. Products with null matrices, 43</i>	
<i>h. Products with diagonal matrices, 43</i>	
<i>i. Identity matrices, 44</i>	
<i>j. The transpose of a product, 44</i>	
<i>k. The trace of a product, 45</i>	
<i>l. Powers of a matrix, 46</i>	
<i>m. Partitioned matrices, 48</i>	
<i>n. Hadamard products, 49</i>	
9. The Laws of algebra	50
<i>a. Associative laws, 50</i>	
<i>b. The distributive law, 51</i>	
<i>c. Commutative laws, 51</i>	
10. Contrasts with scalar algebra	52
Exercises	53
<b>3. Special Matrices</b>	<b>60</b>
1. Symmetric matrices	60
<i>a. Products of symmetric matrices, 61</i>	
<i>b. Properties of <math>AA'</math> and <math>A'A</math>, 61</i>	
<i>c. Products of vectors, 63</i>	
<i>d. Sums of outer products, 64</i>	
<i>e. Elementary vectors, 65</i>	
<i>f. Skew-symmetric matrices, 65</i>	
2. Matrices having all elements equal	65
3. Idempotent matrices	68
4. Orthogonal matrices	69
<i>a. Definitions, 69</i>	
<i>b. Special cases, 71</i>	
(i) Helmert matrices, 71	
(ii) Givens matrices, 72	
(iii) Householder matrices, 72	
5. Quadratic forms	73
6. Positive definite matrices	76
Exercises	78

<b>4. Determinants</b>	<b>84</b>
1. Expansion by minors	84
<i>a. First- and second-order determinants, 85</i>	
<i>b. Third-order determinants, 86</i>	
<i>c. <math>n</math>-order determinants, 89</i>	
2. Formal definition	90
3. Basic properties	92
<i>a. Determinant of a transpose, 92</i>	
<i>b. Two rows the same, 93</i>	
<i>c. Cofactors, 93</i>	
<i>d. Adding multiples of a row (column) to a row (column), 95</i>	
<i>e. Products, 96</i>	
(i) Reduction to triangular form, 96	
(ii) Two useful lemmas, 97	
(iii) Determinant of a product, 98	
4. Elementary row operations	99
<i>a. Factorization, 101</i>	
<i>b. A row (column) of zeros, 102</i>	
<i>c. Interchanging rows (columns), 102</i>	
<i>d. Adding a row to a multiple of a row, 102</i>	
5. Examples	103
6. Diagonal expansion	106
7. The Laplace expansion	109
8. Sums and differences of determinants	111
Exercises	112
<b>5. Inverse Matrices</b>	<b>119</b>
1. Introduction: solving equations	119
2. Products equal to I	123
3. Cofactors of a determinant	125
4. Derivation of the inverse	125
5. Conditions for existence of the inverse	129
6. Properties of the inverse	130
7. Some simple special cases	131
<i>a. Inverses of order 2, 131</i>	
<i>b. Diagonal matrices, 132</i>	
<i>c. I and J matrices, 132</i>	
<i>d. Orthogonal matrices, 132</i>	
<i>e. Idempotent matrices, 133</i>	
8. Equations and algebra	133
<i>a. Solving linear equations, 133</i>	
(i) Age distributions in wild populations, 133	
(ii) Input-output analysis in economics, 134	
(iii) Least squares equations, 136	
<i>b. Algebraic simplifications, 136</i>	



<b>5. Inverse Matrices (Continued)</b>	
9. Computers and inverses	139
<i>a. The arithmetic of linear equations, 140</i>	
<i>b. Rounding error, 142</i>	
(i) Addition, 143	
(ii) Inverting a matrix, 144	
(iii) Solving linear equations, 145	
10. Left and right inverses	147
Exercises	148
<b>6. Rank</b>	<b>155</b>
1. Linear combinations of vectors	155
2. Linear transformations	157
3. Linear dependence and independence	159
<i>a. Definitions, 159</i>	
<i>b. General characteristics, 161</i>	
(i) Sets of vectors, 161	
(ii) Some $a$ 's zero, 161	
(iii) Existence and non-uniqueness of nonzero $a$ 's, 161	
(iv) Null vectors, 161	
4. Linearly dependent vectors	162
<i>a. At least two <math>a</math>'s are nonzero, 162</i>	
<i>b. Vectors are linear combinations of others, 162</i>	
<i>c. Partitioning matrices, 163</i>	
<i>d. Zero determinants, 164</i>	
<i>e. Inverse matrices, 164</i>	
<i>f. Testing for dependence (simple cases), 164</i>	
5. Linearly independent (LIN) vectors	166
<i>a. Nonzero determinants and inverse matrices, 166</i>	
<i>b. Linear combinations of LIN vectors, 167</i>	
<i>c. A maximum number of LIN vectors, 167</i>	
6. The number of LIN rows and columns in a matrix	169
7. The rank of a matrix	171
8. Rank and inverse matrices	172
9. Permutation matrices	173
10. Full-rank factorization	175
<i>a. Basic development, 175</i>	
<i>b. The general case, 177</i>	
<i>c. Matrices of full row (column) rank, 177</i>	
11. Vector spaces	177
<i>a. Euclidean space, 178</i>	
<i>b. Vector spaces, 178</i>	
<i>c. Spanning sets and bases, 179</i>	

<b>6. Rank (Continued)</b>	
<i>d. Many spaces of order <math>n</math>, 179</i>	
<i>e. Subspaces, 180</i>	
<i>f. The range and null space of a matrix, 180</i>	
Exercises	181
<b>7. Canonical Forms</b>	<b>184</b>
1. Elementary operators	184
<i>a. Row operations, 184</i>	
<i>b. Transposes, 185</i>	
<i>c. Column operations, 185</i>	
<i>d. Inverses, 185</i>	
2. Rank and the elementary operators	186
<i>a. Rank, 186</i>	
<i>b. Products of elementary operators, 186</i>	
<i>c. Equivalence, 187</i>	
3. Finding the rank of a matrix	187
<i>a. Some special LIN vectors, 187</i>	
<i>b. Calculating rank, 188</i>	
<i>c. A general procedure, 189</i>	
4. Reduction to equivalent canonical form	190
<i>a. Row operations, 190</i>	
<i>b. Column operations, 191</i>	
<i>c. The equivalent canonical form, 192</i>	
<i>d. Non-uniqueness of <math>\mathbf{P}</math> and <math>\mathbf{Q}</math>, 194</i>	
<i>e. Existence is assured, 194</i>	
<i>f. Full-rank factorization, 194</i>	
5. Rank of a product matrix	196
6. Symmetric matrices	199
<i>a. Row and column operations, 199</i>	
<i>b. The diagonal form, 200</i>	
<i>c. The canonical form under congruence, 201</i>	
<i>d. Two special provisions, 202</i>	
(i) Null rows, 202	
(ii) Negative elements in the diagonal form, 203	
<i>e. Full-rank factorization, 205</i>	
7. Non-negative definite matrices	205
<i>a. Diagonal elements and principal minors, 205</i>	
<i>b. Congruent canonical form, 206</i>	
<i>c. Full-rank factorization, 206</i>	
<i>d. Quadratic forms as sums of squares, 206</i>	
<i>e. Full row (column) rank matrices, 208</i>	
Exercises	209

<b>8. Generalized Inverses</b>	<b>212</b>
1. The Moore–Penrose inverse	212
2. Generalized inverses	212
a. Derivation from row operations, 214	
b. Derivation from the diagonal form, 215	
3. Other names and symbols	216
4. An algorithm	217
a. An easy form, 217	
b. A general form, 217	
5. Arbitrariness in a generalized inverse	219
6. Symmetric matrices	220
a. Non-negative definite matrices, 220	
b. A general algorithm, 221	
c. The matrix $X'X$ , 221	
Exercises	222
<b>9. Solving Linear Equations</b>	<b>227</b>
1. Equations having many solutions	227
2. Consistent equations	228
a. Definition, 228	
b. Existence of solutions, 229	
c. Tests for consistency, 232	
3. Equations having one solution	233
4. Deriving solutions using generalized inverses	235
a. Obtaining a solution, 235	
b. Obtaining many solutions, 236	
c. All possible solutions, 237	
d. Combinations of solutions, 239	
5. Linearly independent solutions	239
6. An invariance property	242
7. Equations $Ax = 0$	245
a. General properties, 245	
b. Orthogonal solutions, 246	
c. Orthogonal vector spaces, 248	
8. A complete example	249
9. Least squares equations	251
Exercises	252

<b>10. Partitioned Matrices</b>	<b>257</b>
1. Orthogonal matrices	257
2. Determinants	258
3. Inverses	260
4. Schur complements	261
5. Generalized inverses	261
6. Direct sums	263
7. Direct products	265
Exercises	267
<b>11. Eigenvalues and Eigenvectors</b>	<b>272</b>
1. Introduction: age distribution vectors	272
2. Derivation of eigenvalues	274
3. Elementary properties of eigenvalues	276
a. Eigenvalues of powers of a matrix, 276	
b. Eigenvalues of a scalar-by-matrix product, 277	
c. Eigenvalues of polynomials, 277	
d. The sum and product of eigenvalues, 278	
4. Calculating eigenvectors	279
a. A general method, 279	
b. Simple roots, 280	
c. Multiple roots, 281	
5. The similar canonical form	282
a. Derivation, 282	
b. Uses, 285	
6. Symmetric matrices	290
a. Eigenvalues all real, 290	
b. Symmetric matrices are diagonal, 290	
c. Eigenvectors are orthogonal, 290	
(i) Different eigenvalues, 291	
(ii) Multiple eigenvalues, 291	
(iii) The canonical form under orthogonal similarity, 291	
d. Rank equals number of nonzero eigenvalues, 292	
7. Dominant eigenvalues	293
8. Factoring the characteristic equation	298
Exercises	299

<b>11A. Appendix to Chapter 11</b>	<b>305</b>
1. Proving the diagonalability theorem	305
<i>a. The number of nonzero eigenvalues never exceeds rank, 305</i>	
<i>b. A lower bound on <math>r(\mathbf{A} - \lambda_k \mathbf{I})</math>, 306</i>	
<i>c. Proof of the diagonalability theorem, 307</i>	
<i>d. All symmetric matrices are diagonalable, 307</i>	
2. Other results for symmetric matrices	308
<i>a. Spectral decomposition, 308</i>	
<i>b. Non-negative definite (n.n.d.) matrices, 309</i>	
<i>c. Simultaneous diagonalization of two symmetric matrices, 312</i>	
3. The Cayley–Hamilton theorem	314
4. The singular-value decomposition	316
Exercises	318
<b>12. Miscellanea</b>	<b>320</b>
1. Orthogonal matrices—a summary	320
2. Idempotent matrices—a summary	320
3. The matrix $a\mathbf{I} + b\mathbf{J}$ —a summary	322
4. Non-negative definite matrices—a summary	322
5. Canonical forms and other decompositions—a summary	323
6. Matrix Functions	325
<i>a. Functions of matrices, 325</i>	
<i>b. Matrices of functions, 325</i>	
7. Iterative solution of nonlinear equations	326
8. Vectors of differential operators	327
<i>a. Scalars, 327</i>	
<i>b. Vectors, 328</i>	
<i>c. Quadratic forms, 329</i>	
9. Vec and vech operators	332
<i>a. Definitions, 332</i>	
<i>b. Properties of vec, 333</i>	
<i>c. Vec-permutation matrices, 334</i>	
<i>d. Relationships between vec and vech, 334</i>	
10. Other calculus results	334
<i>a. Differentiating inverses, 335</i>	
<i>b. Differentiating traces, 335</i>	
<i>c. Differentiating determinants, 336</i>	
<i>d. Jacobians, 338</i>	
<i>e. Aitken's integral, 340</i>	
<i>f. Hessians, 341</i>	
11. Matrices with elements that are complex numbers	341
Exercises	342

<b>13. Applications in Statistics</b>	<b>346</b>
1. Variance–covariance matrices	347
2. Correlation matrices	348
3. Matrices of sums of squares and cross-products	349
<i>a. Data matrices, 349</i>	
<i>b. Uncorrected sums of squares and products, 349</i>	
<i>c. Means, and the centering matrix, 350</i>	
<i>d. Corrected sums of squares and products, 351</i>	
4. The multivariate normal distribution	352
5. Quadratic forms and $\chi^2$ -distributions	355
6. Least squares equations	357
7. Contrasts among means	358
Exercises	360
<b>14. The Matrix Algebra of Regression Analysis</b>	<b>363</b>
1. General description	363
<i>a. Linear models, 364</i>	
<i>b. Observations, 364</i>	
<i>c. Nonlinear models, 365</i>	
2. Estimation	366
3. Several regressor variables	368
4. Deviations from means	369
5. The statistical model	371
6. Unbiasedness and variances	372
7. Predicted $y$ -values	373
8. Estimating the error variance	374
9. Partitioning the total sum of squares	376
10. Multiple correlation	378
11. The no-intercept model	379
12. Analysis of variance	380
13. Testing linear hypotheses	382
<i>a. Stating a hypothesis, 382</i>	
<i>b. The <math>F</math>-statistic, 383</i>	
<i>c. Equivalent statements of a hypothesis, 385</i>	
<i>d. Special cases, 386</i>	
(i) $H: \mathbf{b} = \mathbf{0}$ , 386	
(ii) $H: \mathbf{b} = \mathbf{b}_0$ , 386	
(iii) $H: \lambda' \mathbf{b} = m$ , 386	
14. Confidence intervals	386
15. Fitting subsets of the $x$ -variables	387
16. Reductions in sums of squares: the $R(\cdot   \cdot)$ notation	389

<b>15. An Introduction to Linear Statistical Models</b>	<b>392</b>
1. General description	392
2. The normal equations	395
a. A general form, 395	
b. Many solutions, 396	
3. Solving the normal equations	397
a. Generalized inverses of $X'X$ , 397	
b. Solutions, 397	
4. Expected values and variances	399
5. Predicted $y$ -values	400
6. Estimating the error variance	401
a. Error sum of squares, 401	
b. Expected value, 402	
c. Estimation, 402	
7. Partitioning the total sum of squares	403
8. Coefficient of determination	404
9. Analysis of variance	405
10. The $R(\cdot   \cdot)$ notation	407
11. Estimable functions	408
12. Testing linear hypotheses	411
13. Confidence intervals	415
14. Some particular models	416
a. The one-way classification, 416	
b. Two-way classification, no interactions, balanced data, 418	
c. Two-way classification, no interactions, unbalanced data, 422	
15. The $R(\cdot   \cdot)$ notation (Continued)	424
<b>References</b>	<b>429</b>
<b>Index</b>	<b>433</b>

## Matrix Algebra Useful for Statistics

## CHAPTER 1

# INTRODUCTION

---

### 1. THE SCOPE OF MATRIX ALGEBRA

Science is today becoming increasingly quantitative. Scientists are being confronted more and more with large amounts of numerical data, measurements of one form or another gathered from their laboratories, field experiments and surveys. But the mere collecting and recording of data achieves nothing; having been collected, data must be analyzed and interpreted. Mathematics is made great use of in describing this analysis and interpretation, and one of the most useful branches of mathematics for this purpose is matrix algebra. It is useful not only in both simplifying description and promoting development of many analysis methods but also in organizing computer techniques to execute those methods and to present the results. This modern-day usage for matrices is why quantitative scientists need to have, as part of their mathematical repertoire, an understanding of matrix algebra. As a branch of mathematics, it dates back more than a century, but its applications in today's world are widespread, with particular value to statistics.

Matrices are simply rectangular arrays of numbers set out in rows and columns; matrix algebra is the algebra of these arrays. It is this very nature of matrices, of their being arrays of numbers with each array treated as a single entity, that makes them so useful in handling data. The algebra of such entities is a succinct representation of a wealth of mathematical manipulations.

Numerous examples are cited throughout this book of applying matrix algebra to real-life situations. One application is in population dynamics wherein studying a biological population we want to investigate the distribution of individuals according to their age. If  $n_t$  represents a series of

values describing the age distribution at time  $t$ , then the relationships between successive age distributions can often be described as  $\mathbf{n}_t = \mathbf{M}\mathbf{n}_{t-1}$  for some appropriate matrix  $\mathbf{M}$ . As we shall see, an equation such as this is simple enough in the context of matrix algebra, yet it has contributed greatly to the study of stable age distributions and logistic growth, yielding results that might have been obtained only with great difficulty without the use of matrix algebra. Some of the methods involved are mentioned in Chapter 12. Another example is the calculation of transition probabilities over a range of time intervals. This is achieved easily using matrix algebra, by assembling the transition probabilities in a matrix, known as a transition probability matrix, and by taking successive powers of that matrix. This is discussed in Section 6 of this chapter and again in Chapter 12.

A third example of the brevity and universality of matrix expressions occurs in the statistical method of regression analysis. This involves calculating values  $b_0, b_1, \dots, b_k$  for the equation

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

where we have a series of observations on the variable  $y$  and on each of the  $k$  variables  $x_1, x_2, \dots, x_k$ . With proper definition the  $b$ 's are obtained by solving the equations  $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$  as  $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , where  $\mathbf{X}$  and  $\mathbf{y}$ , both matrices, represent all the observations on the  $x$  and  $y$  variables respectively, and  $\mathbf{b}$  represents the series of  $b$ 's. Again, the brevity of such expressions is evident, as is their universality when one realizes that they apply no matter how many  $x$ -variables there are and no matter how numerous the observations on each. Once the necessary matrix algebra is assimilated [for example, the meanings attached to  $\mathbf{X}$ ,  $\mathbf{X}'$  and  $(\mathbf{X}'\mathbf{X})^{-1}$ ], the method of expressing the results is the same for all manner of situations. Chapter 14 deals with this topic in detail and Chapter 15 is concerned with extensions of it.

The manipulative skill required for using matrices in these and numerous other situations of applying mathematics to biological and other data is not demanding. Achieving some competence in this skill is important, because in terms of describing voluminous calculations with brevity, simplicity and clarity, the value of matrix algebra is enormous. Furthermore, the broad applicability of matrix expressions is exceedingly valuable because matrix results can so often be applied, with only minor changes, to situations involving both small amounts of data and to those involving extremely large amounts. Matrix algebra is therefore a vehicle by which mathematical procedures for many problems, both large and small, can be described independently of the size of the problem. Size affects not the understanding of the procedures, only the amount of computing that will be involved.

## 2. GENERAL DESCRIPTION OF A MATRIX

The feature of a matrix that makes it an aid in organizing data provides a convenient starting point for its description. We begin with an illustration and proceed to a formal definition in Section 6.

**Illustration.** Suppose the percentage of sterile cultures found in four generations of three experimental populations of a laboratory organism are as shown in Table 1.1.

TABLE 1.1 PERCENT STERILE CULTURES

Generation	Population		
	1	2	3
1	18	17	11
2	19	13	6
3	6	14	9
4	9	11	4

Consider the array of numbers in the table extracted and written simply as

$$\begin{bmatrix} 18 & 17 & 11 \\ 19 & 13 & 6 \\ 6 & 14 & 9 \\ 9 & 11 & 4 \end{bmatrix},$$

where the position of an entry in this array determines its meaning; for example, the entry in the second row and third column, 6, represents the percentage of sterile cultures observed in the second generation of the third population. In this way a row represents the percentage of sterile cultures seen in the same generation of all populations, and a column represents the percentage of sterile cultures seen in the same population at all generations. For example, the first row shows the percentage of sterile cultures observed in the first generation of the three populations, and the first column shows the percentage of sterile cultures seen in the four generations of the first population. Such an array of numbers is called a *matrix*.

The first thing to notice about any matrix is that it is a rectangular or square array, with the entries therein set out in rows and columns. The elements of each row usually have something in common, as do those of

each column. The individual entries in the array are called the *elements* or *terms* of the matrix; in general they can be numbers of any sort (real or complex, rational or irrational) or even functions of one or more variables. For our purposes we will usually think of them as being real numbers, positive, negative or zero. In this context this book deals almost exclusively with what are called *real matrices*: matrices having elements that are real numbers (and not complex numbers involving  $\sqrt{-1}$ ).

*Matrix algebra*, being the algebra of matrices, is the algebra of arrays such as we just have described, each treated as an entity and denoted by a single symbol. The algebra involves the elements within the arrays, but it is the handling of the arrays themselves, as separate entities, that constitutes matrix algebra.

Some simple mathematical notation that is widely used in matrix algebra is now described, before proceeding to a formal definition of a matrix. (Readers familiar with subscript and summation notation can readily omit Sections 3, 4 and 5.)

### 3. SUBSCRIPT NOTATION

Algebra is arithmetic with letters of the alphabet representing numbers. Thus the first two rows of the matrix given previously,

$$\begin{bmatrix} 18 & 17 & 11 \\ 19 & 13 & 6 \end{bmatrix}, \text{ might be written as } \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

where  $a, b, c, d, e$  and  $f$  represent the numbers 18, 17, 11, 19, 13 and 6, respectively. To this matrix, the array thought of as a self-contained entity, might be given the name  $\mathbf{A}$ . To emphasize that it is an entity in its own right, representing not just a single number but a whole array of numbers, we use the customary boldface notation for matrices, i.e.,  $\mathbf{A}$ , and then write

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}.$$

More generally,  $\mathbf{A}$  in this form can be thought of as being any matrix of two rows and three columns.

Since a matrix symbolized in this manner using the Roman alphabet is limited to only 26 letters, a notation that is less restrictive is more useful. Instead of specifying each element by a different letter we use one letter for a series of elements, modified by attaching to it a series of integers to

provide the necessary identification. Thus the matrix  $\mathbf{A}$  could be written as

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}.$$

The integers 1, 2, and 3 are called *subscripts* and the elements of the first row, for example, are read as "a one," "a two" and "a three," or as "a sub one," "a sub two" and "a sub three." In this instance the subscripts describe the column in which each element belongs and all elements in the same column have a common subscript; for example, the second column elements have subscript 2.

The principle just described can be applied to the rows and columns of a matrix simultaneously, using a single letter of the alphabet for all elements. Two subscripts written alongside one another are used for each element; the first specifies the row the element is in and the second the column. Thus  $\mathbf{A}$  could be written as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}.$$

The elements are read as "a one one," "a one two" or as "a sub one one," "a sub one two" and so on. In this way an element's subscripts uniquely locate its position in the matrix. For just as we have used the letter  $a$  to denote the value of an element of a matrix, so we can also use other letters to denote other things; for example, we might refer to a row as "row  $i$ ," and were  $i$  equal to 2 "row  $i$ " would be the second row of the matrix. In the same way we can use  $j$  for referring to columns, and "column  $j$ " or the " $j$ th column" is, for  $j = 3$ , the third column. By this means we refer to the element in the  $i$ th row and  $j$ th column of a matrix as  $a_{ij}$ , using  $i$  and  $j$  as subscripts to the letter  $a$  to denote the row and column the element is in, just as  $a_{12}$ , for example, is in the first row and second column. In general then,  $a_{ij}$  is in the  $i$ th row and  $j$ th column; thus  $a_{ij}$  for  $i = 2$  and  $j = 3$  is  $a_{23}$ , the element in the second row and third column. No comma is placed between the two subscripts unless it is needed to avoid confusion. For example, were there 12 columns in a matrix, the element in row 1 and column 12 would be written as  $a_{1,12}$  to distinguish it from  $a_{11,2}$ , the element in the eleventh row and second column.

This notation provides ready opportunity to detail the elements of a row, a column or even a complete matrix, in a very compact manner. Thus the first row of  $\mathbf{A}$ ,  $a_{11}, a_{12}, a_{13}$ , can be described as  $a_{1j}$  for  $j = 1, 2, 3$ . By so doing, we mean that the elements of the row are  $a_{1j}$  for  $j$  taking in turn each

one of the integer values 1, 2, and 3. Columns can be described in like fashion. The whole matrix is accounted for by enclosing  $a_{ij}$  in curly brackets and writing

$$A = \{a_{ij}\} \text{ for } i = 1, 2 \text{ and } j = 1, 2, 3.$$

This notation completely specifies the elements of the matrix, its name and its size. It is used in Section 1.6 to specify

$$A = \{a_{ij}\} \text{ for } i = 1, 2, \dots, r, \text{ and } j = 1, 2, \dots, c$$

as the general matrix of  $r$  rows and  $c$  columns, the dots indicating that  $i$  takes all integer values from 1 through  $r$  and  $j$  takes all integer values from 1 through  $c$ .

#### 4. SUMMATION NOTATION

The most frequently used arithmetic operation is that of adding numbers together. Subscript notation provides a means of describing this operation very succinctly. Suppose we wish to add five numbers represented as  $a_1, a_2, a_3, a_4$  and  $a_5$ . Their sum is clearly

$$a_1 + a_2 + a_3 + a_4 + a_5.$$

It can be put in words as

“the sum of all values of  $a_i$  for  $i = 1, 2, \dots, 5$ .”

The phrase “the sum of all values of” is customarily represented by the capital form of the Greek letter sigma,  $\Sigma$ . Accordingly the sum of the  $a$ 's is written as

$$\sum a_i \text{ for } i = 1, 2, \dots, 5.$$

A further abbreviation is  $\sum_{i=1}^{i=5} a_i$ , where the phrase “for  $i = 1, 2, \dots, 5$ ” has been replaced by the “ $i = 1$ ” and “ $i = 5$ ” below and above the  $\Sigma$  sign. This indicates that the summation is for all integer values of  $i$  from 1 through 5. Thus

$$\sum_{i=1}^{i=5} a_i = a_1 + a_2 + a_3 + a_4 + a_5.$$

[1.4]

Many variations on this summation notation are to be found. Almost always, the “ $i = 5$ ” above the  $\Sigma$  is reduced to just 5, so that the usual form is

$$\sum_{i=1}^5 a_i = a_1 + a_2 + a_3 + a_4 + a_5.$$

Similarly,

$$\sum_{i=1}^3 x_i = x_1 + x_2 + x_3$$

and

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_{n-2} + x_{n-1} + x_n,$$

the sum of  $n$   $x$ 's. Notice the algebraic expressions for subscripts,  $n - 2$  and  $n - 1$ . This is permissible, and occasionally quite extensive expressions are used this way. Sometimes the terms above and below the  $\Sigma$  sign are omitted when they are obvious from the context, and sometimes they are used as subscripts and superscripts to the  $\Sigma$ ; for example,

$$\sum_1^4 y_i = y_1 + y_2 + y_3 + y_4.$$

Another variant of the notation is that the lower limit of  $i$  need not necessarily be the value 1: with  $i = 3$  written below the  $\Sigma$  sign we have

$$\sum_{i=3}^7 y_i = y_3 + y_4 + y_5 + y_6 + y_7.$$

Furthermore, although subscripts in a summation are usually consecutive integers, omissions can be accounted for:

$$\sum_{\substack{i=3 \\ i \neq 4}}^7 y_i = y_3 + y_5 + y_6 + y_7.$$

Summation notation has so far been described in terms of simple sums, but it also encompasses sums of squares and sums of products, and indeed sums of any series of expressions that can be identified by subscript



notations. Thus the sum of squares  $c_1^2 + c_2^2 + c_3^2 + c_4^2$  can be written as

$$\sum_{i=1}^4 c_i^2 = c_1^2 + c_2^2 + c_3^2 + c_4^2.$$

Similarly, if we have two series of numbers  $p_1, p_2, p_3$  and  $q_1, q_2, q_3$  the sum of their products term by term,  $p_1q_1 + p_2q_2 + p_3q_3$ , is written

$$\sum_{i=1}^3 p_iq_i = p_1q_1 + p_2q_2 + p_3q_3.$$

The notation we have described applies equally well to cases involving two subscripts: in summing over either subscript the other remains unchanged. For example,

$$\sum_{j=1}^3 a_{1j} = a_{11} + a_{12} + a_{13} \quad \text{and} \quad \sum_{i=1}^2 a_{i2} = a_{12} + a_{22}$$

are specific instances of the general results

$$\sum_{j=1}^3 a_{1j} = a_{11} + a_{12} + a_{13} \quad \text{and} \quad \sum_{i=1}^2 a_{ij} = a_{1j} + a_{2j}.$$

Summing each of the latter leads to the self-explanatory operation of double summation:

$$\begin{aligned} \sum_{i=1}^2 \left( \sum_{j=1}^3 a_{ij} \right) &= \sum_{i=1}^2 (a_{i1} + a_{i2} + a_{i3}) \\ &= \sum_{i=1}^2 a_{i1} + \sum_{i=1}^2 a_{i2} + \sum_{i=1}^2 a_{i3} \\ &= a_{11} + a_{21} + a_{12} + a_{22} + a_{13} + a_{23}. \end{aligned}$$

Similarly,

$$\begin{aligned} \sum_{j=1}^3 \left( \sum_{i=1}^2 a_{ij} \right) &= \sum_{j=1}^3 (a_{1j} + a_{2j}) \\ &= \sum_{j=1}^3 a_{1j} + \sum_{j=1}^3 a_{2j} \\ &= a_{11} + a_{12} + a_{13} + a_{21} + a_{22} + a_{23}. \end{aligned}$$

which is seen at once to be the same as  $\sum_{i=1}^2 \left( \sum_{j=1}^3 a_{ij} \right)$ . Removing the brackets gives the important result

$$\sum_{i=1}^2 \sum_{j=1}^3 a_{ij} = \sum_{j=1}^3 \sum_{i=1}^2 a_{ij},$$

thus illustrating the general principle that the order of summation in double summation is of no consequence. In general

$$\sum_{i=1}^m \sum_{j=1}^n a_{ij} = \sum_{j=1}^n \sum_{i=1}^m a_{ij}.$$

In terms of a matrix of  $m$  rows and  $n$  columns the left-hand side is the sum of all the row totals and the right-hand side is the sum of all the column totals, both sums equaling the total of all elements.

Sums of squares and products can be written in a similar fashion:

$$\sum_{j=1}^4 a_{1j}^2 = a_{11}^2 + a_{12}^2 + a_{13}^2 + a_{14}^2,$$

$$\sum_{i=1}^3 a_{ij}b_{ij} = a_{1j}b_{1j} + a_{2j}b_{2j} + a_{3j}b_{3j}$$

and

$$\sum_{j=1}^3 a_{1j}b_{j1} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}.$$

Some readers may question the use of  $j$  in  $b_{j1}$  in this last expression since until now we have used  $i$  as the first subscript. There is, however, nothing sacrosanct about  $i$  and  $j$  as subscripts, any letters may be so used. Certainly  $i$  and  $j$  are commonly found in this role, both in this book and elsewhere, but they are by no means the only letters used as subscripts; for example,

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^3 a_{ij} &= \sum_{p=1}^2 \sum_{q=1}^3 a_{pq} = \sum_{k=1}^2 \sum_{t=1}^3 a_{kt} \\ &= a_{11} + a_{12} + a_{13} + a_{21} + a_{22} + a_{23}. \end{aligned}$$

The expression  $\sum_{j=1}^3 a_{1j}b_{j1}$  developed earlier is an example of the more general form

$$\sum_{j=1}^n a_{1j}b_{jk} = a_{11}b_{1k} + a_{12}b_{2k} + a_{13}b_{3k} + \cdots + a_{1n}b_{nk},$$

an expression used in multiplying two matrices together. It is discussed in detail in the next chapter (Section 2.7), but to anticipate for a moment we may note that if

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix},$$

then

$$\sum_{j=1}^3 a_{1j}b_{j1} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}$$

is the sum of the term by term products of the elements of the first row of  $\mathbf{A}$  and the first column of  $\mathbf{B}$ . Similarly,

$$\sum_{j=1}^3 a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + a_{i3}b_{3k}$$

is the sum of the term by term products of the elements of the  $i$ th row of  $\mathbf{A}$  and the  $k$ th column of  $\mathbf{B}$ .

Another attribute of summation operations may be noted. It involves summation of terms that do not have subscripts; for example,

$$\sum_{i=1}^4 x = x + x + x + x = 4x$$

and

$$\sum_{i=1}^3 ky_i = ky_1 + ky_2 + ky_3 = k \left( \sum_{i=1}^3 y_i \right).$$

Generalizations of these results are obvious:

$$\sum_{i=1}^n x = nx \quad \text{and} \quad \sum_{i=1}^n ky_i = k \left( \sum_{i=1}^n y_i \right).$$

It is appropriate to mention at this point a multiplication procedure analogous to  $\Sigma$ . It is denoted by the capital Greek letter pi,  $\Pi$ , and involves multiplying together all the terms to which it applies. Thus whereas

$$\sum_{i=1}^5 b_i = b_1 + b_2 + b_3 + b_4 + b_5,$$

$$\prod_{i=1}^5 b_i = b_1 b_2 b_3 b_4 b_5,$$

the product of all the  $b$ 's. Operationally,  $\Pi$  is similar to  $\Sigma$  except that it denotes multiplication instead of addition.

## 5. DOT NOTATION

A further abbreviation used extensively is

$$\sum_{i=1}^m a_{ij} = a_{.j}.$$

The square of this sum is

$$a_{.j}^2 = (a_{.j})^2 = \left( \sum_{i=1}^m a_{ij} \right)^2 = (a_{1j} + a_{2j} + \cdots + a_{mj})^2,$$

which is not the same as the sum of squares

$$\sum_{i=1}^m a_{ij}^2 = a_{1j}^2 + a_{2j}^2 + \cdots + a_{mj}^2$$

for which there is, generally speaking, no special symbol.

The dot subscript in  $a_{.j}$  in place of  $i$  denotes that summation has taken place over the  $i$  subscript. The symbol  $a_{.j}$  is read "a dot jay." Since the

notation  $a_{ij}$  shows no indication of the limits of  $i$  over which summation has occurred, it is used only when these limits are clear from the context of its use. In line with  $a_{ij}$  we also have

$$a_i = \sum_{j=1}^n a_{ij}$$

and

$$a_{ij} = \sum_{i=1}^m a_i = \sum_{j=1}^n a_j = \sum_{i=1}^m \sum_{j=1}^n a_{ij}.$$

## 6. DEFINITION OF A MATRIX

A matrix is a rectangular or square array of numbers arranged in rows and columns. The rows are of equal length, as are the columns. In terms of the notation of Section 3 we will let  $a_{ij}$  denote the element in the  $i$ th row and  $j$ th column of a matrix  $A$ , and if  $A$  has  $r$  rows and  $c$  columns it is written as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & a_{r3} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix}$$

The three dots indicate, in the first row, for example, that the elements  $a_{11}$ ,  $a_{12}$  and  $a_{13}$  continue in sequence up to  $a_{1j}$  and then on up to  $a_{1c}$ ; likewise in the first column the elements  $a_{11}$ ,  $a_{21}$ , continue in sequence to  $a_{i1}$  and on up to  $a_{r1}$ . The use of three dots to represent omitted values of a long sequence in this manner is standard and will be used extensively. This form of writing a matrix clearly specifies its elements, and also its size, i.e., the number of rows and columns. An alternative and briefer form is

$$A = \{a_{ij}\}, \quad \text{for } i = 1, 2, \dots, r, \quad \text{and } j = 1, 2, \dots, c,$$

the curly brackets indicating that  $a_{ij}$  is a typical element, the limits of  $i$  and  $j$  being  $r$  and  $c$ , respectively.

The element  $a_{ij}$  is called the  $ij$ th element, the first subscript referring to the row the element is in and the second to the column. Thus  $a_{23}$  is the

element in the second row and third column. The size of the matrix, i.e., the number of rows and columns, is referred to as its *order*, or as its *dimensions*. (Sometimes the word *order* is also used for other characteristics of a matrix, but in this book it always refers to the number of rows and columns.) Thus  $A$  with  $r$  rows and  $c$  columns has order  $r \times c$  (read as " $r$  by  $c$ ") and, to emphasize its order, the matrix can be written  $A_{r \times c}$ . A simple example of a matrix of order  $2 \times 3$  is

$$A_{2 \times 3} = \begin{bmatrix} 4 & 0 & -3 \\ -7 & 2.73 & 1 \end{bmatrix}.$$

Notice that zero is legitimate as an element, that the elements need not all have the same sign, and that integers and decimal numbers can both be elements of the same matrix. The element in the first row and first column,  $a_{11}$ , the element in the upper left-hand corner of a matrix is called the *leading element* of the matrix.

When  $r = c$ , the number of rows equals the number of columns,  $A$  is square and is referred to as a *square matrix* and, provided there is no ambiguity, is described as being of order  $r$ . In this case the elements  $a_{11}, a_{22}, a_{33}, \dots, a_{rr}$  are referred to as the *diagonal elements* or *diagonal* of the matrix. For example,

$$A = \begin{bmatrix} 1 & 7 & 0 & 7 \\ 5 & 2 & 9 & 1 \\ 0 & 3 & 4 & 6 \\ 8 & 0 & 5 & 7 \end{bmatrix}$$

is square, of order 4 and its diagonal elements are 1, 2, 4 and 7. The elements of a square matrix that lie in a line parallel to and just below the diagonal are sometimes referred to as being on the *subdiagonal elements*; in the example they are 5, 3, 5. Elements of a square matrix other than the diagonal elements are called *off-diagonal* or *nondiagonal* elements.

A square matrix having zero for all its nondiagonal elements is described as a *diagonal matrix*; for example,

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -17 & 0 \\ 0 & 0 & 99 \end{bmatrix}$$

is a diagonal matrix. When  $a_1, a_2, \dots, a_n$  (some of which may be zero) are the diagonal elements of a diagonal matrix, useful notations for that matrix

are

$$\begin{aligned} D\{a_1, a_2, \dots, a_n\} &= \text{diag}\{a_1, a_2, \dots, a_n\} \\ &= D\{a_i\} = \text{diag}\{a_i\} \quad \text{for } i = 1, 2, \dots, n. \end{aligned}$$

A square matrix with all elements above (or below) the diagonal being zero is called a *triangular matrix*. For example,

$$\mathbf{B} = \begin{bmatrix} 1 & 5 & 13 \\ 0 & -2 & 9 \\ 0 & 0 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 & 0 \\ 8 & 3 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

are triangular matrices.  $\mathbf{B}$  is an *upper triangular matrix* and  $\mathbf{C}$  is a *lower triangular matrix*.

**Illustration.** Suppose an experienced taxicab driver has ascertained that when in Town 1 there is a probability of .2 that the next fare will be within Town 1 and a probability of .8 that it will be to Town 2. But when in Town 2 the probabilities are .4 of going to Town 1 and .6 of staying in Town 2. These probabilities can be assembled in a matrix

$$\mathbf{P} = \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} = \{p_{ij}\} \quad \text{for } i, j = 1, 2. \quad (1)$$

Then  $p_{ij}$  is the probability when in Town  $i$  of the next fare being to Town  $j$ . The matrix in (1) is an example of a transition probability matrix. It is an array of probabilities of making a transition from what is called state  $i$  to state  $j$ . (The states in the example are the two towns.) Thus  $p_{ij}$  is a transition probability, and hence the name: *transition probability matrix*.

In (1), the sum of the elements in each row (the *row sum*) is 1.00. This is so because whenever the taxi is in Town  $i$  it must, with its next fare, either stay in that town or go to the other one. This is a feature of all transition probability matrices. In general, for

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mm} \end{bmatrix} \\ &= \{p_{ij}\} \quad \text{for } i, j = 1, 2, \dots, m, \end{aligned} \quad (2)$$

corresponding to  $m$  different states, the sum of the probabilities of going from state  $i$  to any one of the  $m$  states (including staying in state  $i$ ) must be 1. Hence

$$p_{i1} + p_{i2} + \cdots + p_{ij} + \cdots + p_{im} = 1,$$

i.e.,

$$\sum_{j=1}^m p_{ij} = 1 \quad \text{for all } i = 1, 2, \dots, m. \quad (3)$$

Arranging the  $p_{ij}$ 's as elements of the matrix  $\mathbf{P}$  in (2) allows us, as is shown in Section 3.2, to write the equations in (3) as a product of  $\mathbf{P}$  and another (particular kind of) matrix.

Situations like the taxicab example form a class of probability models known as *Markov chains*. They arise in a variety of ways; for example, in considering the probabilities of changes in the prime interest rate from week to week, or the probabilities of certain types of mating in genetics (see the illustration at the end of Section 2.8b), or in studying the yearly mobility of labor where  $p_{ij}$  can be the probability of moving during a calendar year from category  $i$  in the labor force to category  $j$ . Assembling probabilities of this nature into a matrix then enables matrix algebra to be used as a tool for answering questions of interest; for example, the probabilities of moving from category  $i$  to category  $j$  in two years are given by  $\mathbf{P}^2$ ; in three years, by  $\mathbf{P}^3$ ; and so on (see Section 2.8m).

## 7. VECTORS AND SCALARS

A matrix consisting of only a single column is called a *column vector*; for example,

$$\mathbf{x} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

is a column vector of order 4. Likewise a matrix that is just a single row is a *row vector*:

$$\mathbf{y}' = [4 \quad 6 \quad -7]$$

is a row vector of order 3. (A more specific reason for the prime on  $\mathbf{y}$  is given in Section 2.1.)

A single number such as 2, 6.4 or  $-4$  is called a *scalar*; the elements of a matrix are (usually) scalars although, as discussed in Section 2.2 and in Chapter 10, a matrix can be expressed as a matrix of smaller matrices. Sometimes it is convenient to think of a scalar as a matrix of order  $1 \times 1$  (see Section 2.3).

### 8. GENERAL NOTATION

A well-recognized notation is that of denoting matrices by uppercase letters and their elements by the lowercase counterparts with appropriate subscripts. Vectors are denoted by lowercase letters, often from the end of the alphabet, using the prime superscript to distinguish a row vector from a column vector. Thus  $x$  is a column vector and  $x'$  is a row vector. The lowercase Greek lambda,  $\lambda$ , is often used for a scalar.

Throughout this book the notation for displaying a matrix is

$$A = \begin{bmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{bmatrix},$$

enclosing the array of elements in square brackets. Among the variety of forms that can be found in the literature are

$$\begin{pmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{pmatrix}, \quad \left\{ \begin{matrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{matrix} \right\} \quad \text{and} \quad \left\| \begin{matrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{matrix} \right\|.$$

Single vertical lines are seldom used, since they are usually reserved for determinants (see Chapter 4).

Another useful notation that has already been described is

$$A = \{a_{ij}\} \text{ for } i = 1, 2, \dots, r \text{ and } j = 1, 2, \dots, c.$$

The curly brackets indicate that  $a_{ij}$  is a typical term of the matrix  $A$  for all pairs of values of  $i$  and  $j$  from unity up to the limits shown, in this case  $r$  and  $c$ ; i.e.,  $A$  is a matrix of  $r$  rows and  $c$  columns. This is by no means a universal notation and several variants of it can be found in the literature. Furthermore, there is nothing sacrosanct about the repeated use in this book of the letter  $A$  for a matrix. Any letter may be used.

### 9. ILLUSTRATIVE EXAMPLES

A text on matrix algebra designed for pure mathematicians would deal with many topics that do not appear in this book because they have little

direct connection with practical applied problems. The text for mathematicians might also have few numerical illustrations. This will not be the case here, however, and indeed, simple illustrations have already been used to introduce the concept of a matrix. The advantage of such illustrations is that for many people arithmetic involving actual numbers is easier to follow than long expositions in algebra. For this reason, mathematical results are frequently illustrated with numerical examples.

The mathematical discussion in the ensuing chapters is relatively informal. Proofs and demonstrations of general results are not omitted, however, for to do so would result in a lifeless "cookbook" presentation. Once a general mathematical result has been established it is usually illustrated by a numerical example, with additional examples available as exercises at the end of most chapters. Only a small proportion of these is algebraic; many are numerical.

Numerical illustrations come in two styles. One style draws on real-life backgrounds like the examples of sterile cultures and taxicab fares already used. These two illustrations are solely for demonstrating an idea, but later in the book when matrix tools of sufficient complexity have been developed, illustrations from real life show how matrix algebra can help solve genuine problems. Before reaching that stage, however, there are intermediate mathematical steps for which few real-world problems can be used as illustrations. For this part of the work we rely heavily on the second style of illustration, which is nothing more than numbers "pulled out of a hat" for the sole purpose of illustrating a mathematical procedure. Such numbers have no use other than the very good one of assisting the reader to understand the procedures being demonstrated. Numerical illustrations created in this manner are labeled "Example"; those with a background in real life are labeled "Illustration."

### EXERCISES

1. For  $a_{11} = 17$   $a_{12} = 31$   $a_{13} = 26$   $a_{14} = 11$   
 $a_{21} = 19$   $a_{22} = 27$   $a_{23} = 16$   $a_{24} = 14$   
 $a_{31} = 21$   $a_{32} = 23$   $a_{33} = 15$   $a_{34} = 16$

show that

- (a)  $a_{1.} = 85$ ,  $a_{.2} = 76$ , and  $a_{3.} = 75$ ;  
 (b)  $a_{.1} = 57$ ,  $a_{.2} = 81$ ,  $a_{.3} = 57$ , and  $a_{.4} = 41$ ;  
 (c)  $a_{..} = 236$ ;  
 (d)  $\sum_{i=1}^3 a_{ii} = 59$ ;

- (e)  $\sum_{\substack{j=1 \\ j \neq 2}}^4 a_{ij} = 54, 49$  and  $52$  for  $i = 1, 2$  and  $3$ , respectively;
- (f)  $\sum_{\substack{i=1 \\ i \neq 2}}^3 \sum_{\substack{j=1 \\ j \neq 3}}^4 a_{ij} = 119$ .

2. For the elements of the matrix

$$A = \begin{bmatrix} -1 & 17 & 9 & -2 & 3 \\ 3 & 13 & 10 & 2 & 6 \\ 11 & -9 & 0 & -3 & 2 \\ -6 & -8 & 1 & 4 & 5 \end{bmatrix} = \{a_{ij}\}$$

for  $i = 1, 2, \dots, 4$  and  $j = 1, 2, \dots, 5$  show that

- (a)  $a_{1.} = 26$ ; (b)  $a_{.3} = 20$ ;
- (c)  $\sum_{i=1}^3 a_{i2} = 21$ ; (d)  $\sum_{\substack{i=1 \\ i \neq 2}}^4 a_{i5} = 10$ ;
- (e)  $\sum_{i=3}^4 a_i^2 = \sum_{i=3}^4 (a_{i.})^2 = 17$ ; (f)  $a_{..} = 57$ ;
- (g)  $\sum_{i=1}^4 \frac{\sum_{j=1}^5 a_{ij}^2}{a_i}$  (h)  $\sum_{\substack{i=1 \\ i \neq 2}}^4 \sum_{\substack{j=2 \\ j \neq 4}}^5 a_{ij} = 20$ ;
- (i)  $\prod_{j=1}^3 a_{4j} = \prod_{i=1}^4 a_{i4}$ ; (j)  $\sum_{i=1}^4 a_{i1} a_{i4} = -49$ ;
- (k)  $\sum_{i=1}^4 \sum_{j=1}^5 a_{..} = 1140$ ; (l)  $\prod_{i=1}^3 a_{i4} = 12$ ;
- (m)  $\prod_{j=1}^5 a_{2j} a_{3j} = 0$ ; (n)  $\prod_{i=1}^4 2^{a_{i.}} = 2$ ;
- (o)  $\sum_{i=1}^4 \sum_{j=1}^5 (-1)^j a_{ij} = -29$ ; (p) for  $i = 2, \sum_{\substack{j=1 \\ j \neq 3}}^5 (a_{ij} - a_{i+2,j})^2 = 527$ ;
- (q)  $\prod_{j=1}^5 2^{(-1)^j a_{2j}} = 0.0625$ .

3. From  $A$  of Exercise 2, write down the following matrices:

$$B = \{a_{i+1, j+2}\} \quad \text{for } i = 1, 2, 3 \text{ and } j = 1, 2, 3;$$

$$C = \{a_{2i, 2j-1}\} \quad \text{for } i = 1, 2 \text{ and } j = 1, 2, 3;$$

and

$$D = \{a_{i,j} + i - j\} \quad \text{for } i = 2, 3 \text{ and } j = 1, \dots, 4.$$

4. Prove the following identities, and demonstrate their validity using  $A$  of Exercise 2.

- (a)  $\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2$ .
- (b)  $\sum_{i=1}^m \left( \sum_{j=1}^n a_{ij} \right)^2 = \sum_{i=1}^m a_i^2$ .
- (c)  $\sum_{j=1}^n \sum_{k=1}^n a_{ij} a_{hk} = a_i \cdot a_h$ , for  $i \neq h$ .
- (d)  $\sum_{i=1}^m \sum_{j=1}^n 4a_{ij} = 4a_{..}$ .
- (e)  $\left( \sum_{j=1}^n a_{ij} \right)^2 = \sum_{j=1}^n a_{ij}^2 + 2 \sum_{j=1}^{n-1} \sum_{p=j+1}^n a_{ij} a_{ip}$   
 $= \sum_{j=1}^n a_{ij}^2 + 2 \sum_{j=1}^{n-1} \sum_{p>j}^n a_{ij} a_{ip}$   
 $= \sum_{j=1}^n (a_{ij}^2 + \sum_{\substack{p=1 \\ p \neq j}}^n a_{ij} a_{ip})$
- (f)  $a^2 = \sum_{i=1}^n a_i^2 + 2 \sum_{i=1}^{n-1} \sum_{h=i+1}^n a_i \cdot a_h$   
 $= \sum_{i=1}^n a_i^2 + 2 \sum_{i=1}^{n-1} \sum_{h>i}^n a_i \cdot a_h$   
 $= \sum_{i=1}^n a_i^2 + \sum_{i=1}^n \sum_{\substack{h=1 \\ h \neq i}}^n a_i \cdot a_h$
- (g)  $\sum_{\substack{i=1 \\ i \neq p}}^m \sum_{\substack{j=1 \\ j \neq q}}^n a_{ij} = a_{..} - a_{.p} - a_{.q} + a_{pq}$ .

$$(h) \sum_{i=1}^m \sum_{j=1}^n (a_{ij} - 1) = a_{..} - mn.$$

$$(i) \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n a_{ij} a_{ik} = a_i^2 - \sum_{j=1}^n a_{ij}^2.$$

5. Show that

$$(a) \sum_{i=3}^5 3^i = 351;$$

$$(b) \sum_{k=2}^7 2^k = 252;$$

$$(c) \sum_{r=1}^5 r = 15;$$

$$(d) \sum_{s=1}^5 s(s+1) = 106;$$

$$(e) \prod_{i=1}^4 2^i = 1024;$$

$$(f) \sum_{i=1}^3 \prod_{j=1}^2 i^{2j-1} = 98.$$

6. Write down the following matrices:

$$\mathbf{A} = \{a_{ij}\} \text{ for } a_{ij} = i + j \text{ for } i = 1, 2, 3 \text{ and } j = 1, 2.$$

$$\mathbf{B} = \{b_{kt}\} \text{ for } b_{kt} = k^{t-1} \text{ for } k = 1, \dots, 4 \text{ and } t = 1, \dots, 3.$$

$$\mathbf{C} = \{c_{rs}\} \text{ for } c_{rs} = 3r + 2(s-1) \text{ for } r = 1, \dots, 4$$

and  $s = 1, \dots, 5.$

7. Write down the matrices

$$\mathbf{D}_1 = \mathbf{D}\{1, 2, 3, 4\}, \quad \mathbf{D}_2 = \mathbf{D}\{3^{i-2}\} \quad \text{for } i = 1, \dots, 4$$

and

$$\mathbf{D}_3 = \mathbf{D}\{i + 3^{i-2}\} \quad \text{for } i = 1, \dots, 4.$$

8. When  $x_i = \sum_{j=1}^n \lambda_{ij} v_j$  where  $\sum_{j=1}^n \lambda_{ij} = 1$  for all  $i = 1, 2, \dots, m$ , show that the coefficients of the  $y_j$ 's in  $\sum_{i=1}^m \mu_i x_i$  sum to unity when  $\sum_{i=1}^m \mu_i = 1.$

9. Show that

$$\begin{aligned} & \sum_{i=1}^a \frac{1}{(1+n_i t)^2} \sum_{i=1}^a \frac{n_i}{1+n_i t} - \sum_{i=1}^a \frac{n_i}{(1+n_i t)^2} \sum_{i=1}^a \frac{1}{1+n_i t} \\ &= t \left[ \sum_{i=1}^a \frac{1}{(1+n_i t)^2} \sum_{i=1}^a \frac{n_i^2}{(1+n_i t)^2} - \left\{ \sum_{i=1}^a \frac{n_i}{(1+n_i t)^2} \right\}^2 \right] \end{aligned}$$

10. Prove that

$$1 + \sum_{j=1}^b \frac{1}{n_{1j}} \sum_{i=2}^a \left( \sum_{j=1}^b \frac{1}{n_{ij}} \right)^{-1} = \sum_{j=1}^b \frac{1}{n_{1j}} \sum_{i=1}^a \left( \sum_{j=1}^b \frac{1}{n_{ij}} \right)^{-1}.$$

11. Suppose that for families with the husband aged 20 through 30 family size (number of children) is recorded each year, as of September 30, for 10 years. From these data, the frequencies  $f_{ij}$  are calculated of annual changes in family size for  $i, j = 0, 1, 2, 3$  and 4, where 4 represents 4 or more children. Ignoring deaths, explain why  $\mathbf{F} = \{f_{ij}\}$  is a transition probability matrix, why it is triangular, and why  $f_{44} = 1$ . Comment on other features of the values of  $f_{ij}$  that you think might be apparent.

## CHAPTER 2

# BASIC OPERATIONS

This chapter describes simple operations on matrices arising from their being rectangular arrays of numbers, and also the arithmetic of matrix addition, subtraction and multiplication. Division is sufficiently more complicated for matrices than for scalars as to require a chapter to itself, Chapter 5.

### 1. THE TRANSPOSE OF A MATRIX

In first defining a matrix (Section 1.2) we considered the percentage of sterile cultures of a laboratory organism in four different generations of three different populations. The percentages were arrayed as a  $4 \times 3$  matrix

$$\mathbf{A} = \begin{bmatrix} 18 & 17 & 11 \\ 19 & 13 & 6 \\ 6 & 14 & 9 \\ 9 & 11 & 4 \end{bmatrix}$$

where rows represent percentages for the generations and columns represent those for the populations. This choice of rows is arbitrary as far as presentation of the percentages is concerned, and we could well interchange rows and columns with no loss of information, having rows (rather than columns) for populations and columns for generations. The array of the percentages would be

$$\mathbf{B} = \begin{bmatrix} 18 & 19 & 6 & 9 \\ 17 & 13 & 14 & 11 \\ 11 & 6 & 9 & 4 \end{bmatrix}$$

Although the elements of  $\mathbf{B}$  are the same as those of  $\mathbf{A}$ , the matrices are clearly different; for example, they are not of the same order.  $\mathbf{A}$  has order  $4 \times 3$  and  $\mathbf{B}$  is  $3 \times 4$ . The matrices are nevertheless related to each other, through the rows of one being the columns of the other. Whenever matrices are related in this fashion each is said to be the *transpose* of the other; for example,  $\mathbf{B}$  is said to be the transpose of  $\mathbf{A}$ , and  $\mathbf{A}$  is the transpose of  $\mathbf{B}$ .

A formal description is that the transpose of a matrix  $\mathbf{A}$  is the matrix whose columns are the rows of  $\mathbf{A}$ , with order retained, from first to last. The transpose is written as  $\mathbf{A}'$ , although the notations  $\mathbf{A}'$  and  $\mathbf{A}^T$  are also seen in the literature. We will use  $\mathbf{A}'$ .

An obvious consequence of the transpose operation is that the rows of  $\mathbf{A}'$  are the same as the columns of  $\mathbf{A}$ , and if  $\mathbf{A}$  is  $r \times c$ , the order of  $\mathbf{A}'$  is  $c \times r$ . If  $a_{ij}$  is the term in the  $i$ th row and the  $j$ th column of  $\mathbf{A}$ , it is also the term in the  $j$ th row and  $i$ th column of  $\mathbf{A}'$ . Therefore, on defining  $a'_{ji}$  as the element in the  $j$ th row and  $i$ th column of  $\mathbf{A}'$ , we have  $a'_{ji} = a_{ij}$ . Equivalently, on interchanging  $i$  and  $j$ , we have

$$\mathbf{A}' = \{a'_{ij}\} \quad \text{for } i = 1, \dots, c \quad \text{and } j = 1, \dots, r, \quad \text{and } a'_{ij} = a_{ji}.$$

**Example.** When

$$\mathbf{A} = \begin{bmatrix} 18 & 17 & 11 \\ 19 & 13 & 6 \\ 6 & 14 & 9 \\ 9 & 11 & 4 \end{bmatrix}, \quad \text{then } \mathbf{A}' = \begin{bmatrix} 18 & 19 & 6 & 9 \\ 17 & 13 & 14 & 11 \\ 11 & 6 & 9 & 4 \end{bmatrix}$$

and  $a_{32} = 14 = a'_{23}$ .

A minor difficulty in notation arises here.  $\mathbf{A}_{r \times c}$  has  $r$  rows and  $c$  columns; but combining the notation  $\mathbf{A}'$  for the transpose of  $\mathbf{A}$  with the  $r \times c$  subscript notation for the order of  $\mathbf{A}$ , to get  $\mathbf{A}'_{r \times c}$ , is ambiguous.  $\mathbf{A}'_{r \times c}$  could mean either that  $\mathbf{A}$  of order  $r \times c$  has been transposed or that the transpose of  $\mathbf{A}$  has order  $r \times c$  (and so  $\mathbf{A}$  has order  $c \times r$ ). For clarity, one of the equivalent forms  $(\mathbf{A}_{r \times c})'$  or  $(\mathbf{A}')_{c \times r}$  must be used whenever it is necessary to have subscript notation for the order of a transposed matrix. Fortunately, the need for this clumsy notation seldom arises. The simple  $\mathbf{A}'$  suffices in most contexts.

Two important consequences of the transpose operation are worth noting.

#### a. A reflexive operation

The transpose operation is reflexive: the transpose of a transposed matrix is the matrix itself, i.e.,  $(\mathbf{A}')' = \mathbf{A}$ . This is so because transposing  $\mathbf{A}'$  yields a



matrix whose rows are the columns of  $A'$ , and these are the rows of  $A$ . Hence  $(A')' = A$ . More formally,

$$(A')' = \{a'_{ij}\}' = \{a_{ji}\}' = \{a'_{ji}\} = \{a_{ij}\} = A.$$

### b. Vectors

The transpose of a row vector is a column vector and vice versa. For example, the transpose of

$$x = \begin{bmatrix} 1 \\ 6 \\ 4 \end{bmatrix} \quad \text{is} \quad x' = [1 \quad 6 \quad 4].$$

This explains the notation (see Section 1.7) of denoting a row vector by a superscript prime. It indicates that a row vector is the transpose of the column vector of the same elements and distinguishes it from that column vector.

## 2. PARTITIONED MATRICES

### a. Example

Consider the matrix

$$B = \begin{bmatrix} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{bmatrix}.$$

Suppose we draw dashed lines between certain rows and columns as in

$$B = \begin{bmatrix} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{bmatrix}. \quad (1)$$

Each of the arrays of numbers in the four sections of  $B$  engendered by the dashed lines is a matrix:

$$\begin{aligned} B_{11} &= \begin{bmatrix} 1 & 6 & 8 & 9 \\ 2 & 4 & 1 & 6 \\ 4 & 3 & 6 & 1 \end{bmatrix}, & B_{12} &= \begin{bmatrix} 3 & 8 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \\ B_{21} &= \begin{bmatrix} 9 & 1 & 4 & 6 \\ 6 & 8 & 1 & 4 \end{bmatrix}, & B_{22} &= \begin{bmatrix} 8 & 7 \\ 3 & 2 \end{bmatrix}. \end{aligned} \quad (2)$$

Using the matrices in (2), the matrix  $B$  of (1) can now be written as a matrix of matrices:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}. \quad (3)$$

This specification of  $B$  is called a *partitioning* of  $B$ , and the matrices  $B_{11}$ ,  $B_{12}$ ,  $B_{21}$ ,  $B_{22}$  are said to be *submatrices* of  $B$ ; further,  $B$  of (3) is called a *partitioned matrix*.

Note that  $B_{11}$  and  $B_{21}$  have the same number of columns, as do  $B_{12}$  and  $B_{22}$ . Likewise  $B_{11}$  and  $B_{12}$  have the same number of rows, as do  $B_{21}$  and  $B_{22}$ . This is the usual method of partitioning, as expressed in the general case for an  $r \times c$  matrix:

$$A_{r \times c} = \begin{bmatrix} K_{p \times q} & L_{p \times (c-q)} \\ M_{(r-p) \times q} & N_{(r-p) \times (c-q)} \end{bmatrix}$$

where  $K$ ,  $L$ ,  $M$  and  $N$  are the submatrices with their orders shown as subscripts.

Partitioning is not restricted to dividing a matrix into just four submatrices; it can be divided into numerous rows and columns of matrices. Thus if

$$B_{01} = \begin{bmatrix} 1 & 6 & 8 & 9 \\ 2 & 4 & 1 & 6 \end{bmatrix}, \quad B_{02} = \begin{bmatrix} 3 & 8 \\ 1 & 1 \end{bmatrix},$$

$$B_{03} = [4 \quad 3 \quad 6 \quad 1] \quad \text{and} \quad B_{04} = [2 \quad 1]$$

with  $B_{21}$  and  $B_{22}$  as in (2), then  $B$  can be written in partitioned form as

$$B = \begin{bmatrix} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{bmatrix} = \begin{bmatrix} B_{01} & B_{02} \\ B_{03} & B_{04} \\ B_{21} & B_{22} \end{bmatrix}. \quad (4)$$

It goes without saying that each such line must always go the full length (or breadth) of the matrix. Partitioning in any staggered manner such as

$$\begin{bmatrix} | & | & | & | & | & | \\ | & | & | & | & | & | \\ | & | & | & | & | & | \\ | & | & | & | & | & | \\ | & | & | & | & | & | \end{bmatrix}$$

is not allowed.

**b. General specification**

The example illustrates the simplicity of the operation of partitioning a matrix and, as is evident from (1) and (4), there is no single way in which any matrix can be partitioned. In general, a matrix  $A$  of order  $p \times q$  can be partitioned into  $r$  rows and  $c$  columns of submatrices as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1c} \\ A_{21} & A_{22} & \cdots & A_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rc} \end{bmatrix}$$

where  $A_{ij}$  is the submatrix in the  $i$ th row and  $j$ th column of submatrices. If the  $i$ th row of submatrices has  $p_i$  rows of elements and the  $j$ th column of submatrices has  $q_j$  columns, then  $A_{ij}$  has order  $p_i \times q_j$ , where

$$\sum_{i=1}^r p_i = p \quad \text{and} \quad \sum_{j=1}^c q_j = q.$$

**c. Transposing a partitioned matrix**

The transpose of a partitioned matrix is the transposed matrix of transposed submatrices:

$$[X \quad Y]' = \begin{bmatrix} X' \\ Y' \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} A & B & C \\ D & E & F \end{bmatrix}' = \begin{bmatrix} A' & D' \\ B' & E' \\ C' & F' \end{bmatrix}.$$

The reader should use the example in (1) to verify these results and to be assured that in general the transpose of  $[X \quad Y]$  is neither  $\begin{bmatrix} X \\ Y \end{bmatrix}$  nor  $[X' \quad Y']$ .

**Example**

$$\begin{bmatrix} 2 & 8 & 9 \\ 3 & 7 & 4 \end{bmatrix}' = \begin{bmatrix} 2 & 3 \\ 8 & 7 \\ 9 & 4 \end{bmatrix}.$$

**d. Partitioning into vectors**

Suppose  $a_j$  is the  $j$ th column of  $A_{r \times c}$ . Then

$$A = [a_1 \quad a_2 \quad \cdots \quad a_j \quad \cdots \quad a_c] \quad (5)$$

[2.3]

is  $A$  partitioned into its  $c$  columns. Similarly,

$$A = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \vdots \\ \alpha'_i \\ \vdots \\ \alpha'_r \end{bmatrix} \quad (6)$$

is  $A$  partitioned into its  $r$  rows  $\alpha'_i$  for  $i = 1, \dots, r$ .

**3. THE TRACE OF A MATRIX**

The sum of the diagonal elements of a square matrix is called the *trace* of the matrix, written  $\text{tr}(A)$ ; i.e., for  $A = \{a_{ij}\}$  for  $i, j = 1, \dots, n$

$$\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^n a_{ii}.$$

**Example**

$$\text{tr} \begin{bmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{bmatrix} = 1 + 3 - 8 = -4.$$

When  $A$  is not square, the trace is not defined; i.e., it does not exist.

The trace of a transposed matrix is the same as the trace of the matrix itself:

$$\text{tr}(A') = \text{tr}(A).$$

For example,

$$\text{tr} \begin{bmatrix} 1 & 8 & 4 \\ 7 & 3 & -2 \\ 6 & 9 & -8 \end{bmatrix} = \text{tr} \begin{bmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{bmatrix} = 1 + 3 - 8 = -4.$$

Also, by treating a scalar as a  $1 \times 1$  matrix we have

$$\text{tr}(\text{scalar}) = \text{scalar}; \quad \text{e.g.,} \quad \text{tr}(13) = 13.$$

## 4. ADDITION

We introduce the operation of addition by means of an illustration.

**Illustration.** The numbers of lunches served one Saturday at a country club are shown in Table 2.1.

TABLE 2.1 NUMBER OF LUNCHES

	Beef	Fish	Omelet
Member	98	24	42
Guest	39	15	22

Let us write the data of the table as the  $2 \times 3$  matrix:

$$\mathbf{A} = \begin{bmatrix} 98 & 24 & 42 \\ 39 & 15 & 22 \end{bmatrix}.$$

Then, with the same frame of reference, the data for the Sunday lunches might be

$$\mathbf{B} = \begin{bmatrix} 55 & 19 & 44 \\ 43 & 53 & 38 \end{bmatrix}.$$

Hence over the weekend the total number of members served a beef lunch is the sum of the elements in the first row and first column of each matrix,  $98 + 55 = 153$ ; and the total number of guests served a fish lunch is  $15 + 53 = 68$ . In this way the matrix of all such sums is

$$\begin{bmatrix} 98 + 55 & 24 + 19 & 42 + 44 \\ 39 + 43 & 15 + 53 & 22 + 38 \end{bmatrix} = \begin{bmatrix} 153 & 43 & 86 \\ 82 & 68 & 60 \end{bmatrix},$$

which represents the numbers of lunches served over the weekend. This is the matrix sum  $\mathbf{A}$  plus  $\mathbf{B}$ ; it is the matrix formed by adding the matrices  $\mathbf{A}$  and  $\mathbf{B}$  element by element. Hence if we write  $\mathbf{A} = \{a_{ij}\}$  and  $\mathbf{B} = \{b_{ij}\}$  for  $i = 1, 2, \dots, r$  and  $j = 1, 2, \dots, c$  the matrix representing the sum of  $\mathbf{A}$  and  $\mathbf{B}$  is

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\} \quad \text{for } i = 1, 2, \dots, r \quad \text{and } j = 1, 2, \dots, c;$$

i.e., the sum of two matrices is the matrix of sums, element by element.

[2.5]

It is evident from this definition that matrix addition can take place only when the matrices involved are of the same order; i.e., two matrices can be added only if they have the same number of rows and the same number of columns, in which case they are said to be *comformable for addition*.

Having defined addition, we can now consider the transpose of a sum and the trace of a sum:

- (i) The transpose of a sum of matrices is the sum of the transposed matrices, i.e.,

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'.$$

- (ii) The trace of a sum of matrices is the sum of the traces, i.e.,

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

provided  $\mathbf{A}$  and  $\mathbf{B}$  are square, of the same order.

It is left to the reader to create examples illustrating these results. Formal proofs are as follows.

- (i) If

$$\mathbf{A} + \mathbf{B} = \mathbf{C} = \{c_{ij}\} = \{a_{ij} + b_{ij}\},$$

then

$$(\mathbf{A} + \mathbf{B})' = \mathbf{C}' = \{c'_{ij}\} = \{c_{ji}\} = \{a_{ji} + b_{ji}\} = \{a_{ji}\} + \{b_{ji}\} = \mathbf{A}' + \mathbf{B}'.$$

$$(ii) \quad \text{tr}(\mathbf{A} + \mathbf{B}) = \sum_{i=1}^n (a_{ii} + b_{ii}) = \sum_{i=1}^n a_{ii} + \sum_{i=1}^n b_{ii} = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}).$$

## 5. SCALAR MULTIPLICATION

We have just described matrix addition. A simple use of it shows that

$$\mathbf{A} + \mathbf{A} = \{a_{ij}\} + \{a_{ij}\} = \{2a_{ij}\} = 2\mathbf{A}.$$

Extending this to the case where  $\lambda$  is a positive integer, we have

$$\lambda\mathbf{A} = \mathbf{A} + \mathbf{A} + \mathbf{A} + \dots + \mathbf{A},$$

there being  $\lambda$   $\mathbf{A}$ 's in the sum on the right. Carrying out these matrix

additions gives

$$\lambda \mathbf{A} = \{\lambda a_{ij}\} \text{ for } i = 1, 2, \dots, r \text{ and } j = 1, 2, \dots, c.$$

This result, extended to  $\lambda$  being any scalar, is the definition of *scalar multiplication of a matrix*. Thus the matrix  $\mathbf{A}$  multiplied by the scalar  $\lambda$  is the matrix  $\mathbf{A}$  with every element multiplied by  $\lambda$ . For example,

$$3 \begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 6 & -21 \\ 9 & 15 \end{bmatrix} \quad \text{and} \quad \lambda \begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 2\lambda & -7\lambda \\ 3\lambda & 5\lambda \end{bmatrix}.$$

Although  $\lambda \mathbf{A}$  is the customary notation, rather than  $\mathbf{A}\lambda$ , the latter is more convenient when  $\lambda$  is a fraction; e.g.,  $\mathbf{A}/17$  rather than  $(1/17)\mathbf{A}$  or  $\mathbf{A}(1/17)$ . Thus

$$\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} / 17 = \begin{bmatrix} 2/17 & -7/17 \\ 3/17 & 5/17 \end{bmatrix}.$$

## 6. SUBTRACTION

**Illustration.** An international corporation's gross income for 1981 and 1982 in four countries is shown in Table 2.2.

TABLE 2.2. GROSS INCOME (in \$1,000,000 units)

Year	United States	Canada	Australia	United Kingdom
1981	27	15	18	21
1982	32	14	21	30

Represent the income figures by the matrix

$$\mathbf{A} = \begin{bmatrix} 27 & 15 & 18 & 21 \\ 32 & 14 & 21 & 30 \end{bmatrix}$$

and suppose the corresponding matrix of expenses is

$$\mathbf{B} = \begin{bmatrix} 19 & 9 & 11 & 17 \\ 22 & 10 & 13 & 24 \end{bmatrix}.$$

[2.7]

Gross profit in 1981 in the United States is  $27 - 19 = 8$ , and that for Canada is  $15 - 9 = 6$ . Hence the matrix of all such differences

$$\begin{bmatrix} 27 - 19 & 15 - 9 & 18 - 11 & 21 - 17 \\ 32 - 22 & 14 - 10 & 21 - 13 & 30 - 24 \end{bmatrix} = \begin{bmatrix} 8 & 6 & 7 & 4 \\ 10 & 4 & 8 & 6 \end{bmatrix}$$

represents profits in both years for all four countries. This illustrates the matrix operation of subtraction. Thus if  $\mathbf{A} = \{a_{ij}\}$  and  $\mathbf{B} = \{b_{ij}\}$  for  $i = 1, 2, \dots, r$  and  $j = 1, 2, \dots, c$ , subtraction in matrix algebra is defined as

$$\mathbf{A} - \mathbf{B} = \{a_{ij} - b_{ij}\} \text{ for } i = 1, 2, \dots, r \text{ and } j = 1, 2, \dots, c;$$

i.e., the difference between two matrices is the matrix of differences, element by element. A formal proof of this is a direct outcome of the rules of addition and scalar multiplication. Since  $-\mathbf{B} = (-1)\mathbf{B}$ ,

$$\mathbf{A} - \mathbf{B} = \{a_{ij}\} + (-1)\{b_{ij}\} = \{a_{ij}\} + \{-b_{ij}\} = \{a_{ij} - b_{ij}\}.$$

### Example

$$\begin{bmatrix} 3 & 6 \\ 8 & 2 \\ 4 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 0 & -3 \\ 2 & -5 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 8 & 5 \\ 2 & 6 \end{bmatrix}.$$

As with addition, only matrices that are of the same order can be subtracted from one another. Thus matrices that are conformable for addition are also *conformable for subtraction*, and vice versa.

## 7. EQUALITY AND THE NULL MATRIX

Two matrices are said to be equal when they are identical element by element. Thus  $\mathbf{A} = \mathbf{B}$  when  $\{a_{ij}\} = \{b_{ij}\}$ , meaning that  $a_{ij} = b_{ij}$  for  $i = 1, 2, \dots, r$  and  $j = 1, 2, \dots, c$ . If

$$\mathbf{A} = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 6 & -4 \\ 2 & 0 & 1 \end{bmatrix},$$

$\mathbf{A}$  is equal to  $\mathbf{B}$  but not equal to  $\mathbf{C}$ . It is also apparent that equality of two matrices has no meaning unless they are of the same order.

Combining the ideas of subtraction and equality leads to the definition of zero in matrix algebra. For when  $\mathbf{A} = \mathbf{B}$ , then  $a_{ij} = b_{ij}$  for  $i = 1, 2, \dots, r$

and  $j = 1, 2, \dots, c$ , and so

$$\mathbf{A} - \mathbf{B} = \{a_{ij} - b_{ij}\} = \{0\} = \mathbf{0}.$$

The matrix on the right is a matrix of zeros, i.e., every element is zero. Such a matrix is called a *null matrix*; it is the zero of matrix algebra and is sometimes referred to as a *zero matrix*. It is, of course, not a unique zero because corresponding to every matrix there is a null matrix of the same order. For example, null matrices of order  $2 \times 4$  and  $3 \times 3$  are, respectively,

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

## 8. MULTIPLICATION

The method of multiplying matrices is developed by first considering the product of two vectors and then the product of a matrix and a vector. Each of these products is introduced by means of an illustration and is then given a formal definition.

### a. The inner product of two vectors

**Illustration.** Consider buying supplies of experimental rats, mice and rabbits for laboratory courses in the chemistry, biochemistry, nutrition and physiology departments of a university. Suppose the price per animal of rats, mice and rabbits in the home town is \$3, \$1 and \$10, and that the chemistry department needs 50, 100 and 30 animals, respectively. The total cost to the chemistry department is, very simply,

$$3(50) + 1(100) + 10(30) = 550 \quad (7)$$

dollars. Suppose the prices are written as a row vector  $\mathbf{a}'$  and the numbers of animals needed as a column vector  $\mathbf{x}$ :

$$\mathbf{a}' = [3 \quad 1 \quad 10] \text{ and } \mathbf{x} = \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix}.$$

Then the total cost of the animals needed, 550, is the sum of products of the elements of  $\mathbf{a}'$  each multiplied by the corresponding element of  $\mathbf{x}$ . This is the

definition of the product  $\mathbf{a}'\mathbf{x}$ . It is written as

$$\mathbf{a}'\mathbf{x} = [3 \quad 1 \quad 10] \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = 3(50) + 1(100) + 10(30) = 550. \quad (8)$$

This example illustrates the general procedure for obtaining  $\mathbf{a}'\mathbf{x}$ : multiply each element of the row vector  $\mathbf{a}'$  by the corresponding element of the column vector  $\mathbf{x}$  and add the products. The sum is  $\mathbf{a}'\mathbf{x}$ . Thus if

$$\mathbf{a}' = [a_1 \quad a_2 \cdots a_n] \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

their product  $\mathbf{a}'\mathbf{x}$  is defined as

$$\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^n a_ix_i.$$

It is called the *inner product* of the vectors  $\mathbf{a}$  and  $\mathbf{x}$ . It exists only when  $\mathbf{a}$  and  $\mathbf{x}$  have the same order; when they are not of the same order the product  $\mathbf{a}'\mathbf{x}$  is undefined.

### b. A matrix-vector product

**Illustration.** For the preceding illustration suppose the animal prices in a neighboring town were \$2, \$2 and \$8, respectively. Let us represent them by the row vector  $[2 \quad 2 \quad 8]$ . Then purchasing the chemistry department's requirements in the neighboring town would cost

$$[2 \quad 2 \quad 8] \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = 2(50) + 2(100) + 8(30) = 540 \quad (9)$$

dollars, calculated just like (8).

Now put the two sets of prices as the rows of a matrix,

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}.$$

Then the products (8) and (9) can be represented simultaneously as a single

product of the matrix  $A$  and the vector  $x$ :

$$Ax = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = \begin{bmatrix} 3(50) + 1(100) + 10(30) \\ 2(50) + 2(100) + 8(30) \end{bmatrix} = \begin{bmatrix} 550 \\ 540 \end{bmatrix}. \quad (10)$$

The result is a vector, its elements being the inner products (8) and (9). In terms of the rows of  $A$ , this means that the elements of the vector  $Ax$  are derived in exactly the same way as the product  $a'x$  developed earlier, using the successive rows of  $A$  as the vector  $a'$ . The result is the product  $Ax$ ; that is,  $Ax$  is obtained by repetitions of the product  $a'x$  using the rows of  $A$  successively for  $a'$  and writing the results as a column vector. Hence, on using the notation of (6), with  $\alpha'_1$  and  $\alpha'_2$  being the rows of  $A$ , we see that (10) is

$$Ax = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \end{bmatrix} x = \begin{bmatrix} \alpha'_1 x \\ \alpha'_2 x \end{bmatrix}.$$

This generalizes at once to  $A$  having  $r$  rows:

$$Ax = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \vdots \\ \alpha'_i \\ \vdots \\ \alpha'_r \end{bmatrix} x = \begin{bmatrix} \alpha'_1 x \\ \alpha'_2 x \\ \vdots \\ \alpha'_i x \\ \vdots \\ \alpha'_r x \end{bmatrix}. \quad (11)$$

Thus  $Ax$  is a column vector, with its  $i$ th element being the inner product of the  $i$ th row of  $A$  with the column vector  $x$ . Providing neither  $A$  nor  $x$  is a scalar, it is clear from this definition and from the example that  $Ax$  is defined only when the number of elements in each row of  $A$  (i.e., number of columns) is the same as the number of elements in the column vector  $x$ . And when this occurs  $Ax$  is a column vector having the same number of elements as there are rows in  $A$ . Therefore, when  $A$  has  $r$  rows and  $c$  columns and  $x$  is of order  $c$ ,  $Ax$  is a column vector of order  $r$ ; its  $i$ th element is  $\sum_{k=1}^c a_{ik} x_k$  for  $i = 1, 2, \dots, r$ . More formally, when

$$A = \{a_{ij}\} \quad \text{and} \quad x = \{x_j\} \quad \text{for} \quad i = 1, 2, \dots, r \quad \text{and} \quad j = 1, 2, \dots, c,$$

[2.8]

then

$$Ax = \left\{ \sum_{j=1}^c a_{ij} x_j \right\} \quad \text{for} \quad i = 1, 2, \dots, r.$$

**Example**

$$\begin{bmatrix} 4 & 2 & 1 & 3 \\ 2 & 0 & -4 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4(1) + 2(0) + 1(-1) + 3(3) \\ 2(1) + 0(0) + -4(-1) + 7(3) \end{bmatrix} = \begin{bmatrix} 12 \\ 27 \end{bmatrix}.$$

Each element of  $Ax$  is obtained by moving along a row of  $A$  and down the column  $x$ , multiplying each pair of corresponding elements and adding the products. This is always the procedure for calculating the product  $Ax$ . It is also, as we shall see, the basis for calculating the product  $AB$  of two matrices.

**Illustration.** Typical of many uses of the matrix-vector product is one that occurs in the study of inbreeding, where what is known as the generation matrix is used to relate the frequencies of mating types in one generation to those in another. Kempthorne (1957, p. 108 et seq.), for example, gives the result earlier stated by Fisher (1949) that if after one generation of full-sib mating  $f^{(1)}$  represents the vector of frequencies of the seven distinct types of mating possible in this situation, then  $f^{(1)} = Af^{(0)}$ , where  $f^{(0)}$  is the vector of initial frequencies and where  $A$  is the matrix

$$A = \begin{bmatrix} 1 & \frac{2}{16} & 0 & \frac{1}{4} & 0 & \frac{1}{16} & 0 \\ 0 & \frac{4}{16} & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{16} & \frac{4}{16} \\ 0 & \frac{2}{16} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{16} & 0 & \frac{2}{4} & 0 & \frac{4}{16} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{16} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{6}{16} & \frac{8}{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{4}{16} \end{bmatrix}.$$

This simply means that if  $f_i^{(1)}$  is the relative frequency of the  $i$ th type of mating after a generation of full-sib mating [i.e., the  $i$ th element of the vector  $f^{(1)}$ ] and if  $f_i^{(0)}$  is the corresponding initial frequency [ $i$ th element of  $f^{(0)}$ ], then, for example,

$$f_1^{(1)} = f_1^{(0)} + \frac{2}{16}f_2^{(0)} + \frac{1}{4}f_4^{(0)} + \frac{1}{16}f_6^{(0)};$$

and as another example

$$f_4^{(1)} = \frac{8}{16}f_2^{(0)} + \frac{2}{4}f_4^{(0)} + \frac{4}{16}f_6^{(0)}.$$

These and five similar equations are implied in the vector equation  $\mathbf{f}^{(1)} = \mathbf{A}\mathbf{f}^{(0)}$ . The matrix  $\mathbf{A}$  which represents the relationships between the two sets of frequencies is known in this context as the generation matrix.

### c. A product of two matrices

Multiplying two matrices can be explained as a simple repetitive extension of multiplying a matrix by a vector.

**Illustration.** Continuing the illustration of buying laboratory animals, suppose the biochemistry department needed 60 rats, 80 mice and 40 rabbits. The costs of these, obtained in the manner of (10), are shown in the product

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 60 \\ 80 \\ 40 \end{bmatrix} = \begin{bmatrix} 660 \\ 600 \end{bmatrix}. \quad (12)$$

Similarly, if the nutrition department needed 90, 30 and 20 animals, and the physiology department needed 30, 20 and 10, respectively, their costs would be

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 90 \\ 30 \\ 20 \end{bmatrix} = \begin{bmatrix} 500 \\ 400 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 30 \\ 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 210 \\ 180 \end{bmatrix}. \quad (13)$$

(The reader should verify the validity of these products.) Now, by writing the four column vectors of departmental requirements alongside one another as a matrix,

$$\mathbf{B} = \begin{bmatrix} 50 & 60 & 90 & 30 \\ 100 & 80 & 30 & 20 \\ 30 & 40 & 20 & 10 \end{bmatrix},$$

the products in (10), (12) and (13) can be represented as the single matrix product

$$\mathbf{AB} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 & 60 & 90 & 30 \\ 100 & 80 & 30 & 20 \\ 30 & 40 & 20 & 10 \end{bmatrix} = \begin{bmatrix} 550 & 660 & 500 & 210 \\ 540 & 600 & 400 & 180 \end{bmatrix}. \quad (14)$$

[2.8]

This example illustrates how the product  $\mathbf{AB}$  is simply a case of obtaining the product of  $\mathbf{A}$  with each column of  $\mathbf{B}$  and setting the products alongside one another. Thus for

$$\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \quad \mathbf{b}_4]$$

partitioned into columns in the manner of (5),

$$\mathbf{AB} = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \mathbf{Ab}_3 \quad \mathbf{Ab}_4].$$

Then, with  $\mathbf{A}$  partitioned into its rows

$$\mathbf{A} = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \end{bmatrix} \quad \text{we have} \quad \mathbf{AB} = \begin{bmatrix} \alpha'_1\mathbf{b}_1 & \alpha'_1\mathbf{b}_2 & \alpha'_1\mathbf{b}_3 & \alpha'_1\mathbf{b}_4 \\ \alpha'_2\mathbf{b}_1 & \alpha'_2\mathbf{b}_2 & \alpha'_2\mathbf{b}_3 & \alpha'_2\mathbf{b}_4 \end{bmatrix}.$$

Hence the element of  $\mathbf{AB}$  in its  $i$ th row and  $j$ th column is the inner product  $\alpha'_i\mathbf{b}_j$  of the  $i$ th row of  $\mathbf{A}$  and the  $j$ th column of  $\mathbf{B}$ . And this is true in general:

$$\begin{aligned} \mathbf{A}_{r \times c} \mathbf{B}_{c \times s} &= \{\alpha'_i\mathbf{b}_j\} \\ &= \{a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{ic}b_{cj}\} \\ &= \left\{ \sum_{k=1}^c a_{ik}b_{kj} \right\}, \quad \text{for } i = 1, \dots, r \quad \text{and } j = 1, \dots, s. \end{aligned} \quad (15)$$

The  $ij$ th element of  $\mathbf{AB}$  can therefore be obtained by thinking of moving from element to element along the  $i$ th row of  $\mathbf{A}$  and simultaneously down the  $j$ th column of  $\mathbf{B}$ , summing the products of corresponding elements. The resulting sum is the  $ij$ th element of  $\mathbf{AB}$ . Schematically the operation can be represented as follows:

to obtain  $\mathbf{AB}$ ,

$$\left[ \begin{array}{c} i \text{th row} \\ \longrightarrow \end{array} \right]_{r \times c} \left[ \begin{array}{c} j \text{th} \\ \downarrow \text{column} \end{array} \right]_{c \times s} = \left[ \begin{array}{c} (i, j) \text{th} \\ \text{element} \end{array} \right]_{r \times s}$$

for  $i = 1, 2, \dots, r$  and  $j = 1, 2, \dots, s$ . The arrows indicate moving along the  $i$ th row and simultaneously down the  $j$ th column, summing the products of corresponding elements to get the  $(i, j)$ th element of the product.

Once again, this is a matrix operation defined only if a certain condition is met: the  $i$ th row of  $\mathbf{A}$  (and hence all rows) must have the same number of elements as does the  $j$ th column of  $\mathbf{B}$  (and hence all columns). Since the number of elements in a row of a matrix is the number of columns in the

matrix (and the number of elements in a column is the number of rows), this means that there must be exactly as many columns in **A** as there are rows in **B**. Thus the matrix product **AB** is defined only if the number of columns of **A** equals the number of rows of **B**. Note also, particularly in the numerical examples, that the product **AB** has the same number of rows as **A** and the same number of columns as **B**. This is true in general.

The important consequences of the definition of matrix multiplication are therefore as follows. The product **AB** of two matrices **A** and **B** is defined and therefore exists only if the number of columns in **A** equals the number of rows in **B**; the matrices are then said to be *conformable for multiplication for the product AB*, and **AB** has the same number of rows as **A** and the same number of columns as **B**. And the *ij*th element of **AB** is the inner product of the *i*th row of **A** and *j*th column of **B**.

**Example.** For

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 4 & 3 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 0 & 6 & 1 & 5 \\ 1 & 1 & 0 & 7 \\ 3 & 4 & 4 & 3 \end{bmatrix}$$

the element in the first row and first column of the product **AB** is the inner product of the first row of **A** and the first column of **B** and is

$$1(0) + 0(1) + 2(3) = 6;$$

that in the first row and second column is

$$1(6) + 0(1) + 2(4) = 14;$$

and the element of **AB** in the second row and third column is

$$-1(1) + 4(0) + 3(4) = 11.$$

In this way **AB** is obtained as

$$\mathbf{AB} = \begin{bmatrix} 6 & 14 & 9 & 11 \\ 13 & 10 & 11 & 32 \end{bmatrix}.$$

The reader should verify this result.

#### d. Existence of matrix products

As noted in Section 1.6, subscript notation can be used to denote the order of a matrix; thus  $\mathbf{A}_{r \times c}$  is a matrix of order *r* by *c* (*r* rows and *c*

columns). A product **AB** can be written in this notation as

$$\mathbf{A}_{r \times c} \mathbf{B}_{c \times s} = \mathbf{P}_{r \times s},$$

a form which provides both for checking the conformability of **A** and **B** and for ascertaining the order of their product. Repeated use of this also simplifies determining the order of a matrix derived by multiplying several matrices together. Adjacent subscripts (which must be equal for conformability) "cancel out," leaving the first and last subscripts as the order of the product. For example, the product

$$\mathbf{A}_{r \times c} \mathbf{B}_{c \times s} \mathbf{C}_{s \times t} \mathbf{D}_{t \times u}$$

is a matrix of order  $r \times u$ .

This notation also demonstrates what is by now readily apparent from the definition of matrix multiplication, namely that the product **BA** does not necessarily exist, even if **AB** does. For **BA** can be written as  $\mathbf{B}_{c \times s} \mathbf{A}_{r \times c}$ , which we see at once is a legitimate product only if  $s = r$ . Otherwise **BA** is not defined. There are therefore three situations regarding the product of two matrices **A** and **B**. If **A** is of order  $r \times c$ :

- (i) **AB** exists only if **B** has *c* rows.
- (ii) **BA** exists only if **B** has *r* columns.
- (iii) **AB** and **BA** both exist only if **B** is  $c \times r$ .

A corollary to (iii) is that  $\mathbf{A}^2$  exists only when **A** is square. Another corollary is that both **AB** and **BA** always exist and are of the same order when **A** and **B** are square and of the same order. But as shall be shown subsequently, the two products are not necessarily equal. Their inequality will be discussed when considering the commutative law of multiplication, but meanwhile we simply state that they are not in general equal.

As a means of distinction, **AB** is described as *A postmultiplied by B*, or as *A multiplied on the right by B*; and **BA** is either *A premultiplied by B*, or *A multiplied on the left by B*.

#### e. Products with vectors

Both the inner product of two vectors and the product of a matrix postmultiplied by a column vector (see Sections 2.5a and b) are special cases of the general matrix product  $\mathbf{A}_{r \times c} \mathbf{B}_{c \times s} = \mathbf{P}_{r \times s}$ . Thus for the inner product of two vectors,  $r = 1$  and  $s = 1$ , which means that  $\mathbf{A}_{r \times c}$  becomes  $\mathbf{A}_{1 \times c}$ , a row vector ( $\mathbf{a}'$ ) $_{1 \times c}$ , say; and  $\mathbf{B}_{c \times s}$  becomes  $\mathbf{B}_{c \times 1}$ , a column vector  $\mathbf{b}_{c \times 1}$ . Thus



we have the inner product

$$(\mathbf{a}')_{1 \times c} \mathbf{b}_{c \times 1} = p_{1 \times 1},$$

a scalar. And the product in reverse order, where  $\mathbf{a}$  and  $\mathbf{b}$  can now be of different orders, is

$$\mathbf{b}_{c \times 1} (\mathbf{a}')_{1 \times r} = \mathbf{P}_{c \times r},$$

a matrix: it is called the *outer product* of  $\mathbf{b}$  and  $\mathbf{a}'$ . For the matrix-vector product of Section 2.5b, we have

$$\mathbf{A}_{r \times c} \mathbf{b}_{c \times 1} = \mathbf{p}_{r \times 1},$$

the product being a column vector; similarly, a row vector postmultiplied by a matrix is a row vector:

$$(\mathbf{a}')_{1 \times c} \mathbf{B}_{c \times r} = \mathbf{p}'_{1 \times r}.$$

In words these four results (for conformable products) are as follows:

- (i) A row vector postmultiplied by a column vector is a scalar.
- (ii) A column vector postmultiplied by a row vector is a matrix.
- (iii) A matrix postmultiplied by a column vector is a column vector.
- (iv) A row vector postmultiplied by a matrix is a row vector.

**Examples.** Given

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix}, \quad \mathbf{a}' = [1 \quad 5], \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

then

$$(i) \quad \mathbf{a}'\mathbf{b} = [1 \quad 5] \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 8.$$

$$(ii) \quad \mathbf{b}\mathbf{a}' = \begin{bmatrix} 3 \\ 1 \end{bmatrix} [1 \quad 5] = \begin{bmatrix} 3 & 15 \\ 1 & 5 \end{bmatrix}.$$

$$(iii) \quad \mathbf{A}\mathbf{b} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 25 \end{bmatrix}.$$

$$(iv) \quad \mathbf{a}'\mathbf{B} = [1 \quad 5] \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix} = [1 \quad 13 \quad 26].$$

**Illustration (Markov Chain).** In the taxicab illustration of Section 1.6 suppose the driver lives in Town 1 and starts work from there each day. Denote this by a vector  $\mathbf{x}'_0 = [1 \quad 0]$ . In general  $\mathbf{x}'$  is called a *state probability vector* (or simply *state vector*). The subscript 0 represents, the beginning of the day, and the elements 1 and 0 are probabilities of starting in Town 1 and in Town 2, respectively. This being so, the probabilities of being in Town 1 or Town 2 after the morning's first fare are, using the transition probability matrix  $\mathbf{P}$  of (1) in Section 1.6,

$$\mathbf{x}'_1 = \mathbf{x}'_0 \mathbf{P} = [1 \quad 0] \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} = [.2 \quad .8]. \quad (16)$$

And after the second fare the probabilities are

$$\mathbf{x}'_2 = \mathbf{x}'_1 \mathbf{P} = [.2 \quad .8] \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} = [.36 \quad .64]. \quad (17)$$

**Illustration (Linear Programming).** An order for 1600 boxes of coronas, 500 boxes of half coronas and 2000 boxes of cigarillos comes to a cigar-making company. The daily production and operating costs of the company's two small plants are shown in the following table:

	Daily Production (100's of boxes)		Order Received
	Plant 1	Plant 2	
Product			
Corona	8	2	16
Half corona	1	1	5
Cigarillo	2	7	20
Daily operating cost	\$1000	\$2000	

The company president needs to decide how many days each plant should operate to fill the order.

The decision will be made on the criterion of filling the order with minimum cost. Let Plant 1 operate for  $x_1$  days and Plant 2 for  $x_2$  days to fill the order. Then to minimize cost we need to

$$\text{Minimize } f = 1000x_1 + 2000x_2,$$

and to fill the order  $x_1$  and  $x_2$  must satisfy

$$8x_1 + 2x_2 \geq 16$$

$$x_1 + x_2 \geq 5$$

$$2x_1 + 7x_2 \geq 20,$$

and because plants cannot operate for negative days we must also have

$$x_1 \geq 0 \quad \text{and} \quad x_2 \geq 0.$$

This is an example of what is called a linear programming problem. It can be stated generally as

$$\text{Minimize } f = \mathbf{c}'\mathbf{x}, \quad \text{subject to } \mathbf{Ax} \geq \mathbf{r} \quad \text{and} \quad \mathbf{x} \geq \mathbf{0} \quad (18)$$

where the inequality sign is used in vector and matrix statements in exactly the same manner as the equality sign:  $\mathbf{u} \geq \mathbf{v}$  means each element of  $\mathbf{u}$  is equal to or greater than the corresponding element of  $\mathbf{v}$ , i.e.,  $u_i \geq v_i$  for all  $i$ . In our example (18) has

$$\mathbf{c} = \begin{bmatrix} 1000 \\ 2000 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 8 & 2 \\ 1 & 1 \\ 2 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 16 \\ 5 \\ 20 \end{bmatrix}.$$

Linear programming has its own vast literature with many problems involving more than just two variables (as does our example), solution of which can be quite difficult. Nevertheless, being able to state problems succinctly in matrix terminology, as in (18), provides a basis for being able to work efficiently with characteristics of a problem that lead to its solution. For example, it is known that (18) is equivalent to the problem

$$\text{Maximize } g = \mathbf{r}'\mathbf{z}, \quad \text{subject to } \mathbf{A}'\mathbf{z} \leq \mathbf{c} \quad \text{and} \quad \mathbf{z} \geq \mathbf{0}. \quad (19)$$

Furthermore, if  $\mathbf{x}_0$  and  $\mathbf{z}_0$  are solutions of (18) and (19), respectively, then the minimum  $f$  in (18) and the maximum  $g$  in (19) are equal:  $f_{\min} = \mathbf{c}'\mathbf{x}_0 = g_{\max} = \mathbf{r}'\mathbf{z}_0$ . Derivation of such results and development of solutions  $\mathbf{x}_0$  and  $\mathbf{z}_0$  are beyond the scope of this book, but the ability to state complicated problems and relationships in terms of simple matrix products is to be noted.

#### f. Products with scalars

To the extent that a scalar can be considered a  $1 \times 1$  matrix, certain cases of the scalar multiplication of a matrix in Section 2.4 are included in

[2.8]

the general matrix product  $\mathbf{A}_{r \times c} \mathbf{B}_{c \times s} = \mathbf{P}_{r \times s}$ , to wit  $a_{1 \times 1} b'_{1 \times c} = ab'$  and  $(a')_{r \times 1} b_{1 \times 1} = ba'$ .

#### g. Products with null matrices

For any matrix  $\mathbf{A}_{r \times s}$ , pre- or post-multiplication by a null matrix of appropriate order results in a null matrix. Thus if  $\mathbf{0}_{c \times r}$  is a null matrix of order  $c \times r$ ,

$$\mathbf{0}_{c \times r} \mathbf{A}_{r \times s} = \mathbf{0}_{c \times s} \quad \text{and} \quad \mathbf{A}_{r \times s} \mathbf{0}_{s \times p} = \mathbf{0}_{r \times p}.$$

For example,

$$[0 \quad 0] \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix} = [0 \quad 0 \quad 0].$$

In writing this equality as  $\mathbf{0A} = \mathbf{0}$ , it is important to notice that the two  $\mathbf{0}$ 's are not of the same order.

#### h. Products with diagonal matrices

A diagonal matrix is defined in Section 1.6 as a square matrix having all off-diagonal elements zero. Multiplication by a diagonal matrix is particularly easy: premultiplication of a matrix  $\mathbf{A}$  by a diagonal matrix  $\mathbf{D}$  gives a matrix whose rows are those of  $\mathbf{A}$  multiplied by the respective diagonal elements of  $\mathbf{D}$ .

**Illustration.** In the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}$$

of (10), rows represent prices of rats, mice and rabbits bought locally or in a neighboring town. Suppose prices increase 5% locally and 20% in the neighboring town. Then with

$$\mathbf{D} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.20 \end{bmatrix}$$

the matrix of new prices is

$$\mathbf{DA} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.20 \end{bmatrix} \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 3.15 & 1.05 & 10.50 \\ 2.40 & 2.40 & 9.60 \end{bmatrix}.$$

But rather than prices increasing according to town of purchase, suppose that for both towns rats cost 20% more, mice stay the same and rabbits

decline 10%. Then the new prices would be

$$\mathbf{A}\Delta = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 1.20 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & .90 \end{bmatrix} = \begin{bmatrix} 3.60 & 1 & 9.00 \\ 2.40 & 2 & 7.20 \end{bmatrix}.$$

### i. Identity matrices

A diagonal matrix having all diagonal elements equal to unity is called an *identity matrix*, or sometimes a *unit matrix*. It is usually denoted by the letter **I**, with a subscript for its order when necessary for clarity; e.g.,

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

When **A** is of order  $p \times q$

$$\mathbf{I}_p \mathbf{A}_{p \times q} = \mathbf{A}_{p \times q} \mathbf{I}_q = \mathbf{A}_{p \times q};$$

i.e., multiplication of a matrix **A** by a (conformable) identity matrix does not alter **A**. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix}.$$

Thus just as a null matrix is a zero of matrix algebra, so is an identity matrix a "one" or unit of the algebra. A distinction from scalar algebra is that whereas with scalar algebra there is only one zero and one "one," with matrix algebra there are many zeros (null matrices of every order) and many "ones" (identity matrices of every order).

A matrix of the form  $\lambda \mathbf{I}$  where  $\lambda$  is a scalar is called a *scalar matrix*; e.g.,

$$4\mathbf{I} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

is a scalar matrix. A scalar matrix is simply a diagonal matrix with all diagonal elements the same. Multiplication by a scalar matrix has the same effect as scalar multiplication; e.g.,  $4\mathbf{I}\mathbf{A} = 4\mathbf{A}$ .

### j. The transpose of a product

The transpose of a product matrix is the product of the transposed matrices taken in reverse sequence, i.e.,  $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ .

### Example

$$\mathbf{AB} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \\ 3 & 0 & 7 \end{bmatrix} = \begin{bmatrix} -2 & 1 & -6 \\ 11 & 0 & 19 \end{bmatrix}$$

$$\mathbf{B}'\mathbf{A}' = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 2 & 0 \\ 1 & 4 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} -2 & 11 \\ 1 & 0 \\ -6 & 9 \end{bmatrix} = (\mathbf{AB})'.$$

Consideration of order and conformability for multiplication confirms this result. If **A** is  $r \times s$  and **B** is  $s \times t$ , the product  $\mathbf{P} = \mathbf{AB}$  is  $r \times t$ ; i.e.,  $\mathbf{A}_{r \times s} \mathbf{B}_{s \times t} = \mathbf{P}_{r \times t}$ . But **A'** is  $s \times r$  and **B'** is  $t \times s$  and the only product to be derived from these is  $(\mathbf{B}')_{t \times s} (\mathbf{A}')_{s \times r} = \mathbf{Q}_{t \times r}$  say. That  $\mathbf{Q} = \mathbf{B}'\mathbf{A}'$  is indeed the transpose of  $\mathbf{P} = \mathbf{AB}$  is apparent from the definition of multiplication: the  $(i, j)$ th term of **Q** is the inner product of the  $i$ th row of **B'** and the  $j$ th column of **A'**, which in turn is the inner product of the  $i$ th column of **B** and the  $j$ th row of **A**, and this by the definition of multiplication is the  $(j, i)$ th term of **P**. Hence  $\mathbf{Q} = \mathbf{P}'$ , or  $\mathbf{B}'\mathbf{A}' = (\mathbf{AB})'$ . More formally,

$$\mathbf{AB} = \mathbf{P} = \{p_{ij}\} = \left\{ \sum_{k=1}^s a_{ik} b_{kj} \right\},$$

and so

$$\begin{aligned} (\mathbf{AB})' = \mathbf{P}' &= \{(p')_{ij}\} = \{p_{ji}\} = \left\{ \sum_{k=1}^s a_{jk} b_{ki} \right\} = \left\{ \sum_{k=1}^s (a')_{kj} (b')_{ik} \right\} \\ &= \left\{ \sum_{k=1}^s (b')_{ik} (a')_{kj} \right\} = \mathbf{B}'\mathbf{A}'. \end{aligned}$$

This result for the transpose of the product of two matrices extends directly to the product of more than two. For example,  $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$  and  $(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$ . Proof is left as an exercise for the reader.

### k. The trace of a product

The trace of a matrix is defined in Section 2. Thus if  $\mathbf{A} = \{a_{ij}\}$  for  $i, j = 1, 2, \dots, n$ , the trace of **A** is  $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$ , the sum of the diagonal elements. We now show that for the product **AB**,  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$  and hence  $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB})$ . Note that  $\text{tr}(\mathbf{AB})$  exists only if **AB** is square, which occurs only when **A** is  $r \times c$  and **B** is  $c \times r$ . Then if  $\mathbf{AB} = \mathbf{P} = \{p_{ij}\}$

and  $\mathbf{BA} = \mathbf{T} = \{t_{ij}\}$ ,

$$\begin{aligned}\operatorname{tr}(\mathbf{AB}) &= \sum_{i=1}^r p_{ii} = \sum_{i=1}^r \left( \sum_{j=1}^c a_{ij} b_{ji} \right) = \sum_{i=1}^r \left( \sum_{j=1}^c b_{ji} a_{ij} \right) \\ &= \sum_{j=1}^c \left( \sum_{i=1}^r b_{ji} a_{ij} \right) = \sum_{j=1}^c t_{jj} = \operatorname{tr}(\mathbf{BA}).\end{aligned}$$

Extension to products of three or more matrices is obvious.

Notice that the intermediate result

$$\operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^r \sum_{j=1}^c a_{ij} b_{ji}$$

can also, using  $\mathbf{B}' = \{b'_{ij} = b_{ji}\}$ , be expressed as

$$\operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^r \sum_{j=1}^c a_{ij} b'_{ij},$$

which is the sum of products of each element of  $\mathbf{A}$  multiplied by the corresponding element of  $\mathbf{B}'$ . And if  $\mathbf{B} = \mathbf{A}'$ , we have

$$\operatorname{tr}(\mathbf{AA}') = \operatorname{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^r \sum_{j=1}^c a_{ij}^2,$$

that the trace of  $\mathbf{AA}'$  (and of  $\mathbf{A}'\mathbf{A}$ ) is the sum of squares of elements of  $\mathbf{A}$ . It is left as an exercise for the reader to verify these results numerically.

### 1. Powers of a matrix

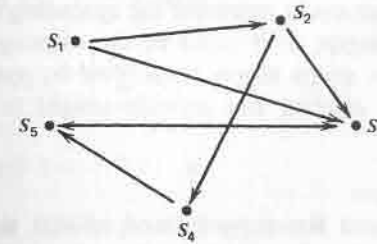
Since  $\mathbf{A}_{r \times c} \mathbf{A}_{r \times c}$  exists only if  $r = c$ , i.e., only if  $\mathbf{A}$  is square, we see that  $\mathbf{A}^2$  exists only when  $\mathbf{A}$  is square; and then  $\mathbf{A}^k$  exists for all positive integers  $k$ . And, in keeping with scalar arithmetic where  $x^0 = 1$ , we take  $\mathbf{A}^0 = \mathbf{I}$  for  $\mathbf{A}$  square.

**Illustration.** In the taxicab illustration, it is clear from (16) that (17) is

$$\begin{aligned}\mathbf{x}'_2 &= \mathbf{x}'_1 \mathbf{P} = \mathbf{x}'_0 \mathbf{P}^2 = [1 \quad 0] \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix}^2 \\ &= [1 \quad 0] \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} = [1 \quad 0] \begin{bmatrix} .36 & .64 \\ .32 & .68 \end{bmatrix} = [.36 \quad .64].\end{aligned}$$

Similarly,  $\mathbf{x}'_3 = \mathbf{x}'_0 \mathbf{P}^3$ , and in general  $\mathbf{x}'_n = \mathbf{x}'_0 \mathbf{P}^n$ .

**Illustration (Graph Theory).** Suppose in a communications network of five stations messages can be sent only in the directions of the arrows of the accompanying diagram.



This chart of the possible message routes can also be represented by a matrix  $\mathbf{T} = \{t_{ij}\}$  say, where  $t_{ij} = 0$  except  $t_{ij} = 1$  if a message can be sent from  $S_i$  to  $S_j$ . Hence

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

In the  $r$ th power of  $\mathbf{T}$ , say  $\mathbf{T}^r = \{t_{ij}^{(r)}\}$ , the element  $t_{ij}^{(r)}$  is then the number of ways of getting a message from station  $i$  to station  $j$  in exactly  $r$  steps. Thus

$$\mathbf{T}^2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{T}^3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

show that messages can be transmitted in two ways from  $S_2$  to  $S_5$  in two steps and in two ways from  $S_2$  to  $S_3$  in three steps. In this manner the pathways through a network can be counted simply by looking at powers of  $\mathbf{T}$ . More than that, if there is no direct path from  $S_i$  to  $S_j$ , i.e.,  $t_{ij} = 0$ , the powers of  $\mathbf{T}$  can be used to ascertain if there are any indirect paths. For example,  $t_{25} = 0$  but  $t_{25}^{(2)} = 2$ , showing that the path from  $S_2$  to  $S_5$  cannot be traversed directly but there are two indirect routes of two steps each. Thus only if  $\sum_{r=0}^{\infty} t_{ij}^{(r)} = 0$  is there no path at all from  $S_i$  to  $S_j$ .

Applications of matrices like  $\mathbf{T}$  arise in many varied circumstances. For example, instead of the  $S$ 's being stations in a communications network they could be people in a social or business group, whereupon each arrow of the figure could represent the dominance of one person by another. Then  $\mathbf{T}^2$  represents the two-stage dominances existing within the group of people; and so on. Or the figure could represent the spreading of a rumor between people or groups of people; or it could be the winning of sports contests; and so on. In this way, graph theory, supported by matrix representations and attendant matrix algebra, can provide insight to a host of real-life problems.

### m. Partitioned matrices

When matrices  $\mathbf{A}$  and  $\mathbf{B}$  are partitioned so that their submatrices are appropriately conformable for multiplication, the product  $\mathbf{AB}$  can be expressed in partitioned form having submatrices that are functions of the submatrices of  $\mathbf{A}$  and  $\mathbf{B}$ . For example, if

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

then

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

This means that the partitioning of  $\mathbf{A}$  along its columns must be the same as that of  $\mathbf{B}$  along its rows. Then  $\mathbf{A}_{11}$  (and  $\mathbf{A}_{21}$ ) has the same number of columns as  $\mathbf{B}_{11}$  has rows, and  $\mathbf{A}_{12}$  (and  $\mathbf{A}_{22}$ ) has the same number of columns as  $\mathbf{B}_{21}$  has rows. We see at once that when two matrices are appropriately partitioned the submatrices of their product are obtained by treating the submatrices of each of them as elements in a normal matrix product, and the individual elements of the product are derived in the usual way from the products of the submatrices.

In general, if  $\mathbf{A}$  is  $p \times q$  and is partitioned as

$$\mathbf{A}_{p \times q} = \{\mathbf{A}_{ij}(p_i \times q_j)\}, \quad \text{for } i = 1, 2, \dots, r \quad \text{and } j = 1, 2, \dots, c$$

with  $\sum_{i=1}^r p_i = p$  and  $\sum_{j=1}^c q_j = q$ , where  $p_i \times q_j$  is the order of the submatrix

[2.8]

$\mathbf{A}_{ij}$ , and likewise if

$$\mathbf{B}_{q \times s} = \{\mathbf{B}_{jk}(q_j \times s_k)\}, \quad \text{for } j = 1, 2, \dots, c \quad \text{and } k = 1, 2, \dots, d$$

with  $\sum_{j=1}^c q_j = q$  and  $\sum_{k=1}^d s_k = s$ , then

$$(\mathbf{AB})_{p \times s} = \left\{ \sum_{j=1}^c \mathbf{A}_{ij} \mathbf{B}_{jk}(p_i \times s_k) \right\},$$

for  $i = 1, 2, \dots, r$  and  $k = 1, 2, \dots, d$ .

**Illustration.** Feller (1968, p. 439) gives an example of a matrix of transition probabilities that can be partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{U}_1 & \mathbf{V}_1 & \mathbf{T} \end{bmatrix},$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are also transition probability matrices. It is readily shown that for  $\mathbf{U}_n = \mathbf{U}_1 \mathbf{A}^{n-1} + \mathbf{T} \mathbf{U}_{n-1}$  and  $\mathbf{V}_n = \mathbf{V}_1 \mathbf{B}^{n-1} + \mathbf{T} \mathbf{V}_{n-1}$ ,

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{A}^n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^n & \mathbf{0} \\ \mathbf{U}_n & \mathbf{V}_n & \mathbf{T}^n \end{bmatrix}.$$

### n. Hadamard products

The definition of a product that has been presented at some length is the one most generally used. But because matrices are arrays of numbers, they provide opportunity for defining products in several different ways. One of these is described here and another in Section 10.7.

The Hadamard product of matrices  $\mathbf{A} = \{a_{ij}\}$  and  $\mathbf{B} = \{b_{ij}\}$  is defined only when  $\mathbf{A}$  and  $\mathbf{B}$  have the same order. It is the matrix of the element-by-element products of corresponding elements in  $\mathbf{A}$  and  $\mathbf{B}$ :

$$\mathbf{A} \cdot \mathbf{B} = \{a_{ij} b_{ij}\}.$$

Thus the  $ij$ th element of the Hadamard product  $\mathbf{A} \cdot \mathbf{B}$  is the product of the  $ij$ th elements of  $\mathbf{A}$  and  $\mathbf{B}$ :

$$\begin{bmatrix} 2 & 0 & -1 \\ 4 & 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 7 \\ 2 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -7 \\ 8 & 0 & 18 \end{bmatrix}.$$

**Illustration.** Suppose in the taxicab illustration that the average fare between the two towns is \$8 and within Town 1 it is \$6 and within Town 2 it is \$5. Then, corresponding to the transition probability matrix, the reward matrix is

$$\mathbf{R} = \begin{bmatrix} 6 & 8 \\ 8 & 5 \end{bmatrix}.$$

Therefore, since the expected reward for a fare from Town  $i$  to Town  $j$  is  $p_{ij}r_{ij}$ , the matrix of expected rewards is the Hadamard product of  $\mathbf{P}$  and  $\mathbf{R}$ :

$$\mathbf{P} \cdot \mathbf{R} = \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} \cdot \begin{bmatrix} 6 & 8 \\ 8 & 5 \end{bmatrix} = \begin{bmatrix} 1.2 & 6.4 \\ 3.2 & 3.0 \end{bmatrix}. \quad (20)$$

This is needed in Exercise 25.

## 9. THE LAWS OF ALGEBRA

We now give formal consideration to the associative, commutative and distributive laws of algebra as they relate to the addition and multiplication of matrices.

### a. Associative laws

The addition of matrices is associative provided the matrices are conformable for addition. For if  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  have the same order,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \{a_{ij} + b_{ij}\} + \{c_{ij}\} = \{a_{ij} + b_{ij} + c_{ij}\} = \mathbf{A} + \mathbf{B} + \mathbf{C}.$$

Also,

$$\{a_{ij} + b_{ij} + c_{ij}\} = \{a_{ij}\} + \{b_{ij} + c_{ij}\} = \mathbf{A} + (\mathbf{B} + \mathbf{C}),$$

so proving the associative law of addition.

In general, the laws of algebra that hold for matrices do so because matrix results follow directly from corresponding scalar results for their elements—as illustrated here. Further proofs in this section are therefore omitted.

The associative law is also true for multiplication, provided the matrices are conformable for multiplication. For if  $\mathbf{A}$  is  $p \times q$ ,  $\mathbf{B}$  is  $q \times r$  and  $\mathbf{C}$  is  $r \times s$ , then  $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$ .

### b. The distributive law

The distributive law holds true. For example,

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

provided both  $\mathbf{B}$  and  $\mathbf{C}$  are conformable for addition (necessarily of the same order) and  $\mathbf{A}$  and  $\mathbf{B}$  are conformable for multiplication (and hence  $\mathbf{A}$  and  $\mathbf{C}$  also).

### c. Commutative laws

Addition of matrices is commutative (provided the matrices are conformable for addition). If  $\mathbf{A}$  and  $\mathbf{B}$  are of the same order

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\} = \{b_{ij} + a_{ij}\} = \mathbf{B} + \mathbf{A}.$$

Multiplication of matrices is not in general commutative. As seen earlier, there are two possible products that can be derived from matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,  $\mathbf{AB}$  and  $\mathbf{BA}$ , and if  $\mathbf{A}$  is of order  $r \times c$  both products exist only if  $\mathbf{B}$  is of order  $c \times r$ .  $\mathbf{AB}$  is then square, of order  $r$ , and  $\mathbf{BA}$  is also square, of order  $c$ . Possible equality of  $\mathbf{AB}$  and  $\mathbf{BA}$  can therefore be considered only where  $r = c$ , in which case  $\mathbf{A}$  and  $\mathbf{B}$  are both square and have the same order  $r$ . The products are then

$$\mathbf{AB} = \left\{ \sum_{k=1}^r a_{ik} b_{kj} \right\} \quad \text{and} \quad \mathbf{BA} = \left\{ \sum_{k=1}^r b_{ik} a_{kj} \right\} \quad \text{for } i, j = 1, \dots, r.$$

It can be seen that the  $(i, j)$ th elements of these products do not necessarily have even a single term in common in their sums of products, let alone are they equal. Therefore, even when  $\mathbf{AB}$  and  $\mathbf{BA}$  both exist and are of the same order, they are not in general equal; for example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 4 & -7 \end{bmatrix} \\ \neq \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -3 & -4 \\ -2 & -2 \end{bmatrix}.$$

But in certain cases  $\mathbf{AB}$  and  $\mathbf{BA}$  are equal; for example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 12 & 18 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Two special cases of matrix multiplication being commutative are  $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$  and  $\mathbf{0A} = \mathbf{A0} = \mathbf{0}$  for  $\mathbf{A}$  being square. When  $\mathbf{A}$  is rectangular the former holds, with identity matrices of different orders, i.e.,  $\mathbf{I}_r \mathbf{A}_{r \times c} = \mathbf{A}_{r \times c} \mathbf{I}_c = \mathbf{A}_{r \times c}$ ; the latter can be expressed more generally as two separate statements:  $\mathbf{0}_{p \times r} \mathbf{A}_{r \times c} = \mathbf{0}_{p \times c}$  and  $\mathbf{A}_{r \times c} \mathbf{0}_{c \times s} = \mathbf{0}_{r \times s}$ .

### 10. CONTRASTS WITH SCALAR ALGEBRA

The definition of matrix multiplication leads to results in matrix algebra that have no counterpart in scalar algebra. In fact, some matrix results contradict their scalar analogues. We give some examples.

First, although in scalar algebra  $ax + bx$  can be factored either as  $x(a + b)$  or as  $(a + b)x$ , this duality is not generally possible with matrices:

$$\mathbf{AX} + \mathbf{BX} = (\mathbf{A} + \mathbf{B})\mathbf{X} \quad \text{and} \quad \mathbf{XA} + \mathbf{XB} = \mathbf{X}(\mathbf{A} + \mathbf{B}),$$

but

$$\mathbf{XP} + \mathbf{QX} \text{ generally does not have } \mathbf{X} \text{ as a factor.}$$

Another example of factoring is that, similar to  $xy - x = x(y - 1)$  in scalar algebra,  $\mathbf{XY} - \mathbf{X} = \mathbf{X}(\mathbf{Y} - \mathbf{I})$  in matrix algebra, but with the second term inside the parentheses being  $\mathbf{I}$  and not the scalar 1. (It cannot be 1 because 1 is not conformable for subtraction from  $\mathbf{Y}$ .) Furthermore, the fact that it is  $\mathbf{I}$  emphasizes, because of conformability for subtraction from  $\mathbf{Y}$ , that  $\mathbf{Y}$  must be square. Of course this can also be gleaned directly from  $\mathbf{XY} - \mathbf{X}$  itself: if  $\mathbf{X}$  is  $r \times c$ , then  $\mathbf{Y}$  must have  $c$  rows in order for  $\mathbf{XY}$  to be defined, and  $\mathbf{Y}$  must have  $c$  columns in order for  $\mathbf{XY}$  and  $\mathbf{X}$  to be conformable for subtraction. This simple example illustrates the need for constantly keeping conformability in mind, especially when matrix symbols do not have their orders attached.

Another consequence of matrix multiplication is that even when  $\mathbf{AB}$  and  $\mathbf{BA}$  both exist, they may not be equal. Thus for

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \quad (21)$$

we have

$$\mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \neq \mathbf{BA} = \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix}.$$

Notice here that even though  $\mathbf{AB} = \mathbf{0}$ , neither  $\mathbf{A}$  nor  $\mathbf{B}$  is  $\mathbf{0}$ . This illustrates an extremely important feature of matrices: the equation  $\mathbf{AB} = \mathbf{0}$  does not always lead to the conclusion that  $\mathbf{A}$  or  $\mathbf{B}$  is  $\mathbf{0}$ , as would be the case with scalars. A further illustration arises from observing that for  $\mathbf{A}$  and  $\mathbf{B}$  of (21), we have  $\mathbf{BA} = 2\mathbf{B}$ . This can be rewritten as  $\mathbf{BA} - 2\mathbf{B} = \mathbf{0}$ , equivalent to  $\mathbf{B}(\mathbf{A} - 2\mathbf{I}) = \mathbf{0}$ . But from this we cannot conclude either that  $\mathbf{A} - 2\mathbf{I}$  is  $\mathbf{0}$  or that  $\mathbf{B} = \mathbf{0}$ . Similarly, not even the equation  $\mathbf{X}^2 = \mathbf{0}$  means that  $\mathbf{X} = \mathbf{0}$ ; for example, with

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \quad \text{we have} \quad \mathbf{X}^2 = \mathbf{0}.$$

Likewise  $\mathbf{Y}^2 = \mathbf{I}$  implies neither  $\mathbf{Y} = \mathbf{I}$  nor  $\mathbf{Y} = -\mathbf{I}$ ; e.g.,

$$\mathbf{Y} = \begin{bmatrix} 1 & 0 \\ 4 & -1 \end{bmatrix}, \quad \text{but} \quad \mathbf{Y}^2 = \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly, we can have  $\mathbf{M}^2 = \mathbf{M}$  with both  $\mathbf{M} \neq \mathbf{I}$  and  $\mathbf{M} \neq \mathbf{0}$ ; e.g.,

$$\mathbf{M} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} = \mathbf{M}^2.$$

Such a matrix  $\mathbf{M}$  with  $\mathbf{M} = \mathbf{M}^2$  is said to be *idempotent*.

### EXERCISES

1. For  $\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 2 & 1 \end{bmatrix}$ ,  $\mathbf{B} = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & -1 & -1 & 1 \end{bmatrix}$ ,

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

show that

(a)  $\mathbf{AB} = \begin{bmatrix} 3 & -6 & 3 & 12 \\ 2 & -1 & 5 & 5 \end{bmatrix}$

and  $\mathbf{A'B} = \begin{bmatrix} 3 & -2 & 7 & 8 \\ 6 & -1 & 17 & 13 \end{bmatrix}$ ;

(b)  $(\mathbf{A} + \mathbf{A'})\mathbf{B} = \begin{bmatrix} 6 & 8 \\ 8 & 2 \end{bmatrix} \mathbf{B} = \begin{bmatrix} 6 & -8 & 10 & 20 \\ 8 & -2 & 22 & 18 \end{bmatrix} = \mathbf{AB} + \mathbf{A'B}$ ;

$$(c) \mathbf{BB}' = \begin{bmatrix} 14 & -1 \\ -1 & 3 \end{bmatrix} \text{ and } \mathbf{B}'\mathbf{B} = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & 1 & -1 \\ 3 & 1 & 10 & 5 \\ 2 & -1 & 5 & 5 \end{bmatrix};$$

$$(d) \text{trace } \mathbf{BB}' = \text{trace } \mathbf{B}'\mathbf{B} = 17;$$

$$(e) \mathbf{Bx} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \quad \mathbf{B}'\mathbf{Bx} = \begin{bmatrix} -1 \\ 2 \\ -1 \\ -4 \end{bmatrix}, \quad \mathbf{x}'\mathbf{B}'\mathbf{Bx} = 5 = (\mathbf{Bx})'\mathbf{Bx};$$

$$(f) \mathbf{Ay} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \quad \mathbf{A}'\mathbf{Ay} = \begin{bmatrix} -7 \\ -17 \end{bmatrix}, \quad \mathbf{y}'\mathbf{A}'\mathbf{Ay} = 10 = (\mathbf{Ay})'\mathbf{Ay};$$

$$(g) \mathbf{A}^2 - 4\mathbf{A} - 9\mathbf{I} = \mathbf{0};$$

$$(h) \frac{1}{9}\mathbf{A} \begin{bmatrix} -1 & 6 \\ 2 & -3 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -1 & 6 \\ 2 & -3 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

2. Confirm:

$$(a) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix};$$

$$(b) \text{ if } \mathbf{A} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \text{ then } \mathbf{A}^2 = \mathbf{A};$$

$$(c) \text{ if } \mathbf{B} = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{3} & 1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{3} & 0 \\ 1/\sqrt{6} & 1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}, \mathbf{BB}' = \mathbf{B}'\mathbf{B} = \mathbf{I}_3;$$

$$(d) \text{ if } \mathbf{C} = \begin{bmatrix} 6 & -4 \\ 9 & -6 \end{bmatrix}, \text{ then } \mathbf{C}^2 \text{ is null};$$

$$(e) \frac{1}{9} \begin{bmatrix} 4 & -5 & -1 \\ 1 & 1 & 2 \\ 4 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 4 & -1 \end{bmatrix} \text{ is an identity matrix.}$$

$$3. \text{ For } \mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ -1 & 0 & 7 \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} 6 & 0 & 0 \\ -3 & 4 & 0 \\ 0 & -5 & 2 \end{bmatrix}$$

find  $\mathbf{X}^2$ ,  $\mathbf{Y}^2$ ,  $\mathbf{XY}$  and  $\mathbf{YX}$ , and show that

$$(\mathbf{X} + \mathbf{Y})^2 = \mathbf{X}^2 + \mathbf{XY} + \mathbf{YX} + \mathbf{Y}^2 = \begin{bmatrix} 40 & 5 & 44 \\ -28 & 13 & -33 \\ -1 & -62 & 88 \end{bmatrix}.$$

$$4. \text{ Given } \mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 0 & 2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & -1 \\ 2 & 3 & 0 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 6 & 5 & 7 \\ 2 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix},$$

show that  $\mathbf{AX} = \mathbf{BX}$  even though  $\mathbf{A} \neq \mathbf{B}$ .

5. Show that

$$(a) \begin{bmatrix} 6 & 3 \\ 0 & 7 \\ -5 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 8 \\ 2 & -1 \\ 6 & -4 \end{bmatrix} = \begin{bmatrix} 12 & 17 \\ -11 & 2 \\ 3 & 9 \end{bmatrix} - \begin{bmatrix} 3 & 6 \\ -13 & -4 \\ 2 & 12 \end{bmatrix};$$

$$(b) 4 \begin{bmatrix} 3 & 8 \\ 1 & 9 \end{bmatrix} - 3 \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} = 2 \begin{bmatrix} 2 & -1 \\ 7 & 4 \end{bmatrix} + 5 \begin{bmatrix} 1 & 8 \\ -2 & 5 \end{bmatrix};$$

$$(c) \text{ if } \mathbf{A} = \begin{bmatrix} 3 & 8 & 4 \\ 8 & 7 & -1 \\ 4 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & 4 \\ 3 & 4 & 6 \end{bmatrix}, \text{ then } \mathbf{A} = \mathbf{A}'.$$

6. What are the transposes of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 8 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 6 \\ 8 & 9 \\ 2 & -7 \end{bmatrix}?$$

Calculate  $\mathbf{A} + \mathbf{B}'$  and  $\mathbf{A}' + \mathbf{B}$ , and explain the relationship between these two sums.

7. (a) If, in general,  $\mathbf{A}$  is  $r \times c$ , what must be the order of  $\mathbf{B}$  so that  $\mathbf{A} + \mathbf{B}'$  exists? Why? For  $\mathbf{A}_{r \times c}$ , prove the relationship between  $\mathbf{A} + \mathbf{B}'$  and  $\mathbf{A}' + \mathbf{B}$  illustrated in Exercise 6.

(b) Explain why  $\mathbf{AVA}' = \mathbf{BVB}'$  implies that  $\mathbf{A}$  and  $\mathbf{B}$  have the same order and  $\mathbf{V}$  is square.

8. Prove that  $(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$ .

**Note:** Whenever asked to prove a result such as this, one can assume that the matrices involved are conformable for the expressions given; e.g.,  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  are here conformable for the product  $\mathbf{ABCD}$ .

9. Under what conditions do both  $\text{tr}(\mathbf{ABC})$  and  $\text{tr}(\mathbf{BAC})$  exist? When they exist, do you expect them to be equal? Why? Calculate  $\text{tr}(\mathbf{ABC})$  and  $\text{tr}(\mathbf{BAC})$  for

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



10. Explain why each of the following statements is true.

(a)  $(A + A')B = AB + A'B$ .  
 (b)  $\text{tr}(B'B) = \text{tr}(BB')$ .  
 (c)  $x'B'Bx = (Bx)'Bx = \text{tr}(B'Bxx')$ .

Verify each statement for the following values of  $A$ ,  $B$  and  $x$ :

$$A = \begin{bmatrix} -1 & 2 \\ 7 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 8 & -1 & 0 \\ 2 & 3 & 1 & 4 \end{bmatrix}, \quad x' = [1 \quad 1 \quad 2 \quad 3].$$

11. For  $A$  and  $B$  having the same order, explain why

$$\begin{aligned} (A + B)(A + B)' &= (A + B)(A' + B') \\ &= AA' + AB' + BA' + BB'. \end{aligned}$$

Will these expressions also generally equal  $(A + B)'(A + B)$ ? Why must  $A$  and  $B$  have the same order here? Verify these equalities for

$$A = \begin{bmatrix} 2 & 3 & 7 \\ 6 & 1 & -2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 7 \end{bmatrix}.$$

12. For a matrix  $A$  from one of the preceding exercises, calculate  $AA'$  and  $A'A$  and verify that the trace of each is the sum of squares of elements of  $A$ .
13. The theory of one-person games against nature uses what is called a payoff matrix. Its rows correspond to possible actions, its columns correspond to possible outcomes (degrees of success or failure) and its elements are the payoffs. From the payoff matrix is derived a new matrix having every element in column  $j$  equal to the largest element of column  $j$  of the payoff matrix. The payoff matrix is subtracted from this new matrix, the result being called the regret matrix. Its elements represent the maximum losses that could be incurred by not choosing the best action for a particular outcome. What is called the *minimax* strategy of the game is to choose the action corresponding to the row of the regret matrix that contains the minimum of the maximum entries of all rows. Find the regret matrix and the minimax strategy for the following payoff matrix:

$$P = \begin{bmatrix} 0 & 2 & 100 & 0 \\ 7 & 8 & 100 & 0 \\ 10 & 40 & 1 & 90 \\ 1 & 3 & 1 & 1 \end{bmatrix}.$$

14. Shipping costs (in dollars) from three manufacturing plants to four destinations are shown in the following matrix:

$$\begin{bmatrix} 10 & 15 & 9 & 7 \\ 14 & 8 & 12 & 8 \\ 6 & 14 & 22 & 17 \end{bmatrix}.$$

Production costs are \$40, \$38 and \$41 at the three plants. What is the matrix of total costs, production plus shipping?

15. (a) When does  $(A + B)(A - B) = A^2 - B^2$ ?  
 (b) When  $A = A'$ , prove that  $\text{tr}(AB) = \text{tr}(AB')$ .  
 (c) When  $X'XGX'X = X'X$ , prove that  $X'XG'X'X = X'X$ .
16. In discussing the problem of calculating frequencies for different relatives, of pairs of genotypes at a single two-allele locus, Li (1955, p. 34) uses two matrices of conditional probabilities. They are

$$P = \begin{bmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} p & q & 0 \\ \frac{1}{2}p & \frac{1}{2} & \frac{1}{2}q \\ 0 & p & q \end{bmatrix}.$$

$P$  for when the relatives have no genes identical by descent and  $T$  for when they have one gene identical by descent. With  $p + q = 1$ , and with the vector  $\mathbf{1}' = [1 \quad 1 \quad 1]$ , show that

- (a)  $P\mathbf{1} = \mathbf{1}[p^2 \quad 2pq \quad q^2]$ , (b)  $P\mathbf{1} = \mathbf{1}$ ,  
 (c)  $T\mathbf{1} = \mathbf{1}$ , (d)  $PT = P = TP$ ,  
 (e)  $T^2 = \frac{1}{2}(P + T)$ , (f)  $P^2 = P$ ,  
 (g)  $T^n = P + (\frac{1}{2})^{n-1}(T - P)$ .

And for  $S = \frac{1}{4}I + \frac{1}{2}T + \frac{1}{4}P$ , prove that

- (h)  $ST = TS = T^2$  and (i)  $S^2 = \frac{1}{16}(I + 6T + 9P)$ .

17. A generation matrix given by Kempthorne (1957, p. 120) is

$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}. \quad \text{Given that } \mathbf{f}^{(i)} = A\mathbf{f}^{(i-1)}, \quad \text{show that}$$

$$\mathbf{f}^{(2)} = \begin{bmatrix} 1 & \frac{3}{4} \\ 0 & \frac{1}{4} \end{bmatrix} \mathbf{f}^{(0)}, \quad \mathbf{f}^{(3)} = \begin{bmatrix} 1 & \frac{7}{8} \\ 0 & \frac{1}{8} \end{bmatrix} \mathbf{f}^{(0)}$$

and  $\mathbf{f}^{(n)} = \mathbf{A}^n \mathbf{f}^{(0)} = \begin{bmatrix} 1 & 1 - 1/2^n \\ 0 & 1/2^n \end{bmatrix} \mathbf{f}^{(0)}$ .

18. With  $\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$  and  $\mathbf{B} = \begin{bmatrix} 3 & 3 & 1 \\ -4 & 2 & -1 \\ -2 & 0 & 0 \end{bmatrix}$ :

(a) Partition  $\mathbf{A}$  and  $\mathbf{B}$  as

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

where both  $\mathbf{A}_{21}$  and  $\mathbf{B}_{21}$  have order  $1 \times 2$ .

- (b) Calculate  $\mathbf{AB}$  both with and without the partitioning, to demonstrate the validity of multiplication of partitioned matrices.  
 (c) Calculate  $\mathbf{AB}'$ , showing that

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

19. Partition the following matrices in any manner conformable for the products indicated. Calculate the products both with and without using the partitioned forms, to test their validity.

(a)  $\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ ,  $\mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \\ 0 & 1 \end{bmatrix}$ , and the product  $\mathbf{AB}$ .

(b)  $\mathbf{C} = \begin{bmatrix} 4 & 1 \\ -1 & 1 \end{bmatrix}$ ,  $\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 2 & 3 & -5 \end{bmatrix}$ , and  $\mathbf{CD}$ .

(c)  $\mathbf{E} = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}$ ,  $\mathbf{F} = \begin{bmatrix} 0 & 1 & 3 \\ 4 & 4 & -2 \end{bmatrix}$ ,

$\mathbf{G} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & -3 & -2 \\ -1 & 0 & -1 \end{bmatrix}$ , and the product  $\mathbf{EFG}$ .

20. Calculate  $\mathbf{P}^2$ ,  $\mathbf{P}^3$ ,  $\mathbf{P}^4$  and  $\mathbf{P}^5$  for  $\mathbf{P}$  of (1) in Section 1.6.

21. Four friends play racquetball regularly every week. In a particular year,  $A$  dominates  $B$  and  $D$ ,  $B$  dominates  $C$ ,  $C$  dominates  $A$ , and  $D$  dominates  $C$ . Draw a chart showing these dominations, write down a matrix to represent them and find the first five powers of that matrix. What do you notice about these powers?

22. Prove that for any  $\mathbf{P}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$  and  $\mathbf{S}$  (of what orders?)

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{A} \end{bmatrix} \mathbf{B} \begin{bmatrix} \mathbf{I} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{P} \\ \mathbf{A} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{C} \\ \mathbf{R} & \mathbf{S} \end{bmatrix}.$$

23. By considering the inner product of  $\mathbf{y} - (\mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x})\mathbf{x}$  with itself, prove that

$$\left( \sum_{i=1}^n x_i y_i \right)^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2.$$

This is the Cauchy-Schwarz inequality. Use it to prove that a product-moment correlation (apart from sign) can never exceed unity.

24. Suppose for some large positive integer  $s$  and a transition probability matrix  $\mathbf{P}$  that the  $s$ th and  $(s+1)$ th powers of  $\mathbf{P}$  are effectively equal. Then

$$\mathbf{x}'_{s+1} = \mathbf{x}'_0 \mathbf{P}^{s+1} = \mathbf{x}'_0 \mathbf{P}^s = \mathbf{x}'_s$$

and so the standard relationship  $\mathbf{x}'_{s+1} = \mathbf{x}'_s \mathbf{P}$  becomes  $\mathbf{x}'_s = \mathbf{x}'_s \mathbf{P}$ . Then  $\mathbf{x}'_s$  is known as the *steady state vector* and can be determined (so long as  $s$  exists, i.e., so long as  $\mathbf{P}^n$  tends to a limit as  $n$  tends to infinity) from the equations  $\mathbf{x}'_s = \mathbf{x}'_s \mathbf{P}$ . The equation  $\mathbf{x}'_s \mathbf{1} = 1$  must also be used. Find  $\mathbf{x}'_s$  for  $\mathbf{P}$  of Section 2.8f.

25. Suppose  $v_{i,n}$  is the expected reward for being in state  $i$  of a Markov chain after  $n$  transitions. Then

$$v_{i,n} = \sum_j p_{ij} (r_{ij} + v_{j,n-1}).$$

- (a) Write this equation in matrix notation.  
 (b) Suppose  $\mathbf{v}_0$  and  $\mathbf{v}_n$  are the vectors of expected rewards after zero and  $n$  transitions, respectively. Express  $\mathbf{v}_n$  as a function of  $\mathbf{v}_0$ .  
 (c) For  $\mathbf{P}$  and  $\mathbf{R}$  of Section 2.8n, calculate  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  and  $\mathbf{v}_3$ , assuming  $\mathbf{v}_0 = \mathbf{0}$ .

## CHAPTER 3

# SPECIAL MATRICES

Numerous matrices with particular properties have attracted special names. Although their historical origins are in specific mathematical problems or applications, they were later found to have properties of broader interest. Of the vast number of such matrices, this chapter presents but a small selection of general use and interest in basic matrix algebra and elementary applications. Others appear in subsequent chapters.

### 1. SYMMETRIC MATRICES

**Illustration.** From observations taken at intervals from a steam plant in a manufacturing concern, Draper and Smith (1981, p. 616) report correlations of .94 and  $-.13$  of monthly stored fatty acid with pounds of glycerin made and with average wind velocity, respectively, and a correlation of  $-.14$  between these last two variables. Since the correlation of a variable  $X_1$  with another variable  $X_2$  is the same as that of  $X_2$  with  $X_1$ , and because the correlation of a variable with itself can be taken as 1.00, all these correlations can be assembled in a matrix as

$$\mathbf{R} = \begin{bmatrix} 1.00 & .94 & -.13 \\ .94 & 1.00 & -.14 \\ -.13 & -.14 & 1.00 \end{bmatrix} = \{r_{ij}\} \quad \text{for } i, j = 1, 2, 3,$$

where  $r_{ij}$  is the correlation of  $X_i$  with  $X_j$ .

A noticeable feature of  $\mathbf{R}$  is that every off-diagonal element  $r_{ij}$  above the diagonal equals its counterpart  $r_{ji}$  below the diagonal; e.g., the .94 in row 1, column 2 is the same as the .94 in row 2, column 1. Indeed, this symmetry exists for all off-diagonal elements. As a result, the elements of any row of  $\mathbf{R}$

[3.1]

are the same as those of the corresponding column. Consequently, the transpose of  $\mathbf{R}$  is  $\mathbf{R}$  itself. Such a matrix is called a *symmetric matrix*; it is, of course, square.

A square matrix is defined as symmetric when it equals its transpose; i.e.,

$$\mathbf{A} \text{ is symmetric when } \mathbf{A} = \mathbf{A}', \text{ with } a_{ij} = a_{ji} \quad (1)$$

for  $i, j = 1, \dots, r$  for  $\mathbf{A}_{r \times r}$ .

Symmetric matrices have many useful special properties, a few of which (relating to matrix products) are noted here. Others are noted in later chapters.

#### a. Products of symmetric matrices

Products of symmetric matrices are not generally symmetric. If  $\mathbf{A} = \mathbf{A}'$  and  $\mathbf{B} = \mathbf{B}'$ , then  $\mathbf{A}$  and  $\mathbf{B}$  are symmetric but the transpose of their product (when it exists) is

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' = \mathbf{BA}.$$

Since  $\mathbf{BA}$  is generally not the same as  $\mathbf{AB}$ , this means  $\mathbf{AB}$  is generally not symmetric.

**Example.** With

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 3 & 7 \\ 7 & 6 \end{bmatrix},$$

$$(\mathbf{AB})' = \begin{bmatrix} 17 & 19 \\ 27 & 32 \end{bmatrix}' = \begin{bmatrix} 17 & 27 \\ 19 & 32 \end{bmatrix} = \mathbf{BA} \neq \mathbf{AB}.$$

#### b. Properties of $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$

Products of a matrix and its transpose always exist and are symmetric:

$$(\mathbf{AA}')' = (\mathbf{A}')'\mathbf{A}' = \mathbf{AA}' \quad \text{and} \quad (\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}. \quad (2)$$

Observe the method used for showing that a matrix is symmetric: transpose the matrix and show that the result equals the matrix itself.

Although both products  $\mathbf{AA}'$  and  $\mathbf{A}'\mathbf{A}$  are symmetric, they are not necessarily equal. In fact, only when  $\mathbf{A}$  is square might they be equal, because it is only then that both products have the same order. Thus for  $\mathbf{A}_{r \times c}$ , the product  $\mathbf{AA}' = \mathbf{A}_{r \times c}(\mathbf{A}')_{c \times r}$  has order  $r \times r$ , and  $\mathbf{A}'\mathbf{A} = (\mathbf{A}')_{c \times r}\mathbf{A}_{r \times c}$  has order  $c \times c$ .

**Example.** For

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

$$AA' = \begin{bmatrix} 6 & 2 \\ 2 & 10 \end{bmatrix}, \text{ and } A'A = \begin{bmatrix} 10 & 2 & 2 \\ 2 & 4 & -2 \\ 2 & -2 & 2 \end{bmatrix}.$$

Note that matrix multiplication ensures that elements of  $AA'$  are inner products of rows of  $A$  with themselves and with each other:

$$AA' = \{\text{inner product of } i\text{th and } k\text{th rows of } A\} \text{ for } i, k = 1, \dots, r. \quad (3)$$

In particular, the  $i$ th diagonal element of  $AA'$  is the sum of squares of the elements of the  $i$ th row of  $A$ , namely  $\sum_{j=1}^c a_{ij}^2$  for  $A$  of order  $r \times c$ . Confining attention to real matrices (Section 1.2), we can use the property of real numbers that a sum of squares of them is positive (unless they are all zero), and so observe that for real matrices  $A$ , the matrix  $AA'$  has diagonal elements that are positive (or zero). Similar results hold for  $A'A$  in terms of columns of  $A$ :

$$A'A = \{\text{inner product of } j\text{th and } m\text{th columns of } A\} \text{ for } j, m = 1, \dots, c. \quad (4)$$

The reader can verify these results for the example.

Recall that if a sum of squares of real numbers is zero, then each of the numbers is zero; i.e., for real numbers  $x_1, x_2, \dots, x_n$ ,  $\sum_{i=1}^n x_i^2 = 0$  implies  $x_1 = 0 = x_2 = \dots = x_n$ . This is the basis for proving, for any real matrix  $A$ , that

$$A'A = \mathbf{0} \text{ implies } A = \mathbf{0} \quad (5)$$

and

$$\text{tr}(A'A) = 0 \text{ implies } A = \mathbf{0}. \quad (6)$$

Result (5) is true because  $A'A = \mathbf{0}$  implies that every diagonal element of  $A'A$  is zero; and for the  $j$ th such element this means, by (3), that the inner product of the  $j$ th column of  $A$  with itself is zero; i.e., for  $A$  of order  $r \times c$ ,

that  $\sum_{k=1}^r a_{kj}^2 = 0$  for  $j = 1, \dots, c$ . Hence  $a_{kj} = 0$  for  $k = 1, \dots, r$ ; and this is true for  $j = 1, \dots, c$ . Therefore,  $A = \mathbf{0}$ . Proof of (6) is similar:

$$\text{tr}(A'A) = \sum_{j=1}^c (\textit{jth diagonal element of } A'A) = \sum_{j=1}^c \sum_{k=1}^r a_{kj}^2$$

and so  $\text{tr}(A'A) = 0$  implies that every  $a_{kj}$  is zero; i.e., that every element of  $A$  is zero. Hence  $A = \mathbf{0}$ .

Results (5) and (6) are seldom useful for the sake of some particular matrix  $A$ , but they are often helpful in developing other results in matrix algebra when  $A$  is a function of other matrices. For example, by means of (5) we can prove, for real matrices  $P, Q$  and  $X$ , that

$$PXX' = QXX' \text{ implies } PX = QX. \quad (7)$$

The proof consists of observing that

$$(PXX' - QXX')(P' - Q') \equiv (PX - QX)(PX - QX)'. \quad (8)$$

Hence if  $PXX' = QXX'$ , the left-hand side of (8) is null and so, therefore, is the right-hand side; hence by (5),  $PX - QX = \mathbf{0}$ ; i.e.,  $PX = QX$ , and (7) is established.

### c. Products of vectors

The inner product of two vectors is a scalar and is therefore symmetric:  $\mathbf{x}'\mathbf{y} = (\mathbf{x}'\mathbf{y})' = \mathbf{y}'\mathbf{x}$ . In contrast, the outer product of two vectors (see Section 2.8e) is not necessarily symmetric:  $\mathbf{xy}' = (\mathbf{yx}')' \neq (\mathbf{xy}')'$ . Indeed, such a product is not necessarily even square.

**Example.** For  $\mathbf{x}' = [1 \ 0 \ 2]$ ,  $\mathbf{y}' = [2 \ 2 \ 1]$  and  $\mathbf{z}' = [3 \ 2]$ ,

$$\mathbf{xy}' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} [2 \ 2 \ 1] = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 0 \\ 4 & 4 & 2 \end{bmatrix} \quad (9)$$

is not symmetric. Neither is

$$\mathbf{yx}' = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} [1 \ 0 \ 2] = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 0 & 4 \\ 1 & 0 & 2 \end{bmatrix}, \quad (10)$$

but its transpose does, of course, equal  $\mathbf{xy}'$ ; i.e., (10) equals the transpose of

(9); namely  $yx' = (xy)'$ . And  $xz'$  and  $zx'$  are rectangular:

$$xz' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & 0 \\ 6 & 4 \end{bmatrix}$$

and

$$zx' = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 \\ 2 & 0 & 4 \end{bmatrix}.$$

#### d. Sums of outer products

In the notation of (5) and (6) of Section 2.2d, where  $\mathbf{a}_j$  is the  $j$ th column of  $\mathbf{A}$  and  $\beta'_j$  is the  $j$ th row of  $\mathbf{B}$ , an application of multiplying partitioned matrices (Section 2.8m) is

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_c \end{bmatrix} \begin{bmatrix} \beta'_1 \\ \beta'_2 \\ \vdots \\ \beta'_c \end{bmatrix} = \sum_{j=1}^c \mathbf{a}_j \beta'_j. \quad (11)$$

Thus  $\mathbf{AB}$  is the sum of outer products of columns of  $\mathbf{A}$  with corresponding rows in  $\mathbf{B}$ .

#### Example.

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} = \begin{bmatrix} 43 & 48 \\ 59 & 66 \\ 75 & 84 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 7 & 8 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \begin{bmatrix} 9 & 10 \end{bmatrix} = \begin{bmatrix} 7 & 8 \\ 14 & 16 \\ 21 & 24 \end{bmatrix} + \begin{bmatrix} 36 & 40 \\ 45 & 50 \\ 54 & 60 \end{bmatrix}. \end{aligned}$$

A special case of (11) is when  $\mathbf{B} = \mathbf{A}'$ :

$$\mathbf{AA}' = \sum_{j=1}^c \mathbf{a}_j \mathbf{a}'_j. \quad (12)$$

$\mathbf{AA}'$  is thus the sum of outer products of each column of  $\mathbf{A}$  with itself; e.g.,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} \begin{bmatrix} 2 & 5 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \end{bmatrix} \begin{bmatrix} 3 & 6 \end{bmatrix}.$$

#### e. Elementary vectors

A special case of (12) is

$$\mathbf{I}_n = \sum_{i=1}^n \mathbf{e}_i \mathbf{e}'_i$$

for  $\mathbf{e}_i$  being the  $i$ th column of  $\mathbf{I}_n$ , namely a vector with unity for its  $i$ th element and zeros elsewhere.  $\mathbf{e}_i$  is called an *elementary vector*. When necessary, its order  $n$  can be identified by denoting  $\mathbf{e}_i$  as  $\mathbf{e}_i^{(n)}$ .

The outer product of one elementary vector with another is a null matrix except for one element of unity: for example, with

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{E}_{12} = \mathbf{e}_1 \mathbf{e}'_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In general,  $\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}'_j$  is null except for element  $(i, j)$  being unity; and, of course,  $\mathbf{I} = \sum_i \mathbf{E}_{ii}$ . These  $\mathbf{E}_{ij}$ -matrices are particularly useful in applications of calculus to matrix algebra (see Section 12.9c).

The e-vectors are also useful for delineating individual rows and columns of a matrix. Thus  $\mathbf{e}'_i \mathbf{A} = \alpha'_i$ , the  $i$ th row of  $\mathbf{A}$  and  $\mathbf{A} \mathbf{e}_j = \mathbf{a}_j$ , the  $j$ th column of  $\mathbf{A}$ ; e.g., for

$$\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}, \quad \mathbf{e}'_2 \mathbf{A} = [5 \quad 7] \quad \text{and} \quad \mathbf{A} \mathbf{e}_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

#### f. Skew-symmetric matrices

A symmetric matrix  $\mathbf{A}$  has the property  $\mathbf{A} = \mathbf{A}'$ ; in contrast there are also matrices  $\mathbf{B}$  having the property  $\mathbf{B}' = -\mathbf{B}$ . Their diagonal elements are zero and each off-diagonal element is minus its symmetric partner; i.e.,  $b_{ii} = 0$  and  $b_{ij} = -b_{ji}$ . An example is

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 3 & -2 & 0 \end{bmatrix}.$$

Such matrices, having  $\mathbf{B}' = -\mathbf{B}$ , are called *skew-symmetric*.

## 2. MATRICES HAVING ALL ELEMENTS EQUAL

Vectors whose every element is unity are called *summing vectors* because they can be used to express a sum of numbers in matrix notation as an inner product.

**Example.** The row vector  $\mathbf{1}' = [1 \ 1 \ 1 \ 1]$  is the summing vector of order 4 and for  $\mathbf{x}' = [3 \ 6 \ 8 \ -2]$

$$\begin{aligned} \mathbf{1}'\mathbf{x} &= [1 \ 1 \ 1 \ 1] \begin{bmatrix} 3 \\ 6 \\ 8 \\ -2 \end{bmatrix} = 3 + 6 + 8 - 2 \\ &= 15 = [3 \ 6 \ 8 \ -2] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{x}'\mathbf{1}. \end{aligned}$$

When necessary to avoid confusion the order of a summing vector can be denoted in the usual way:  $\mathbf{1}'_4 = [1 \ 1 \ 1 \ 1]$ . For example,

$$\mathbf{1}'_3\mathbf{X} = [1 \ 1 \ 1] \begin{bmatrix} 2 & -1 \\ -5 & -3 \\ 4 & 5 \end{bmatrix} = [1 \ 1] = \mathbf{1}'_2.$$

Were this equation in  $\mathbf{X}$  to be written as  $\mathbf{1}'\mathbf{X} = \mathbf{1}'$ , one might be tempted to think that both  $\mathbf{1}'$  vectors had the same order. That they do not is made clear by denoting  $\mathbf{X}$  of order  $r \times c$  as  $\mathbf{X}_{r \times c}$  so that  $\mathbf{1}'\mathbf{X}_{r \times c} = \mathbf{1}'$  is  $\mathbf{1}'_r\mathbf{X}_{r \times c} = \mathbf{1}'_c$ .

The inner product of a summing vector with itself is a scalar, the vector's order:

$$\mathbf{1}'_n\mathbf{1}_n = n. \quad (13)$$

And outer products are matrices with all elements unity:

$$\mathbf{1}_3\mathbf{1}'_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [1 \ 1] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \mathbf{J}_{3 \times 2}.$$

In general,  $\mathbf{1}_r\mathbf{1}'_s$  has order  $r \times s$  and is often denoted by the symbol  $\mathbf{J}$  or  $\mathbf{J}_{r \times s}$ :

$$\mathbf{1}_r\mathbf{1}'_s = \mathbf{J}_{r \times s}, \text{ having all elements unity.} \quad (14)$$

Clearly,  $\lambda\mathbf{J}_{r \times s}$  has all elements  $\lambda$ .

Products of  $\mathbf{J}$ 's with each other and with  $\mathbf{1}$ 's are, respectively,  $\mathbf{J}$ 's and  $\mathbf{1}$ 's (multiplied by scalars):

$$\mathbf{J}_{r \times s}\mathbf{J}_{s \times t} = s\mathbf{J}_{r \times t}, \quad \mathbf{1}'_r\mathbf{J}_{r \times s} = r\mathbf{1}'_s \text{ and } \mathbf{J}_{r \times s}\mathbf{1}_s = s\mathbf{1}_r. \quad (15)$$

Particularly useful are square  $\mathbf{J}$ 's and a variant thereof:

$$\mathbf{J}_n = \mathbf{1}_n\mathbf{1}'_n \text{ with } \mathbf{J}_n^2 = n\mathbf{J}_n; \quad (16)$$

and

$$\bar{\mathbf{J}}_n = \frac{1}{n}\mathbf{J}_n \text{ with } \bar{\mathbf{J}}_n^2 = \bar{\mathbf{J}}_n.$$

And for statistics

$$\mathbf{C}_n = \mathbf{I} - \bar{\mathbf{J}}_n = \mathbf{I} - \frac{1}{n}\mathbf{J}_n, \quad (17)$$

known as a *centering matrix*, is especially useful, as is now illustrated. First observe that

$$\mathbf{C} = \mathbf{C}' = \mathbf{C}^2, \quad \mathbf{C}\mathbf{1} = \mathbf{0} \text{ and } \mathbf{C}\mathbf{J} = \mathbf{J}\mathbf{C} = \mathbf{0}, \quad (18)$$

which the reader can easily verify.

**Illustration.** The mean and sum of squares about the mean for data  $x_1, x_2, \dots, x_n$  can be expressed in terms of  $\mathbf{1}$ -vectors and  $\mathbf{J}$ -matrices. Define

$$\mathbf{x}' = [x_1 \ x_2 \ \dots \ x_n].$$

Then the mean of the  $x$ 's is

$$\bar{x} = (x_1 + x_2 + \dots + x_n)/n = \sum_{i=1}^n x_i/n = \frac{1}{n}\mathbf{x}'\mathbf{1} = \frac{1}{n}\mathbf{1}'\mathbf{x},$$

the last equality arising from  $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$  of Section 1c. And using  $\mathbf{C}$  of (17) and (18)

$$\mathbf{x}'\mathbf{C} = \mathbf{x}' - \mathbf{x}'\bar{\mathbf{J}} = \mathbf{x}' - \frac{1}{n}\mathbf{x}'\mathbf{1}\mathbf{1}' = \mathbf{x}' - \bar{x}\mathbf{1}' = \{x_i - \bar{x}\}$$

is the data vector with each observation expressed as a deviation from  $\bar{x}$ .

(This is the origin of the name centering matrix for  $C$ .) Postmultiplying  $x'C$  by  $x$  then gives

$$x'Cx = (x' - \bar{x}\mathbf{1}')x = x'x - \bar{x}(\mathbf{1}'x) = x'x - n\bar{x}^2.$$

Hence, using a standard result in statistics, we get

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = x'x - n\bar{x}^2 = x'Cx. \quad (19)$$

Thus for  $x'$  being a data vector,  $x'\mathbf{1}/n$  is the mean,  $x'C$  is the vector of deviations from the mean, and  $x'Cx$  is the sum of squares about the mean.

Expression (19) is a special case of the form  $x'Ax$ , known as a quadratic form (see Section 5), which can be used for sums of squares generally. Expressed in this manner, and with the aid of other matrix concepts (notably idempotency; see Section 3), sums of squares of normally distributed  $x$ 's are known to be distributed as  $\chi^2$  under very simply stated (in matrix notation) conditions (see Section 13.5).

**Illustration.** It was noted in Section 1.6 that row sums of a transition probability matrix are always unity. For example, with

$$P = \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix}, \quad \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \text{ i.e., } \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

This last result is

$$P\mathbf{1} = \mathbf{1},$$

which is true generally for any transition probability matrix. Furthermore, because  $P^2\mathbf{1} = P(P\mathbf{1}) = P\mathbf{1} = \mathbf{1}$  this in turn extends to

$$P^n\mathbf{1} = \mathbf{1},$$

showing that row sums of any power of a transition probability matrix are also unity. Furthermore, since the elements of a state vector (see Section 2.8e) add to unity,  $x'_0\mathbf{1} = 1$  and so do those of any state vector derived from it because  $x'_n\mathbf{1} = x'_0P^n\mathbf{1} = x'_0\mathbf{1} = 1$ .

### 3. IDEMPOTENT MATRICES

The matrix  $\bar{J}_n$  in (17) has the characteristic that its square equals itself. Many different matrices are of this nature; they are called idempotent matrices. Thus when  $K$  is such that  $K^2 = K$ , we say  $K$  is *idempotent* (from

Latin, *idem* meaning "same," and potent "power"). All idempotent matrices are square (otherwise  $K^2$  does not exist); identity matrices and square null matrices are idempotent. When  $K$  is idempotent, all powers of  $K$  equal  $K$ ; i.e.,  $K^r = K$  for  $r$  being a positive integer, and  $(I - K)$  is idempotent. Thus

$$K^2 = K \text{ implies } (I - K)^2 = I - K,$$

but  $K - I$  is not idempotent. A product of two idempotent matrices is idempotent if the matrices commute in multiplication.

Idempotent matrices occur in many applications of matrix algebra and they play an especially important role in statistics.

**Examples.** The matrix  $I - \bar{J}$  in (17) is idempotent; so is  $GA$  whenever  $G$  is such that  $AGA = A$  [because then  $(GA)^2 = GAGA = GA$ ]. A matrix  $G$  of this nature is called a generalized inverse of  $A$ ; its properties are considered in Chapter 8.

A matrix  $A$  satisfying  $A^2 = 0$  is called *nilpotent*, and that for which  $A^2 = I$  could be called *unipotent*.

**Examples.**

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \text{ is nilpotent; } B = \begin{bmatrix} I & X \\ 0 & -I \end{bmatrix} \text{ is unipotent.}$$

Variations on these definitions are  $A^k = A$ ,  $A^k = 0$ , and  $A^k = I$  for some positive integer  $k > 2$ . An example is the matrix

$$B = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix},$$

arising from characteristics of Bernadelli's (1941) beetles:  $B^3 = I$  but  $B^2 \neq I$ .

## 4. ORTHOGONAL MATRICES

### a. Definitions

Another useful class of matrices is that for which  $A$  has the property  $AA' = I = A'A$ . Such matrices are called *orthogonal*. We lead up to them with the following definitions.

The *norm* of a real vector  $x' = [x_1 \ x_2 \ \cdots \ x_n]$  is defined as

$$\text{norm of } x = \sqrt{x'x} = \left( \sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \quad (20)$$

For example, the norm of  $\mathbf{x}' = [1 \ 2 \ 2 \ 4]$  is  $(1 + 4 + 4 + 16)^{\frac{1}{2}} = 5$ . (The square root is taken as positive.) A vector is said to be either *normal* or a *unit vector* when its norm is unity; i.e., when  $\mathbf{x}'\mathbf{x} = 1$ . An example is  $\mathbf{x}' = [.2 \ .4 \ .4 \ .8]$ . Any non-null vector  $\mathbf{x}$  can be changed into a unit vector by multiplying it by the scalar  $1/\sqrt{\mathbf{x}'\mathbf{x}}$ ; i.e.,

$$\mathbf{u} = \left( \frac{1}{\sqrt{\mathbf{x}'\mathbf{x}}} \right) \mathbf{x}$$

is the *normalized* form of  $\mathbf{x}$  (because  $\mathbf{u}'\mathbf{u} = 1$ ).

Non-null vectors  $\mathbf{x}$  and  $\mathbf{y}$  are described as being *orthogonal* when  $\mathbf{x}'\mathbf{y} = 0$  (equivalent, of course, to  $\mathbf{y}'\mathbf{x} = 0$ ); e.g.,  $\mathbf{x}' = [1 \ 2 \ 2 \ 4]$  and  $\mathbf{y}' = [6 \ 3 \ -2 \ -2]$  are orthogonal vectors because

$$\mathbf{x}'\mathbf{y} = [1 \ 2 \ 2 \ 4] \begin{bmatrix} 6 \\ 3 \\ -2 \\ -2 \end{bmatrix} = 6 + 6 - 4 - 8 = 0.$$

Two vectors are defined as *orthonormal vectors* when they are orthogonal and normal. Thus  $\mathbf{u}$  and  $\mathbf{v}$  are orthonormal when  $\mathbf{u}'\mathbf{u} = 1 = \mathbf{v}'\mathbf{v}$  and  $\mathbf{u}'\mathbf{v} = 0$ ; for example,  $\mathbf{u}' = \frac{1}{5}[1 \ 1 \ 3 \ 3 \ 4]$  and  $\mathbf{v}' = [-.1 \ -.9 \ -.1 \ -.1 \ .4]$  are orthonormal vectors.

A group, or collection, of vectors all of the same order is called a *set* of vectors. A set of vectors  $\mathbf{x}_i$ , for  $i = 1, 2, \dots, n$  is said to be an *orthonormal set* of vectors when every vector in the set is normal,  $\mathbf{x}'_i\mathbf{x}_i = 1$  for all  $i$ , and when every pair of different vectors in the set is orthogonal,  $\mathbf{x}'_i\mathbf{x}_j = 0$  for  $i \neq j = 1, 2, \dots, n$ . We can say that the vectors of an orthonormal set are all normal, and pairwise orthogonal.

A matrix  $\mathbf{P}_{r \times c}$  whose rows constitute an orthonormal set of vectors is said to have orthonormal rows, whereupon  $\mathbf{P}\mathbf{P}' = \mathbf{I}_r$ . But then  $\mathbf{P}'\mathbf{P}$  is not necessarily an identity matrix  $\mathbf{I}_c$ , as the following example shows:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}\mathbf{P}' = \mathbf{I}_2 \quad \text{but} \quad \mathbf{P}'\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq \mathbf{I}_3.$$

Conversely, when  $\mathbf{P}_{r \times c}$  has orthonormal columns  $\mathbf{P}'\mathbf{P} = \mathbf{I}_c$  but  $\mathbf{P}\mathbf{P}'$  may not be an identity matrix.

Square matrices having orthonormal rows are in a special class: their columns are also orthonormal. (This is proved in Section 5.7d.) The matrix

$\mathbf{P}$ , say, is then such that

$$\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}. \quad (21)$$

This equation defines  $\mathbf{P}$  as being an *orthogonal matrix*; it implies that  $\mathbf{P}$  is square and that  $\mathbf{P}$  has orthonormal rows and orthonormal columns. These are characteristics of any orthogonal matrix  $\mathbf{P}$ . Actually, as is shown in Sec. 5.7d, any two of the conditions (i)  $\mathbf{P}$  square, (ii)  $\mathbf{P}'\mathbf{P} = \mathbf{I}$  and (iii)  $\mathbf{P}\mathbf{P}' = \mathbf{I}$  implies the third, i.e., implies that  $\mathbf{P}$  is orthogonal.

A simple property of orthogonal matrices is that products of them are orthogonal. (See Exercise 13.)

**Example.**

$$\mathbf{A} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & -\sqrt{3} & 0 \\ 1 & 1 & -2 \end{bmatrix}$$

is an orthogonal matrix, as the reader may easily verify.

The boundless variety of matrices that are orthogonal includes many that are carefully proscribed, three of which are now briefly illustrated and described.

#### b. Special cases

Helmert, Givens and Householder matrices are all orthogonal matrices.

(i) *Helmert matrices.* The Helmert matrix of order 4 is

$$\mathbf{H}_4 = \begin{bmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix}.$$

The Helmert matrix  $\mathbf{H}_n$  of order  $n$  has  $n^{-\frac{1}{2}}\mathbf{1}'_n$  for its first row, and each of its other  $n - 1$  rows for  $i = 1, \dots, n - 1$ , has the partitioned form

$$[\mathbf{1}'_i \mid -i \mid 0\mathbf{1}'_{n-i-1}] / \sqrt{\lambda_i} \quad \text{with } \lambda_i = i(i + 1).$$

The notation  $0\mathbf{1}'$  used here simply indicates a row vector with every element zero.



An important use for Helmert matrices is in statistics for partitioning a total sum of squares  $\sum_{i=1}^n x_i^2 = \mathbf{x}'\mathbf{x}$  for  $\mathbf{x}' = [x_1 \dots x_n]$  (see Exercise 20 and Chapter 6).

(ii) *Givens matrices.* The orthogonal matrix

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

is a Givens matrix of order 2. It is the basis of Givens matrices of order higher than 2. Those of order 3 are

$$\mathbf{G}_{12} = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{G}_{13} = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix}$$

and

$$\mathbf{G}_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma & \sin \gamma \\ 0 & -\sin \gamma & \cos \gamma \end{bmatrix}. \quad (22)$$

The general form of  $\mathbf{G}_{rs} = \mathbf{G}_{sr}$  of order  $n$  is an identity matrix except for four elements:  $g_{rr} = g_{ss} = \cos \theta_{rs}$  and, for  $r > s$ ,  $-g_{rs} = g_{sr} = \sin \theta_{rs}$ . All such matrices are orthogonal, as are products of any number of them.

A sometimes useful operation is that of triangularizing a square matrix; i.e., if  $\mathbf{A}$  is square, of premultiplying it by some matrix  $\mathbf{G}$  so that  $\mathbf{GA}$  is an upper triangular matrix. This can be done by deriving  $\mathbf{G}$  as a product of Givens matrices (see Exercise 21). The operation of going from  $\mathbf{A}$  to  $\mathbf{GA}$  is called a *Givens transformation*. A lower triangular matrix can be obtained from  $\mathbf{A}$  in similar fashion, by postmultiplying  $\mathbf{A}$  by Givens matrices.

(iii) *Householder matrices.* Other orthogonal matrices useful for triangularizing square matrices are the Householder matrices, the general form of which is

$$\mathbf{H} = \mathbf{I} - 2\mathbf{h}\mathbf{h}' \quad \text{for } \mathbf{h}'\mathbf{h} = 1, \quad (23)$$

with  $\mathbf{h}$ , obviously, being a non-null column vector. Then  $\mathbf{H}$  is not only orthogonal but also symmetric; and for any non-null vector

$$\mathbf{x}' = [x_1 \quad x_2 \quad \dots \quad x_n] \quad \text{there exists } \mathbf{h}' = [h_1 \quad h_2 \quad \dots \quad h_n]$$

such that

$$\mathbf{H}\mathbf{x} = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{for } \begin{cases} \lambda = -(\text{sign of } x_1)\sqrt{\mathbf{x}'\mathbf{x}} \\ h_1 = \sqrt{\frac{1}{2}(1 - x_1/\lambda)} \\ h_i = -x_i/2h_1\lambda \quad \text{for } i = 2, 3, \dots, n. \end{cases}$$

**Example.** For  $\mathbf{x}' = [1 \quad 2 \quad 3 \quad 5 \quad 5]$  we find  $\mathbf{h}' = \frac{1}{12}[9 \quad 2 \quad 3 \quad 5 \quad 5]$  and  $\mathbf{H}\mathbf{x} = [-8 \quad 0 \quad 0 \quad 0 \quad 0]'$ .

Triangularization of a square matrix  $\mathbf{A}$  proceeds by developing Householder matrices from successive columns of  $\mathbf{A}$  (see Exercise 22).

## 5. QUADRATIC FORMS

The illustration (19),

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \mathbf{x}'\mathbf{C}\mathbf{x},$$

is the product of a row vector  $\mathbf{x}'$ , a matrix  $\mathbf{C}$  and the column vector  $\mathbf{x}$ . It is called a *quadratic form*; its general form for any matrix  $\mathbf{A}$  is  $\mathbf{x}'\mathbf{A}\mathbf{x}$ . Expressions of this form have many uses, particularly in the general theory of analysis of variance in statistics wherein, with appropriate choice of  $\mathbf{A}$ , any sum of squares can be represented as  $\mathbf{x}'\mathbf{A}\mathbf{x}$ .

Consider the example

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Straightforward multiplication gives

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= x_1^2 + 4x_2x_1 + 2x_3x_1 + 2x_1x_2 \\ &\quad + 7x_2^2 - 2x_3x_2 + 3x_1x_3 + 6x_2x_3 + 5x_3^2, \end{aligned} \quad (24)$$

which simplifies to  $\mathbf{x}'\mathbf{A}\mathbf{x} =$

$$x_1^2 + x_1x_2(4 + 2) + x_1x_3(2 + 3) + 7x_2^2 + x_2x_3(-2 + 6) + 5x_3^2 \quad (25)$$

$$= x_1^2 + 7x_2^2 + 5x_3^2 + 6x_1x_2 + 5x_1x_3 + 4x_2x_3. \quad (26)$$

This is a quadratic function of the  $x$ 's; hence the name quadratic form. Two characteristics of its development are noteworthy. First, in (24) we see that  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is the sum of products of all possible pairs of the  $x_i$ 's, each multiplied by an element of  $\mathbf{A}$ ; thus in (24) the second term,  $4x_2x_1$ , is  $x_2x_1$  multiplied by the element of  $\mathbf{A}$  in the second row and first column. Second, in simplifying (24) to (26) we see in (25) that the coefficient of  $x_1x_2$ , for example, is the sum of two elements in  $\mathbf{A}$ : the one in the first column and second row plus that in the second column and first row. These results are true generally.

If  $\mathbf{x}$  is a vector of order  $n$  with elements  $x_i$  for  $i = 1, 2, \dots, n$ , and if  $\mathbf{A}$  is a square matrix of order  $n$  with elements  $a_{ij}$  for  $i, j = 1, 2, \dots, n$ , then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i \sum_j x_i x_j a_{ij} \quad [\text{similar to (24)}]$$

$$= \sum_i x_i^2 a_{ii} + \sum_{i \neq j} x_i x_j a_{ij}$$

and as in (25) and (26) this is

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + \sum_{j>i} x_i x_j (a_{ij} + a_{ji}). \quad (27)$$

Thus  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is the sum of squares of the elements of  $\mathbf{x}$ , each square multiplied by the corresponding diagonal element of  $\mathbf{A}$ , plus the sum of products of the elements of  $\mathbf{x}$ , each product multiplied by the sum of the corresponding elements of  $\mathbf{A}$ ; i.e., the product of the  $i$ th and  $j$ th element of  $\mathbf{x}$  is multiplied by  $(a_{ij} + a_{ji})$ .

Returning to the example, notice that just as (26) equals (25) so also does

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= x_1^2 + 7x_2^2 + 5x_3^2 + 6x_1x_2 + 5x_1x_3 + 4x_2x_3 \\ &= x_1^2 + 7x_2^2 + 5x_3^2 + x_1x_2(1 + 5) + x_1x_3(1 + 4) + x_2x_3(0 + 4) \end{aligned}$$

$$= \mathbf{x}' \begin{bmatrix} 1 & 1 & 1 \\ 5 & 7 & 0 \\ 4 & 4 & 5 \end{bmatrix} \mathbf{x}.$$

In this way we see that

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}' \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix} \mathbf{x} \text{ is the same as } \mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{x}' \begin{bmatrix} 1 & 1 & 1 \\ 5 & 7 & 0 \\ 4 & 4 & 5 \end{bmatrix} \mathbf{x},$$

where  $\mathbf{B}$  is different from  $\mathbf{A}$ . Note that the quadratic form is the same, even though the associated matrix is not the same. In fact, there is no unique matrix  $\mathbf{A}$  for which any particular quadratic form can be expressed as  $\mathbf{x}'\mathbf{A}\mathbf{x}$ . Many matrices can be so used. Each one has the same diagonal elements, and in each of them the sum of each pair of symmetrically placed off-diagonal elements  $a_{ij}$  and  $a_{ji}$  must be the same; for example, (24) can also be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}' \begin{bmatrix} 1 & 2342 & -789 \\ -2336 & 7 & 1.37 \\ 794 & 2.63 & 5 \end{bmatrix} \mathbf{x}. \quad (28)$$

In particular, if we write

$$\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + 7x_2^2 + 5x_3^2 + x_1x_2(3 + 3) + x_1x_3(2\frac{1}{2} + 2\frac{1}{2}) + x_2x_3(2 + 2)$$

we see that it can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}' \begin{bmatrix} 1 & 3 & 2\frac{1}{2} \\ 3 & 7 & 2 \\ 2\frac{1}{2} & 2 & 5 \end{bmatrix} \mathbf{x} \quad (29)$$

where  $\mathbf{A}$  is now a symmetric matrix. As such it is unique; that is to say, for any particular quadratic form there is a unique symmetric matrix  $\mathbf{A}$  for which the quadratic form can be expressed as  $\mathbf{x}'\mathbf{A}\mathbf{x}$ . It can be found in any particular case by rewriting the quadratic  $\mathbf{x}'\mathbf{A}\mathbf{x}$  where  $\mathbf{A}$  is not symmetric as  $\mathbf{x}'[\frac{1}{2}(\mathbf{A} + \mathbf{A}')] \mathbf{x}$ , because  $\frac{1}{2}(\mathbf{A} + \mathbf{A}')$  is symmetric. For example, if  $\mathbf{A}$  is the matrix used in (28), it is easily observed that  $\frac{1}{2}(\mathbf{A} + \mathbf{A}')$  is the symmetric matrix used in (29).

Taking  $\mathbf{A}$  as symmetric with  $a_{ij} = a_{ji}$ , we see from (27) that the quadratic form  $\mathbf{x}'\mathbf{A}\mathbf{x}$  can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + 2 \sum_{j>i} x_i x_j a_{ij}. \quad (30)$$

For example,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}' \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \mathbf{x}$$

$$= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2(a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3).$$

The importance of the symmetric  $\mathbf{A}$  is that, when writing any particular quadratic function as  $\mathbf{x}'\mathbf{A}\mathbf{x}$ , there are many different matrices that can be used, but there is only one symmetric matrix—and it is unique for that particular function. Therefore, when dealing with quadratic forms  $\mathbf{x}'\mathbf{A}\mathbf{x}$ , we always take  $\mathbf{A}$  as symmetric. This is convenient not only because the symmetric  $\mathbf{A}$  is unique for any particular quadratic form, but also because symmetric matrices have many properties that are useful in studying quadratic forms, particularly those associated with analysis of variance. Hereafter, whenever we deal with a quadratic form  $\mathbf{x}'\mathbf{A}\mathbf{x}$ , we assume  $\mathbf{A} = \mathbf{A}'$ .

A slightly more general (but not so useful) function is the second-degree function in two sets of variables  $\mathbf{x}$  and  $\mathbf{y}$ , say. For example,

$$\mathbf{x}'\mathbf{M}\mathbf{y} = [x_1 \quad x_2] \begin{bmatrix} 2 & 4 & 3 \\ 7 & 6 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= 2x_1y_1 + 4x_1y_2 + 3x_1y_3 + 7x_2y_1 + 6x_2y_2 + 5x_2y_3.$$

It is called a *bilinear form* and, as illustrated here, its matrix  $\mathbf{M}$  does not have to be square as does the matrix in a quadratic form. Clearly, quadratic forms are special cases of bilinear forms—when  $\mathbf{M}$  is square and  $\mathbf{y} = \mathbf{x}$ .

## 6. POSITIVE DEFINITE MATRICES

All quadratic forms  $\mathbf{x}'\mathbf{A}\mathbf{x}$  are zero for  $\mathbf{x} = \mathbf{0}$ . For some matrices  $\mathbf{A}$  the corresponding quadratic form is zero *only* for  $\mathbf{x} = \mathbf{0}$ . For example,

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3 \\ &= (x_1 + 2x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2, \end{aligned}$$

and, by the nature of this last expression, we see that (for elements of  $\mathbf{x}$  being real numbers) it is positive unless all elements of  $\mathbf{x}$  are zero, i.e.,  $\mathbf{x} = \mathbf{0}$ . Such a quadratic form is described as being positive definite. More formally,

when  $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$  for all  $\mathbf{x}$  other than  $\mathbf{x} = \mathbf{0}$ ,

then  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is a *positive definite* quadratic form, and  $\mathbf{A} = \mathbf{A}'$  is correspondingly a *positive definite* (p.d.) matrix.

There are also symmetric matrices  $\mathbf{A}$  for which  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is zero for some non-null  $\mathbf{x}$  as well as for  $\mathbf{x} = \mathbf{0}$ ; e.g.,

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_1x_3 - 6x_2x_3 \\ &= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2. \end{aligned}$$

This is zero for  $\mathbf{x}' = [2 \quad 1 \quad 3]$ , and for any scalar multiple thereof, as well as for  $\mathbf{x} = \mathbf{0}$ . This kind of quadratic form is called positive semidefinite and has for its formal definition:

when  $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$  for all  $\mathbf{x}$  and  $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$  for some  $\mathbf{x} \neq \mathbf{0}$

then  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is a *positive semidefinite* quadratic form and hence  $\mathbf{A} = \mathbf{A}'$  is a *positive semidefinite* (p.s.d.) matrix. The two classes of (forms and) matrices taken together, positive definite and positive semidefinite, are called *non-negative definite* (n.n.d.).

**Illustration.** The sum of squares of (19),

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \mathbf{x}'\mathbf{C}\mathbf{x},$$

is a positive semidefinite quadratic form because it is positive, except for being zero when all the  $x_i$ 's are equal. Its matrix,  $\mathbf{I} - \mathbf{J}$ , which is idempotent, is also p.s.d., as are all symmetric idempotent matrices (except  $\mathbf{I}$ , which is the only p.d. idempotent matrix).

Unfortunately there is no universal agreement on the definition of positive semidefinite. Most writers use it with the meaning defined here, but

some use it in the sense of meaning non-negative definite. But on one convention there is universal agreement: p.d., p.s.d., and n.n.d. matrices are always taken as being symmetric. This is so because the definitions of these matrices are in terms of quadratic forms which can always be expressed utilizing symmetric matrices.

## EXERCISES

- Show that  $\begin{bmatrix} 3 & 8 & 4 \\ 8 & 7 & -1 \\ 4 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & 4 \\ 3 & 4 & 6 \end{bmatrix}$  is symmetric.
- If  $\mathbf{x}$  and  $\mathbf{y}$  are  $n \times 1$  column vectors and  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  matrices, which of the following expressions are undefined? Of those that are defined, which are outer products, bilinear forms, or quadratic forms?
  - $\mathbf{y} = \mathbf{A}\mathbf{x}$ .
  - $\mathbf{x}'\mathbf{B}\mathbf{x}$ .
  - $\mathbf{y}'\mathbf{B}'\mathbf{A}\mathbf{x}$ .
  - $\mathbf{x}'\mathbf{A}\mathbf{y}$ .
  - $\mathbf{xy}' = \mathbf{B}'$ .
  - $\mathbf{xy} = \mathbf{A}'\mathbf{B}$ .
  - $\mathbf{yBx}$ .
  - $\mathbf{x}' = \mathbf{y}'\mathbf{B}'$ .
  - $\mathbf{y}'\mathbf{A}'\mathbf{B}\mathbf{y}$ .
  - $\mathbf{y} + \begin{bmatrix} 30 \\ 40 \end{bmatrix} = \mathbf{AB}'\mathbf{x}$  for  $n = 2$ .
- With  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$ ,  $\mathbf{B} = \begin{bmatrix} 3 & 6 \\ 4 & 8 \end{bmatrix}$ ,  $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$  and  $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$  in Exercise 2
  - check the validity of those equations that are defined;
  - calculate the value of the other defined expressions.
- If  $\mathbf{A}$  is skew symmetric, prove that (a)  $a_{ii} = 0$  and  $a_{ij} = -a_{ji}$ ; and (b)  $\mathbf{I} + \mathbf{A}$  is p.d.
- Write down  $\mathbf{M} = \{i(j+1)\}$  for  $i, j = 1, 2, 3$ .
  - Show that  $\mathbf{M}$  is the outer product of the vectors  $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}$ . What is the inner product?
  - Calculate  $\mathbf{M} + \mathbf{M}'$ .
  - What characteristic must a matrix  $\mathbf{T}$  have so that  $\mathbf{T} + \mathbf{T}'$  exists?
  - Prove that  $\mathbf{T} + \mathbf{T}'$  is symmetric, but  $\mathbf{T} - \mathbf{T}'$  is skew-symmetric.
  - For square  $\mathbf{A}$ , prove it is the sum of a symmetric and a skew-symmetric matrix.
- If the product of two symmetric matrices is symmetric, prove that the matrices commute in multiplication.
- Show that if  $\mathbf{X}'\mathbf{X} = \mathbf{X}$ , then  $\mathbf{X} = \mathbf{X}' = \mathbf{X}^2$ .

- Show that the only real symmetric matrix whose square is null is the null matrix itself.
  - Explain why  $\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$  implies  $\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}$ .
- If  $\mathbf{X}_{r \times c} = \{\mathbf{x}_j\}$  for  $j = 1, \dots, c$ , prove for symmetric  $\mathbf{A}$  that  $\text{tr}(\mathbf{A}\mathbf{X}\mathbf{X}')^2 = \sum_{j=1}^c \sum_{k=1}^c (\mathbf{x}'_j \mathbf{A} \mathbf{x}_k)^2$ .
- A matrix  $\mathbf{X}$  arising in the design of certain experiments can be described as follows: every element of the first column of  $\mathbf{X}$  is unity, and all other elements are either  $+1$  or  $-1$ . Then  $\mathbf{X}$  can be changed to being another matrix  $\mathbf{Z}$  whose every element is either 0 or 1 by the following operations: add  $\mathbf{X}$ 's first column to all its other columns, then double the first column and halve the resulting matrix. The result is  $\mathbf{Z}$ .
  - Make up a small example of  $\mathbf{X}$  and derive  $\mathbf{Z}$ .
  - Express  $\mathbf{Z}$  as a product of  $\mathbf{X}$  and other matrices (involving  $\mathbf{I}$  and  $\mathbf{I}$ ).
  - If  $\mathbf{Y}$  is the submatrix of  $\mathbf{X}$  other than its first column, express  $\mathbf{Z}$  as a function of  $\mathbf{Y}$ .
- For  $\mathbf{I}$  and  $\mathbf{J}$  of order  $n$ , and for  $p, q, r$  and  $s$  being scalars with  $p \neq 0$  and  $p + qn \neq 0$ :
  - Write down  $p\mathbf{I} + q\mathbf{J}$  of order 4.
  - Show that  $(p\mathbf{I} + q\mathbf{J})(r\mathbf{I} + s\mathbf{J}) = pr\mathbf{I} + (ps + qr + nqs)\mathbf{J}$ .
  - Simplify  $(p\mathbf{I} + q\mathbf{J})\frac{1}{p}\left(\mathbf{I} - \frac{q}{p + qn}\mathbf{J}\right)$ .
  - Explain why  $\mathbf{J}\mathbf{X}\mathbf{J} = \mathbf{x}_n\mathbf{J}$ .
  - Prove that  $\mathbf{I} - \bar{\mathbf{J}}_n$  is idempotent, but  $\mathbf{I} - \mathbf{J}$  is not.
- Suppose the probabilities of members of three political parties maintaining or changing their affiliation during a year are represented by the probability transition matrix
 
$$\mathbf{P} = \begin{bmatrix} .4 & .3 & .3 \\ .3 & .6 & .1 \\ .5 & .1 & .4 \end{bmatrix}.$$

The 1000 members of a country club are divided in their political leanings so that 600 support the first party, 300 the second and 100 the third. Let  $\mathbf{x}'_0 = [600 \ 300 \ 100]$  represent these political affiliations. The expected political affiliations one year from now will be  $\mathbf{x}'_0\mathbf{P}$ .

  - Calculate  $\mathbf{x}'_0\mathbf{P}$ .

- (b) Calculate the expected political affiliations of the 1000 country club members 2, 3 and 4 years from now.
- (c) Observe that the elements in each row of  $\mathbf{P}$  add to unity. Express this in a matrix (or vector) equation, making use of  $\mathbf{1}_3$  to do so.
- (d) Prove that the elements of each row of  $\mathbf{P}^k$  also add to unity, where  $\mathbf{P}^k$  is the  $k$ th power of  $\mathbf{P}$  and  $k$  is a positive integer.
- (e) Prove that all elements of  $\mathbf{x}'_0 \mathbf{P}^k$  sum to 1000.
- (f) Find the steady-state vector corresponding to  $\mathbf{P}$ . (See Exc. 2.24)
13. Prove that a product of orthogonal matrices is orthogonal.
14. (a) If  $\mathbf{A}$  is idempotent and symmetric, prove that it is non-negative definite.  
 (b) When  $\mathbf{X}$  and  $\mathbf{Y}$  are idempotent, prove that  $\mathbf{XY}$  is, provided  $\mathbf{X}$  and  $\mathbf{Y}$  commute in multiplication.  
 (c) Prove that  $\mathbf{I} + \mathbf{KK}'$  is p.d., for real  $\mathbf{K}$ .
15. Using  $\mathbf{x}' = [1 \ 3 \ 5 \ 7 \ 9]$ , derive or state the numerical value of  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  such that  
 (a)  $1^2 + 3^2 + 5^2 + 7^2 + 9^2 = \mathbf{x}'\mathbf{A}\mathbf{x}$ ;  
 (b)  $(1 + 3 + 5 + 7 + 9)^2/5 = \mathbf{x}'\mathbf{B}\mathbf{x}$ ;  
 (c)  $(1 - 5)^2 + (3 - 5)^2 + (5 - 5)^2 + (7 - 5)^2 + (9 - 5)^2 = \mathbf{x}'\mathbf{C}\mathbf{x}$ .
16. Find  $\mathbf{A} = \mathbf{A}'$  and one matrix  $\mathbf{B} \neq \mathbf{B}'$  for each of the following quadratic forms such that they can be written as  $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$ :  
 (a)  $7x_1^2 + 28x_1x_2 + 4x_2^2$ ;  
 (b)  $(2x_1 - 3x_2)^2 + (x_2 - 2x_3)^2 + (4x_1 - 12x_3)^2$ ;  
 (c)  $2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3$ ;  
 (d)  $2x_1^2 + 4x_1x_2 + 8x_2^2 + 6x_1x_3 + 18x_3^2 + 12x_2x_3$ .
17. Suppose  $\mathbf{Y} = \{y_{ij}\}$  for  $i = 1, \dots, p$  and  $j = 1, \dots, q$  represents experimental data for which the following averages are calculated.

$$\bar{y}_i = \sum_{j=1}^q y_{ij}/q, \quad \bar{y}_j = \sum_{i=1}^p y_{ij}/p \quad \text{and} \quad \bar{y}_{..} = \sum_{i=1}^p \sum_{j=1}^q y_{ij}/pq.$$

Define

$$\mathbf{y}' = [y_{11} \ y_{12} \dots y_{1q} \ y_{21} \ y_{22} \dots y_{2q} \dots y_{i1} \dots y_{iq} \dots y_{p1} \dots y_{pq}],$$

$$u_{ij} = y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..},$$

and

$$\mathbf{u}' = [u_{11} \ u_{12} \dots u_{1q} \ u_{21} \ u_{22} \dots u_{2q} \dots u_{i1} \dots u_{iq} \dots u_{p1} \dots u_{pq}].$$

Derive the following matrices:

- (a)  $\mathbf{A}$  such that  $\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_i \sum_j (\bar{y}_i - \bar{y}_{..})^2$ ;  
 (b)  $\mathbf{B}$  such that  $\mathbf{y}'\mathbf{B}\mathbf{y} = \sum_i \sum_j (\bar{y}_j - \bar{y}_{..})^2$ ;  
 (c)  $\mathbf{C}$  such that  $\mathbf{C}\mathbf{y} = \mathbf{u}$ ;  
 (d)  $\mathbf{D}$  such that  $\mathbf{y}'\mathbf{D}\mathbf{y} = \sum_i \sum_j u_{ij}^2$ .

First make the derivations for  $p = 3$  and  $q = 4$  (or  $p = 2, q = 3$ ) and then infer the results for general  $p$  and  $q$ .

18. Explain why  
 (a)  $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}'\mathbf{x}$ , even when  $\mathbf{A}$  is not symmetric;  
 (b)  $\mathbf{x}'\mathbf{B}\mathbf{x} = \text{tr}(\mathbf{x}'\mathbf{B}\mathbf{x})$ ;  
 (c)  $\mathbf{x}'\mathbf{C}\mathbf{x} = \text{tr}(\mathbf{C}\mathbf{x}\mathbf{x}')$ .
19. (a) Prove that  $\mathbf{A}'\mathbf{A}$  is the sum of outer products of rows of  $\mathbf{A}$  with themselves.  
 (b) Use  $\mathbf{I} = \sum \mathbf{e}_i \mathbf{e}_i'$  of Section 1e to prove (11).
20. (Helmert matrices). The Helmert matrix of order  $n$  is described in Section 4b(i). Partition it as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}' \\ \mathbf{K}_n \end{bmatrix}$$

where  $\mathbf{h}' = n^{-1/2} \mathbf{1}'$  is the first row and  $\mathbf{K}_n$  represents the last  $n - 1$  rows.

- (a) Prove that  $\mathbf{H}\mathbf{H}' = \mathbf{I}_n$  (and therefore, because  $\mathbf{H}$  is square, it is orthogonal).  
 (b) For  $\mathbf{x}$  of order  $n$ , prove that  $SS_n = \sum_{i=1}^n (x_i - \sum_{i=1}^n x_i/n)^2$  can be expressed as

$$SS_n = \mathbf{x}'\mathbf{K}'_n \mathbf{K}_n \mathbf{x}.$$

[Hint:  $\sum x_i^2 = \mathbf{x}'\mathbf{H}'\mathbf{H}\mathbf{x}$  and  $n\bar{x}^2 = \mathbf{x}'\mathbf{h}\mathbf{h}'\mathbf{x}$ .]

- (c) Establish the recursive formula

$$SS_n = SS_{n-1} + (1 - 1/n)(\bar{x}_{n-1} - x_n)^2$$

where  $\bar{x}_n$  is the mean of all  $n$   $x_i$ 's and  $\bar{x}_{n-1}$  is the mean of

$x_1, x_2, \dots, x_{n-1}$ . [Hint: the first  $n-2$  rows of  $\mathbf{K}_n$  contain the rows of  $\mathbf{K}_{n-1}$ .] Check this recursive formula for

$$\mathbf{x}' = [6 \ 8 \ 4 \ 10 \ 17].$$

21. (Givens matrices).

(a) Use Givens matrices to triangularize

$$\mathbf{A} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}$$

by the following steps:

(i)  $x_{12} = \sqrt{a_1^2 + a_2^2}$ ,  $\cos \alpha = a_1/x_{12}$ ,

$$\mathbf{G}_{12}[a_1 \ a_2 \ a_3]' = [x_{12} \ 0 \ a_3]'$$

(ii)  $x_{13} = \sqrt{x_{12}^2 + a_3^2}$ ,  $\cos \beta = x_{12}/x_{13}$ ,

$$\mathbf{G}_{13}[x_{12} \ 0 \ a_3]' = [x_{13} \ 0 \ 0]'$$

(iii) Define  $\mathbf{G}_{13}\mathbf{G}_{12}[b_1 \ b_2 \ b_3]' \equiv [d_1 \ d_2 \ d_3]'$ .

(iv)  $x_{23} = \sqrt{d_2^2 + d_3^2}$ ,  $\cos \gamma = d_2/x_{23}$ .

(v)  $\mathbf{G}_{23}\mathbf{G}_{13}\mathbf{G}_{12}\mathbf{A} = \begin{bmatrix} x_{13} & d_1 & f_1 \\ 0 & x_{23} & f_2 \\ 0 & 0 & f_3 \end{bmatrix}$  for appropriate values  $f_1$ ,  $f_2$  and  $f_3$ .

(b) Explain why  $f_1^2 + f_2^2 + f_3^2 = c_1^2 + c_2^2 + c_3^2$ .

(c) Suppose  $\mathbf{GA}$  is the triangular form of  $\mathbf{A}$  obtained by using  $\mathbf{G}$  as a product of Givens matrices. Let  $\mathbf{y}$  be the column of  $\mathbf{GA}$  corresponding to the column  $\mathbf{a}$  of  $\mathbf{A}$ . Explain why  $\mathbf{y}'\mathbf{y} = \mathbf{a}'\mathbf{a}$ , and why this is true for every column  $\mathbf{a}$  of  $\mathbf{A}$ .

(d) Triangularize  $\begin{bmatrix} 3 & 0 & 10 \\ 4 & 20 & 29 \\ 12 & 15 & 16 \end{bmatrix}$  to  $\begin{bmatrix} 13 & 20 & 26 \\ 0 & 15 & 20 \\ 0 & 0 & -11 \end{bmatrix}$  using Givens matrices.

22. (Householder matrices). For the vector  $\mathbf{h}$  and matrix  $\mathbf{H}$  defined in Section 4b(iii), assume that the first element of  $\mathbf{x}$  is positive and

prove that

(a)  $\mathbf{h}$  is normal;

(b)  $\mathbf{H}$  is symmetric and orthogonal;

(c)  $\mathbf{h}_1\lambda + \mathbf{h}'\mathbf{x} = 0$ ;

(d)  $\mathbf{H}\mathbf{x} = [\lambda \ 0]'$ .

(e) Triangularize  $\mathbf{A} = \begin{bmatrix} 1 & -3 & 88 \\ 2 & 4 & 1 \\ 2 & 5 & 0 \end{bmatrix}$  using Householder matrices in the following steps:

(i)  $\mathbf{H}_1$  derived from  $\mathbf{x}'_1 = [1 \ 2 \ 2]$ .

(ii)  $\mathbf{H}_1\mathbf{A} = \begin{bmatrix} -3 & -5 & -30 \\ 0 & 3 & -58 \\ 0 & 4 & -59 \end{bmatrix}$ .

(iii)  $\mathbf{H}_2^*$  derived from  $\mathbf{x}'_2 = [3 \ 4]$ .

(iv)  $\mathbf{H}_2\mathbf{H}_1\mathbf{A} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^* \end{bmatrix} \mathbf{H}_1\mathbf{A} = \begin{bmatrix} -3 & -5 & -30 \\ 0 & -5 & 82 \\ 0 & 0 & 11 \end{bmatrix}$ .

23. (a) For any transition probability matrix  $\mathbf{P}$  and  $\mathbf{A} = \mathbf{P} - \bar{\mathbf{J}}$ , prove that  $\mathbf{A}^n = \mathbf{A}\mathbf{P}^{n-1}$  for  $n$  a positive integer.

(b) Explain why  $\text{tr}(\mathbf{A}\mathbf{B}') = \mathbf{1}'(\mathbf{A} \cdot \mathbf{B})\mathbf{1}$ .

(c) For  $\mathbf{X}$  symmetric and idempotent and  $\mathbf{TX}$  symmetric, prove that  $\mathbf{TX} = \mathbf{XTX}$ .

24. A transition probability matrix which has both its row sums and its column sums being unity is called doubly stochastic. Prove that powers of such a matrix are also doubly stochastic.

## CHAPTER 4

# DETERMINANTS

Many of the operations associated with matrices have now been introduced. One operation not yet considered—division, or matrix inversion as it is called—is the subject of Chapter 5. As a prelude to that we discuss calculation of a scalar value known as the *determinant* of a (square) matrix. Knowledge of this calculation is necessary for our discussion of matrix inversion and it is also useful in succeeding chapters for certain aspects of matrix theory proper that are used in applied problems. Moreover, since readers will undoubtedly encounter determinants elsewhere in their readings on matrices, it is appropriate to have an introduction to them here.

The literature of determinants is extensive and forms part of many texts on matrices. The presentation in this book is relatively brief, deals with elementary methods of evaluation and discusses selected topics arising therefrom. Extensive use is made of small numerical examples rather than rigorous mathematical proof. Several portions of the chapter can well be omitted at a first reading.

We begin with some general descriptions and proceed from them to the calculating of a determinant from its associated matrix. In practice, the method is useful mainly for deriving determinants of matrices of small order, but it is more informative about the structure of a determinant than is a formal definition. Nevertheless, a formal definition, which perforce is lengthy and quite difficult to follow, is given in Section 2.

### 1. EXPANSION BY MINORS

A determinant is a polynomial of the elements of a square matrix. It is a scalar. It is the sum of certain products of the elements of the matrix from

which it is derived, each product being multiplied by +1 or -1 according to certain rules.

Determinants are defined only for square matrices—the determinant of a nonsquare matrix is undefined, and therefore does not exist. The determinant of a (square) matrix of order  $n$  is referred to as an  $n$ -order determinant and the customary notation for the determinant of the matrix  $A$  is  $|A|$ , it being assumed that  $A$  is square. Some texts use the notation  $\|A\|$  or  $\det A$ , but  $|A|$  is common and is used throughout this book. Obtaining the value of  $|A|$  by adding the appropriate products of the elements of  $A$  (with the correct +1 or -1 factor included in the product) is variously referred to as *evaluating* the determinant, *expanding* the determinant or *reducing* the determinant. The procedure for evaluating a determinant is first illustrated by a series of simple numerical examples.

#### a. First- and second-order determinants

To begin, we have the elementary result that the determinant of a  $1 \times 1$  matrix is the value of its sole element.

The value of a second-order determinant is scarcely more difficult. For example, the determinant of

$$A = \begin{bmatrix} 7 & 3 \\ 4 & 6 \end{bmatrix} \text{ is written } |A| = \begin{vmatrix} 7 & 3 \\ 4 & 6 \end{vmatrix}$$

and is calculated as

$$|A| = 7(6) - 3(4) = 30.$$

This illustrates the general result for expanding a second-order determinant: from the product of the diagonal terms subtract the product of the off-diagonal terms. In other words, the determinant of a  $2 \times 2$  matrix consists of the product (multiplied by +1) of the diagonal terms plus the product (multiplied by -1) of the off-diagonal terms. Hence in general

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} + (-1)a_{12}a_{21} = a_{11}a_{22} - a_{12}a_{21}.$$

Further examples are

$$\begin{vmatrix} 6 & 8 \\ 17 & 21 \end{vmatrix} = 6(21) - 8(17) = -10$$

and

$$\begin{vmatrix} 9 & -3.81 \\ 7 & -1.05 \end{vmatrix} = 9(-1.05) - (-3.81)(7) = 17.22.$$

The evaluation of a second-order determinant is patently simple.

### b. Third-order determinants

A third-order determinant can be expanded as a linear function of three second-order determinants derived from it. Their coefficients are elements of a row (or column) of the main determinant, each product being multiplied by +1 or -1. For example, the expansion of

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix}$$

based on the elements of the first row, 1, 2 and 3, is

$$\begin{aligned} |\mathbf{A}| &= 1(+1) \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 2(-1) \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} + 3(+1) \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 1(50 - 48) - 2(40 - 42) + 3(32 - 35) = -3. \end{aligned}$$

The determinant that multiplies each element of the chosen row (in this case the first row) is the determinant derived from  $|\mathbf{A}|$  by crossing out the row and column containing the element concerned. For example, the first element, 1, is multiplied by the determinant  $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$  which is obtained from  $|\mathbf{A}|$  through crossing out the first row and column; and the element 2 is multiplied (apart from the factor of -1) by the determinant derived from  $|\mathbf{A}|$  by deleting the row and column containing that element—namely, the first row and second column—so leaving  $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$ . Determinants obtained in this way are called *minors* of  $|\mathbf{A}|$ , that is to say,  $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$  is the minor of the element 1 in  $|\mathbf{A}|$ , and  $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$  is the minor of the element 2.

The (+1) and (-1) factors in the expansion are decided on according to the following rule: if  $\mathbf{A}$  is written in the form  $\mathbf{A} = \{a_{ij}\}$ , the product of  $a_{ij}$  and its minor in the expansion of the determinant  $|\mathbf{A}|$  is multiplied by  $(-1)^{i+j}$ . Hence because the element 1 in the example is the element  $a_{11}$ , its product with its minor is multiplied by  $(-1)^{1+1} = +1$ ; and for the element 2, which is  $a_{12}$ , its product with its minor is multiplied by  $(-1)^{1+2} = -1$ .

Let us denote the minor of the element  $a_{11}$  by  $|\mathbf{M}_{11}|$ , where  $\mathbf{M}_{11}$  is a submatrix of  $\mathbf{A}$  obtained by deleting the first row and first column. Then in the preceding example  $|\mathbf{M}_{11}| = \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$ . Similarly, if  $|\mathbf{M}_{12}|$  is the minor of  $a_{12}$ , then  $|\mathbf{M}_{12}| = \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$ ; and if  $|\mathbf{M}_{13}|$  is the minor of  $a_{13}$ , then  $|\mathbf{M}_{13}| = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$ . With this notation, the preceding expansion of  $|\mathbf{A}|$  is

$$|\mathbf{A}| = a_{12}(-1)^{1+1} |\mathbf{M}_{11}| + a_{12}(-1)^{1+2} |\mathbf{M}_{12}| + a_{13}(-1)^{1+3} |\mathbf{M}_{13}|.$$

This method of expanding a determinant is known as *expansion by the elements of a row (or column)* or sometimes as *expansion by minors*. It has been illustrated using elements of the first row but can also be applied to the elements of any row (or column). For example, the expansion of  $|\mathbf{A}|$  just considered, using elements of the second row gives

$$\begin{aligned} |\mathbf{A}| &= 4(-1) \begin{vmatrix} 2 & 3 \\ 7 & 10 \end{vmatrix} + 5(+1) \begin{vmatrix} 1 & 3 \\ 7 & 10 \end{vmatrix} + 6(-1) \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} \\ &= -4(-4) + 5(-11) - 6(-6) = -3 \end{aligned}$$

as before, and using elements of the first column the expansion also gives the same result:

$$\begin{aligned} |\mathbf{A}| &= 1(+1) \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 4(-1) \begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} + 7(+1) \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\ &= 1(2) - 4(-4) + 7(-3) = -3. \end{aligned}$$

The minors in these expansions are derived in exactly the same manner as in the first, by crossing out rows and columns of  $|\mathbf{A}|$ , and so are the (+1) and (-1) factors; for example, the minor of the element 4 is  $|\mathbf{A}|$  with second row and first column deleted, and since 4 is  $a_{21}$  its product with its minor is multiplied by  $(-1)^{2+1} = -1$ . Other terms are obtained in a similar manner.

The foregoing example illustrates the expansion of the general third-order determinant

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$



Expanding this by elements of the first row gives

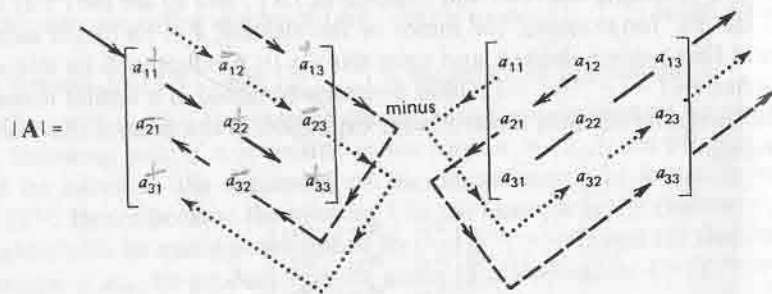
$$\begin{aligned} |\mathbf{A}| &= a_{11}(+1) \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1) \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(+1) \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

The reader should verify that expansion by the elements of any other row or column leads to the same result. No matter by what row or column the expansion is made, the value of the determinant is the same. Note that once a row or column is decided on and the sign calculated for the product of the first element therein with its minor, the signs for the succeeding products alternate from plus to minus and minus to plus.

The diagonal of a square matrix has already been defined. The *secondary diagonal* (as it is sometimes called) is the line of elements in a square matrix going from the lower left corner to the upper right, the "other" diagonal, so to speak. Thus the determinant of a matrix of order 2 is the product of elements on the diagonal minus that of the elements on the secondary diagonal. This extends to the determinant of a  $3 \times 3$  matrix. From the end of the preceding paragraph, the expression for  $|\mathbf{A}|$  can be rewritten with its three positive terms followed by its three negative terms:

$$\begin{aligned} |\mathbf{A}| &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} \\ &\quad - (a_{31}a_{22}a_{13} + a_{11}a_{32}a_{23} + a_{12}a_{21}a_{33}). \end{aligned}$$

This can be represented as in the accompanying figure. The positive terms in  $|\mathbf{A}|$  are the products of the three elements of  $\mathbf{A}$  joined by lines "parallel" to the diagonal; the negative terms are the products of elements joined by



lines "parallel" to the secondary diagonal. Unfortunately, this diagrammatic expansion does not work for determinants beyond those of order 3.

### c. $n$ -order determinants

The expansion of an  $n$ -order determinant by this method is an extension of the expansion of a third-order determinant as just given. Thus the determinant of the  $n \times n$  matrix  $\mathbf{A} = \{a_{ij}\}$  for  $i, j = 1, 2, \dots, n$ , is obtained as follows. Consider the elements of any one row (or column): multiply each element,  $a_{ij}$ , of this row (or column) by its minor,  $|\mathbf{M}_{ij}|$ , the determinant derived from  $|\mathbf{A}|$  by crossing out the row and column containing  $a_{ij}$ ; multiply the product by  $(-1)^{i+j}$ ; add the signed products and their sum is the determinant  $|\mathbf{A}|$ ; i.e., when expanding by elements of a row

$$|\mathbf{A}| = \sum_{j=1}^n a_{ij}(-1)^{i+j} |\mathbf{M}_{ij}| \quad \text{for any } i, \quad (1)$$

and when expanding by elements of a column

$$|\mathbf{A}| = \sum_{i=1}^n a_{ij}(-1)^{i+j} |\mathbf{M}_{ij}| \quad \text{for any } j. \quad (2)$$

This expansion is used recurrently when  $n$  is large, i.e., each  $|\mathbf{M}_{ij}|$  is expanded by the same procedure. Thus a fourth-order determinant is first expanded as four signed products each involving a third-order minor, and each of these is expanded as a sum of three signed products involving a second-order determinant. Consequently, a fourth-order determinant ultimately involves  $4 \times 3 \times 2 = 24$  products of its elements, each product containing four elements. This leads us to the general statement that the determinant of a square matrix of order  $n$  is a sum of  $(n)!$  signed products, each product involving  $n$  elements of the matrix. The determinant is referred to as an  $n$ -order determinant. Utilizing methods given by Aitken (1948) it can be shown that each product has one and only one element from each row and column and that all such products are included and none occurs more than once.

This method of evaluation requires lengthy computing for determinants of order exceeding 3 to 4, say, because for order  $n$  there are  $n!$  terms to be calculated, and even for  $n$  as small as  $n = 10$  this means 3,628,800 terms! Fortunately, easier methods exist, but because the method already discussed forms the basis of these easier methods, it has been considered first. Furthermore, it is useful in presenting various properties of determinants.

<sup>1</sup> $n!$ , read as " $n$  factorial" is the product of all integers 1 through  $n$ ; e.g.,  $4! = 1(2)(3)(4) = 24$ .

## 2. FORMAL DEFINITION

As has been indicated, the determinant of an  $n$ -order square matrix is the sum of  $n!$  signed products of the elements of the matrix, each product containing one and only one element from every row and column of the matrix. Evaluation through expansion by elements of a row or column, as represented in equations (1) and (2), yields the requisite products with their correct signs when the minors are successively expanded at each stage by the same procedure. We have already referred to this as the method of expansion by minors.

There are many formal definitions of a determinant, all, of course, consistent with each other, and all leading to the same results as expansion by minors. The following is adapted from Ferrar (1941). The determinant of a square matrix of order  $n$ ,  $A = \{a_{ij}\}$   $i, j = 1, 2, \dots, n$ , is (i) the sum of all possible products of  $n$  elements of  $A$  such that (ii) each product has one and only one element from every row and column of  $A$ , (iii) the sign of a product being  $(-1)^p$  for  $p = \sum_{i=1}^n n_i$ , where (iv) by writing the product with its  $i$  subscripts in natural order  $a_{1j_1} a_{2j_2} \dots a_{ij_i} \dots a_{nj_n}$ , the  $j$  subscripts  $j_i, i = 1, 2, \dots, n$ , being the first  $n$  integers in some order,  $n_i$  is defined as the number of  $j$ 's less than  $j_i$  that follow  $j_i$  in this order.

**Example.** In the expansion of

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

the  $j_1, j_2, j_3$  array in the product  $a_{12}a_{23}a_{31}$  is  $j_1 = 2, j_2 = 3$  and  $j_3 = 1$ . Now  $n_1$  is the number of  $j$ 's in this array that are less than  $j_1$  but follows it; i.e., that are less than 2 but follow it. There is only one, namely  $j_3 = 1$ ; therefore  $n_1 = 1$ . Likewise  $n_2 = 1$  because  $j_2 = 3$  is followed by only one  $j$  less than 3, and finally  $n_3 = 0$ . Thus

$$n_1 + n_2 + n_3 = 1 + 1 + 0 = 2$$

and the sign of the product  $a_{12}a_{23}a_{31}$  in  $|A|$  is therefore  $(-1)^2 = +1$ . That the sign is also positive in the expansion of  $|A|$  by minors is easily shown, for in expanding by elements of the first row the term involving  $a_{12}$  is

$$-a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -a_{12}(a_{21}a_{33} - a_{23}a_{31}),$$

which includes the term  $+a_{12}a_{23}a_{31}$ .

We shall now indicate the generality of the result implicit in the above example, that the formal definition of  $|A|$  agrees with the procedure of expansion by minors

The definition states that each product of  $n$  elements in  $|A|$  contains one element from every row and column. Suppose we set out to form one such product by selecting elements for it one at a time from each of the rows of  $A$ . Starting with the first row there are  $n$  possible choices of an element as the first element of the product. Having made a choice there are then only  $n - 1$  possible choices in the second row because the column containing the element chosen from the first row must be excluded from the choices in subsequent rows (in order to have one and only one element from each column as well as from each row). Similarly, the column containing the element chosen from the second row has to be excluded from the possible choices in the third and subsequent rows. Thus there are  $(n - 2)$  possible choices in the third row,  $(n - 3)$  in the fourth row and so on. Hence, based on the definition, the total number of products in  $|A|$  is

$$n(n-1)(n-2) \dots (3)(2)(1) = (n!).$$

On the other hand, expansion of  $|A|$  by minors initially gives  $|A|$  as a sum of  $n$  elements each multiplied (apart from sign) by its minor, a determinant of order  $n - 1$ . Each minor can be expanded in similar fashion as the sum of  $(n - 1)$  elements multiplied (apart from sign) by minors that are determinants of order  $n - 2$ . In this way we see that the complete determinant consists of  $n(n - 1)(n - 2) \dots (3)(2)(1) = (n!)$  signed products each containing  $n$  elements. And at each stage of the expansion the method of deleting a row and a column from a determinant to obtain the minor of an element ensures that in each product the  $n$  elements consist of one from every row and column of  $A$ . Hence (apart from sign) both the definition and the procedure of expansion by minors lead to the same set of products.

We now show that the sign of each product is the same in both cases. The sign of any product in a determinant expanded by minors is the product of the signs applied to each minor involved in the derivation of the product. Consider the product as

$$a_{1j_1} a_{2j_2} a_{3j_3} \dots a_{nj_n}.$$

Supposing the first expansion of  $|A|$  is by elements of the first row, the sign attached to the minor of  $a_{1j_1}$  is  $(-1)^{1+j_1}$ . If the expansion of this minor is by elements of the second row of  $|A|$  which is now the first row of the minor, the minor therein of  $a_{2j_2}$  will have attached to it the sign  $(-1)^{1+j_2-1}$  if  $j_1$  is less than  $j_2$  and  $(-1)^{1+j_2}$  if  $j_1$  exceeds  $j_2$ . Likewise the sign attached to the minor of  $a_{3j_3}$  in the minor of  $a_{2j_2}$  in the minor of  $a_{1j_1}$  in the expansion of  $|A|$  will be  $(-1)^{1+j_2-2}$  if both  $j_1$  and  $j_2$  are less than  $j_3$ . It will be  $(-1)^{1+j_2-1}$  if either of  $j_1$  or  $j_2$  is less than  $j_3$  and the other exceeds it, and it will be  $(-1)^{1+j_2}$  if both  $j_1$  and  $j_2$  exceed  $j_3$ . Hence the sign will be  $(-1)^{1+j_1-m_3}$  where  $m_3$  is the number of  $j$ 's less than  $j_3$ .

which precede it in the array  $j_1 j_2 j_3 \cdots j_n$ . In general the sign of the  $i$ th minor involved will be  $(-1)^{1+j_i-m_i}$  where  $m_i$  is the number of  $j$ 's less than  $j_i$  which precede it in the array. Therefore the combined sign is

$$\prod_{i=1}^n (-1)^{1+j_i-m_i} = (-1)^q \quad \text{for } q = \sum_{i=1}^n (1 + j_i - m_i).$$

Now, since the  $j_i$ 's are the first  $n$  integers in some order, the number of them that are less than some particular  $j_i$  is  $j_i - 1$ . Therefore, by the definitions of  $n_i$  and  $m_i$  their sum is  $j_i - 1$ , i.e.,  $n_i + m_i = j_i - 1$ , which gives

$$q = \sum_{i=1}^n (1 + j_i - m_i) = \sum_{i=1}^n (2 + n_i) = 2n + \sum_{i=1}^n n_i.$$

Therefore  $(-1)^q = (-1)^p$  for  $p = \sum_{i=1}^n n_i$ , so that when the determinant is expanded by minors the sign of a product is  $(-1)^p$  as specified by the definition. Hence the definition and the method of expansion by minors are equivalent.

### 3. BASIC PROPERTIES

Several basic properties of determinants are useful for circumventing both the tedious expansion by elements of a row (or column) described in Section 1 and the formal definition. Each is stated in the form of a theorem, followed by an example and proof.

#### a. Determinant of a transpose

**Theorem.** The determinant of the transpose of a matrix equals the determinant of the matrix itself:  $|A'| = |A|$ .

**Example.** On using expansion by the first row,

$$|A| = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 1 & 2 \\ 4 & 4 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 4 & 9 \end{vmatrix} + \begin{vmatrix} 2 & 2 \\ 4 & 9 \end{vmatrix} = 1 + 10 = 11$$

and

$$\begin{aligned} |A'| &= \begin{vmatrix} 1 & 2 & 4 \\ -1 & 1 & 2 \\ 0 & 2 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 4 \\ 2 & 9 \end{vmatrix} - 2 \begin{vmatrix} -1 & 4 \\ 0 & 9 \end{vmatrix} + 4 \begin{vmatrix} -1 & 1 \\ 0 & 2 \end{vmatrix} \\ &= 1 + 18 - 8 = 11. \end{aligned}$$

**Proof.** The formal definition of Section 2 and its equivalence to expansion by elements of a row provide the means for proving equality of that expansion with expansion by elements of a column. Therefore expansion of  $|A|$  by elements of its  $i$ th row is the same as expansion of  $|A'|$  by elements of its  $i$ th column, except that in  $|A|$  all minors will be of matrices that are the transpose of those in the corresponding minors in  $|A'|$ . This will be true right down to the minors of the  $2 \times 2$  matrices in each case. But these are equal: e.g.,

$$\begin{vmatrix} a & b \\ x & y \end{vmatrix} = ay - bx = \begin{vmatrix} a & x \\ b & y \end{vmatrix}.$$

Therefore  $|A| = |A'|$ .

QED

**Corollary.** All properties of  $|A|$  in terms of rows can be stated equivalently in terms of columns—because expanding  $|A'|$  by rows is identical to expanding  $|A|$  by columns.

#### b. Two rows the same

**Theorem.** If two rows of  $A$  are the same,  $|A| = 0$ .

**Example**

$$\begin{vmatrix} 1 & 4 & 3 \\ 7 & 5 & 2 \\ 7 & 5 & 2 \end{vmatrix} = \begin{vmatrix} 5 & 2 \\ 5 & 2 \end{vmatrix} - 4 \begin{vmatrix} 7 & 2 \\ 7 & 2 \end{vmatrix} + 3 \begin{vmatrix} 7 & 5 \\ 7 & 5 \end{vmatrix} = 0,$$

because

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0. \quad (3)$$

**Proof.** If  $A$  has two rows the same, expand  $|A|$  by minors so that the  $2 \times 2$  minors in the last step of the expansion are from two equal rows. Then (3) shows that all those minors are zero, and so  $|A| = 0$ . QED

#### c. Cofactors

The signed minor  $(-1)^{i+j} |M_{ij}|$  used in (1) and (2) is called a cofactor:

$$c_{ij} = (-1)^{i+j} |M_{ij}|, \quad (4)$$

where  $M_{ij}$  is  $A$  with its  $i$ th row and  $j$ th column deleted.

Two properties of cofactors are important. The first is that the sum of products of the elements of a row with their own cofactors is the determinant:

$$\sum_j a_{ij}c_{ij} = \sum_j a_{ij}(-1)^{i+j} |M_{ij}| = |A| \quad \text{for each } i. \quad (5)$$

This is just (1). Second, the sum of products of elements of a row with cofactors of some other row is zero:

$$\sum_j a_{ij}c_{kj} = 0 \quad \text{for } i \neq k. \quad (6)$$

This is so because, for  $i \neq k$ ,

$$\sum_j a_{ij}c_{kj} = \sum_j a_{ij}(-1)^{k+j} |M_{kj}| = 0$$

with the last equality being true because it represents, by comparison with (5), expansion of  $|A|$  with its  $k$ th row replaced by its  $i$ th row and so is a determinant having two rows the same and hence is zero. [Equations (5) and (6) can also be restated in terms of columns of  $A$ ; see Section 5.3]

**Example.** In the determinant

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix}$$

the cofactors of the elements of the first row are as follows:

$$\text{that of the 1 is } (-1)^{1+1} \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} = 50 - 48 = 2;$$

$$\text{that of the 2 is } (-1)^{1+2} \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} = -40 + 42 = 2;$$

$$\text{that of the 3 is } (-1)^{1+3} \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 32 - 35 = -3.$$

Multiplying these elements of the first row by their cofactors gives  $|A|$ :

$$|A| = 1(2) + 2(2) + 3(-3) = -3.$$

But multiplying the elements of another row, the second say, by these cofactors gives zero:

$$4(2) + 5(2) + 6(-3) = 0.$$

The determinantal form of this last expression is

$$4(-1)^{1+1} \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 5(-1)^{1+2} \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} + 6(-1)^{1+3} \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = \begin{vmatrix} 4 & 5 & 6 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix},$$

which is clearly zero because two rows are the same.

#### d. Adding multiples of a row (column) to a row (column)

**Theorem.** Adding to one row (column) of a determinant any multiple of another row (column) does not affect the value of the determinant.

**Example**

$$|A| = \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix} \\ = 1(17 - 147) - 3(8 - 42) + 2(56 - 34) = 16.$$

And adding four times row 1 to row 2 does not affect the value of  $|A|$ :

$$|A| = \begin{vmatrix} 1 & 3 & 2 \\ 8 + 4 & 17 + 12 & 21 + 8 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{vmatrix} \\ = 1(29 - 203) - 3(12 - 58) + 2(84 - 58) \\ = -174 + 138 + 52 = 16. \quad (7)$$

**Proof.** Suppose  $B$  of order  $n$  has  $[b_{11} \ b_{12} \ \cdots \ b_{1n}]$  and  $[b_{21} \ b_{22} \ \cdots \ b_{2n}]$  as its first two rows. Let  $A$  be  $B$  with  $\lambda$  times its second row added to its first. We show that  $|A| = |B|$ .

With  $|M_{ij}|$  being the minor of  $b_{ij}$  in  $|B|$ , expansion of  $|A|$  by elements of its first row, using (1), gives

$$\begin{aligned} |A| &= \sum_{j=1}^n (b_{1j} + \lambda b_{2j})(-1)^{1+j} |M_{1j}| \\ &= \sum_{j=1}^n b_{1j}(-1)^{1+j} |M_{1j}| + \lambda \sum_{j=1}^n b_{2j}(-1)^{1+j} |M_{1j}|. \end{aligned}$$

The first sum here is  $|B|$ , directly from (1). The second sum is an example of (6) and hence is zero. Therefore  $|A| = |B|$ . QED

### e. Products

**Theorem.**  $|AB| = |A||B|$  when  $A$  and  $B$  are square and of the same order.

**Example.** With

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 6 & 6 \end{bmatrix} = \begin{bmatrix} 16 & 15 \\ 60 & 57 \end{bmatrix}, |AB| = 912 - 900 = 12,$$

and

$$|A||B| = \begin{vmatrix} 1 & 2 \\ 3 & 8 \end{vmatrix} \begin{vmatrix} 4 & 3 \\ 6 & 6 \end{vmatrix} = 2(6) = 12.$$

Before proving the theorem, we first show that any determinant can be reduced to the determinant of a triangular matrix, and then consider two useful lemmas.

(i) *Reduction to triangular form.* The determinant of a lower triangular matrix is the product of its diagonal elements. For example,

$$\begin{vmatrix} 6 & 0 & 0 \\ 3 & -1 & 0 \\ 7 & 3 & -5 \end{vmatrix} = 6(-1)(-5) = 30.$$

Verification is self-evident—through expansion by elements of successive rows.

Any determinant can be evaluated as the determinant of a lower triangular matrix. This is done by adding multiples of rows to other rows to reduce the determinant to triangular form. For example, consider

$$|P| = \begin{vmatrix} 3 & 8 & 7 \\ 1 & 2 & 4 \\ -1 & 3 & 2 \end{vmatrix}.$$

[4.3]

We use row operations to reduce all elements above the diagonal to zero. Start first with the 2 of the (3,3) element to reduce elements above it to zero; i.e., to row 1 add  $(-7/2)$  times row 3 and to row 2 add  $(-2)$  times row 3. This gives

$$|P| = \begin{vmatrix} 6\frac{1}{2} & -2\frac{1}{2} & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix}.$$

Now use the (2,2) element to reduce the elements above it to zero, by subtracting  $-2\frac{1}{2}/(-4)$  times row 2 from row 1. Then

$$|P| = \begin{vmatrix} 6\frac{1}{2} - 2\frac{1}{2}(3)/4 & 0 & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 37/8 & 0 & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix},$$

from which  $|P| = (37/8)(-4)(2) = -37$ , as can be verified by any other form of expansion.

The method used in the example extends quite naturally to a determinant of any order  $n$ . Starting with the  $n$ th diagonal element, the elements above it are reduced to zero, then so are those above the  $(n-1)$ th,  $(n-2)$ th, ..., 3rd, 2nd diagonal elements. This yields a lower triangular matrix. [Similar calculations starting with the first diagonal element and using it to reduce all elements below it to zero, followed by doing the same to those below the 2nd, 3rd, ...,  $(n-1)$ th and  $n$ th diagonal elements, yields an upper triangular matrix.] Then the determinant is the product of the resultant diagonal elements.

(ii) *Two useful lemmas.*

**Lemma 1.** For square matrices  $P$ ,  $Q$  and  $X$  of the same order  $n$

$$\begin{vmatrix} P & 0 \\ X & Q \end{vmatrix} = |P||Q|. \quad (8)$$

**Proof.** By row operations on the first  $n$  rows of the left-hand side of (8) reduce  $P$  to lower triangular form. Then expand the left-hand side by elements of successive rows for the first  $n$  rows. The result is  $|P||Q|$ , and so (8) is upheld. QED

**Example.** Using  $P$  of the preceding example and

$$Q = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix},$$

we have

$$\begin{vmatrix} 3 & 8 & 7 & 0 & 0 & 0 \\ 1 & 2 & 4 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 & 0 \\ x & y & z & 1 & 2 & 3 \\ p & q & r & 4 & 5 & 4 \\ a & b & c & 3 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 37/8 & 0 & 0 & 0 & 0 & 0 \\ 3 & -4 & 0 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 & 0 \\ x & y & z & 1 & 2 & 3 \\ p & q & r & 4 & 5 & 4 \\ a & b & c & 3 & 2 & 1 \end{vmatrix} = |\mathbf{P}||\mathbf{Q}|$$

$$= -37 \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix}.$$

**Lemma 2.** For  $\mathbf{R}$  and  $\mathbf{S}$  square and of the same order  $n$

$$\begin{vmatrix} \mathbf{0} & \mathbf{R} \\ -\mathbf{I} & \mathbf{S} \end{vmatrix} = |\mathbf{R}|. \quad (9)$$

**Proof.** Expand the left-hand side of (9) by elements of the successive columns through the  $-\mathbf{I}$  to get

$$\begin{vmatrix} \mathbf{0} & \mathbf{R} \\ -\mathbf{I} & \mathbf{S} \end{vmatrix} = [(-1)^{n+1+1}(-1)]^n |\mathbf{R}| = (-1)^{n(n+3)} |\mathbf{R}| = |\mathbf{R}|. \quad \text{QED}$$

(iii) *Determinant of a product.* Consider the determinant of the matrix product

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix},$$

namely

$$\left| \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} \right| = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix}. \quad (10)$$

Note that using

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

in premultiplication in (10) merely adds multiples of rows of

$$\left| \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} \right|$$

to other rows and so

$$\text{L.H.S. of (10)} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = |\mathbf{A}||\mathbf{B}|, \text{ by (8).}$$

But

$$\text{R.H.S. of (10)} = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = |\mathbf{AB}|, \text{ by (9).}$$

Therefore  $|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}|$  or

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|. \quad (11)$$

Note that for all three determinants in (11) to exist,  $\mathbf{A}$  and  $\mathbf{B}$  must both be square, of the same order. This is also implied by (10). The result (11) does not hold when  $\mathbf{A}$  or  $\mathbf{B}$  or both are rectangular; even if  $|\mathbf{AB}|$  exists in the form  $|\mathbf{A}_{r \times c} \mathbf{B}_{c \times r}|$ ,  $|\mathbf{A}|$  and  $|\mathbf{B}|$  do not and (11) does not apply.

There are several useful corollaries to this result.

#### Corollaries

1.  $|\mathbf{AB}| = |\mathbf{BA}|$  (because  $|\mathbf{A}||\mathbf{B}| = |\mathbf{B}||\mathbf{A}|$ ).
2.  $|\mathbf{A}^2| = |\mathbf{A}|^2$  (each equals  $|\mathbf{A}||\mathbf{A}|$ );  $|\mathbf{A}^k| = |\mathbf{A}|^k$  is the extension.
3. For orthogonal  $\mathbf{A}$ ,  $|\mathbf{A}| = \pm 1$  (because  $\mathbf{AA}' = \mathbf{I}$  implies  $|\mathbf{A}|^2 = 1$ ).
4. For idempotent  $\mathbf{A}$ ,  $|\mathbf{A}| = 0$  or  $1$  (because  $\mathbf{A}^2 = \mathbf{A}$  implies  $|\mathbf{A}|^2 = |\mathbf{A}|$ ).

#### 4. ELEMENTARY ROW OPERATIONS

In Section 3d we introduced the operation of adding a multiple of a row (column) of a determinant to another row (column). It did not alter the value of the determinant. That same operation on a matrix can be represented by a matrix product, e.g.,

$$\begin{bmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} \quad (12)$$

is the matrix product representation of adding 4 times row 1 to row 2, as used in (7). Applying the product rule of Section 3e gives

$$\begin{vmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix}$$

because

$$\begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1. \quad (13)$$

The matrix involved in (12) and (13) is known as an *elementary operator matrix*. For the operation of adding  $\lambda$  times one row of a matrix to another, it is always an identity matrix with  $\lambda$  in an off-diagonal element; and its determinant is always unity. We denote such a matrix by  $P_{ij}(\lambda)$ . Then  $P_{ij}(\lambda)A$  is the operation on  $A$  of adding to its  $i$ th row  $\lambda$  times its  $j$ th row; and

$$|P_{ij}(\lambda)| = 1; \quad \text{e.g., in (13), } |P_{21}(4)| = 1. \quad (14)$$

Two other elementary operations are (i) interchanging two rows of a matrix, and (ii) multiplying a row by a scalar. These are represented by matrices

$$E_{ij} = I \text{ with } i\text{th and } j\text{th rows interchanged}$$

and

$$R_{ii}(\lambda) = I \text{ with } i\text{th diagonal element replaced by } \lambda.$$

Then  $E_{ij}A$  is  $A$  with its  $i$ th and  $j$ th rows interchanged, and  $R_{ii}(\lambda)A$  is  $A$  with its  $i$ th row multiplied by  $\lambda$ .

**Examples.**

$$E_{12}A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 17 & 21 \\ 1 & 3 & 2 \\ 2 & 7 & 1 \end{bmatrix}$$

and

$$R_{33}(5)A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 10 & 35 & 5 \end{bmatrix}.$$

All three of these elementary operator matrices have simple determinants:

$$|P_{ij}(\lambda)| = 1, \quad |R_{ii}(\lambda)| = \lambda \quad \text{and} \quad |E_{ij}| = -1. \quad (15)$$

Special cases of these elementary operator matrices used in combination with the product rule  $|AB| = |A||B|$  provide useful techniques for simplifying the expansion of a determinant.

#### a. Factorization

**Theorem.** When  $\lambda$  (a nonzero scalar) is a factor of a row (column) of  $|A|$  then it is also a factor of  $|A|$ :

$$|A| = \lambda |A \text{ with } \lambda \text{ factored out of a row (column)}|. \quad (16)$$

**Example.**

$$\begin{vmatrix} 4 & 6 \\ 1 & 7 \end{vmatrix} = 2 \begin{vmatrix} 2 & 3 \\ 1 & 7 \end{vmatrix}, \quad \text{i.e., } 4(7) - 1(6) = 22 = 2[2(7) - 1(3)].$$

**Proof.** When  $\lambda$  is a factor of every element in the  $i$ th row of  $A$ ,

$$A = R_{ii}(\lambda)[A \text{ with } \lambda \text{ factored out of its } i\text{th row}]. \quad (17)$$

In the determinant of (17), use (11) and (15), to get (16). QED

**Corollary.** If  $A$  is an  $n \times n$  matrix and  $\lambda$  is a scalar, the determinant of the matrix  $\lambda A$  is  $\lambda^n |A|$ ; i.e.,  $|\lambda A| = \lambda^n |A|$ . (In this case  $\lambda$  is a factor of each of the  $n$  rows of  $\lambda A$ , so that after factoring  $\lambda$  from each row the determinant  $|A|$  remains.)

**Example.**

$$\begin{vmatrix} 3 & 0 & 27 \\ -9 & 3 & 0 \\ 15 & 6 & -3 \end{vmatrix} = 3^3 \begin{vmatrix} 1 & 0 & 9 \\ -3 & 1 & 0 \\ 5 & 2 & -1 \end{vmatrix} = -2700.$$

**Corollary.** If one row of a determinant is a multiple of another row, the determinant is zero. (Factoring out the multiple reduces the determinant to having two rows the same. Hence the determinant is zero.)

**Example.**

$$\begin{vmatrix} -3 & 6 & 12 \\ 2 & -4 & -8 \\ 7 & 5 & 9 \end{vmatrix} = -(1.5) \begin{vmatrix} 2 & -4 & -8 \\ 2 & -4 & -8 \\ 7 & 5 & 9 \end{vmatrix} = 0.$$

## b. A row (column) of zeros

**Theorem.** When a determinant has zero for every element of a row (or column), the determinant is zero.

**Example.**

$$\begin{vmatrix} 0 & 0 & 0 \\ 3 & 6 & 5 \\ 2 & 9 & 7 \end{vmatrix} = 0 \begin{vmatrix} 6 & 5 \\ 9 & 7 \end{vmatrix} - 0 \begin{vmatrix} 3 & 5 \\ 2 & 7 \end{vmatrix} + 0 \begin{vmatrix} 3 & 6 \\ 2 & 9 \end{vmatrix} = 0.$$

**Proof.** This is just (16) with  $\lambda = 0$ .

QED

## c. Interchanging rows (columns)

**Theorem.** Interchanging two rows (columns) of a determinant changes its sign.

**Example.** Subtracting twice column 2 from column 1,

$$\begin{vmatrix} 6 & 3 & 0 \\ -1 & 4 & 7 \\ 2 & 5 & 9 \end{vmatrix} = \begin{vmatrix} 0 & 3 & 0 \\ -9 & 4 & 7 \\ -8 & 5 & 9 \end{vmatrix} = -3(-81 + 56) = 75,$$

whereas

$$\begin{vmatrix} 2 & 5 & 9 \\ -1 & 4 & 7 \\ 6 & 3 & 0 \end{vmatrix} = \begin{vmatrix} -8 & 5 & 9 \\ -9 & 4 & 7 \\ 0 & 3 & 0 \end{vmatrix} = -3(-56 + 81) = -75.$$

**Proof.** Using (11) and (15)

$$|E_{ij}A| = |E_{ij}||A| = -|A|.$$

QED

## d. Adding a row to a multiple of a row

In Section 3d we added a multiple of a row to a row. The reader is cautioned against doing the reverse: adding a row to a multiple of a row is not the same thing. It leads to a different result. For example, instead of adding  $\lambda$  times row 2 to row 1 of  $|A|$ , and *not* altering  $|A|$ , the operation of adding row 2 to  $\lambda$  times row 1 gives  $\lambda|A|$ , e.g.,

$$\begin{vmatrix} 1 & \lambda \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2 + 5\lambda & 3 + 9\lambda \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = 3;$$

the operation which gives  $\lambda|A|$  is

$$\begin{vmatrix} \lambda & 1 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2\lambda + 5 & 3\lambda + 9 \\ 5 & 9 \end{vmatrix} = 3\lambda.$$

## 5. EXAMPLES

The foregoing properties can be applied in endless variation in expanding determinants. Efficiency in perceiving a procedure that leads to a minimal amount of effort in any particular case is largely a matter of practice, and beyond describing the possible steps available there is little more that can be said. The underlying method might be summarized as follows. By adding multiples of one row to other rows of the determinant, reduce a column to having only one nonzero element. Expansion by elements of that column then involves only one minor, which is a determinant of order one less than the original determinant. Successive applications of this method reduce the determinant to one of order  $2 \times 2$  whose expansion is obvious. If at any stage these reductions result in elements of a row containing a common factor, this can be factored out, and if they result in a row of zeros, or in two rows being identical, the determinant is zero.

## Example 1

$$\begin{aligned} |A| &= \begin{vmatrix} 1 & 4 & 9 \\ -4 & 7 & 25 \\ 7 & 5 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 9 \\ -4 + 4(1) & 7 + 4(4) & 25 + 4(9) \\ 7 - 7(1) & 5 - 7(4) & 2 - 7(9) \end{vmatrix} \\ &= \begin{vmatrix} 1 & 4 & 9 \\ 0 & 23 & 61 \\ 0 & -23 & -61 \end{vmatrix} = 0. \end{aligned}$$

The 1 in the leading element of  $A$  makes it easy to reduce the other elements of the first column to zero, whereupon adding the last two rows gives  $|A| = 0$ .

## Example 2

$$|B| = \begin{vmatrix} x & y & y & y \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{vmatrix}.$$



Observe that every column of  $\mathbf{B}$  sums to  $x + 3y$ . Therefore add every row of  $\mathbf{B}$  to its first row and factor out  $x + 3y$ :

$$|\mathbf{B}| = (x + 3y) \begin{vmatrix} 1 & 1 & 1 & 1 \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{vmatrix}.$$

Now use the leading element of  $\mathbf{B}$  to reduce the remainder of the first row to zeros. This is achieved by subtracting the first column from each of the other columns:

$$|\mathbf{B}| = (x + 3y) \begin{vmatrix} 1 & 0 & 0 & 0 \\ y & x - y & 0 & 0 \\ y & 0 & x - y & 0 \\ y & 0 & 0 & x - y \end{vmatrix} = (x + 3y)(x - y)^3,$$

the last step being obtained by simple expansion by minors. It involves the more general result that the determinant of a diagonal matrix (Section 1.6) is the product of its diagonal elements:

$$|\mathbf{D}\{d_i\}| = \begin{vmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & & d_n \end{vmatrix} = \prod_{i=1}^n d_i = d_1 d_2 \dots d_n.$$

### Example 3

$$|\mathbf{C}| = \begin{vmatrix} 6 & 8 & 1 & 4 & 2 \\ 18 & 27 & 3 & 13 & 5 \\ -11 & -17 & -2 & -9 & 2 \\ 7 & 10 & 1 & 4 & 7 \\ 4 & 3 & 13 & 8 & 5 \end{vmatrix} = \begin{vmatrix} 6 & 8 & 1 & 4 & 2 \\ 0 & 3 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 & 6 \\ 1 & 2 & 0 & 0 & 5 \\ -74 & -101 & 0 & -44 & -21 \end{vmatrix}$$

The occurrence of a 1 as the (1, 3) element of  $\mathbf{C}$  prompts using it to reduce all other elements in column 3 to zero. This done, expansion by elements of the third column gives

$$|\mathbf{C}| = (-1)^{1+3} \begin{vmatrix} 0 & 3 & 1 & -1 \\ 1 & -1 & -1 & 6 \\ 1 & 2 & 0 & 5 \\ -74 & -101 & -44 & -21 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 0 & 0 & -1 \\ 1 & 17 & 5 & 6 \\ 1 & 17 & 5 & 5 \\ -74 & -164 & -65 & -21 \end{vmatrix} = 0.$$

The  $(-1)$  of the (1, 4) element has been used here to reduce other elements of the first row to zero; then expansion by elements of the first row gives a value of zero because the only nonzero element is associated with a  $3 \times 3$  minor having two rows the same.

Alternatively, in the original form of  $|\mathbf{C}|$ , we can observe that row 2 plus row 3 equals row 4. Hence  $|\mathbf{C}| = 0$ . This is obviously easier—but this observation can be used only when it presents itself.

### Example 4

$$\begin{vmatrix} y & 7 & 7 & 7 \\ 7 & y & 7 & 7 \\ 7 & 7 & y & 7 \\ 7 & 7 & 7 & y \end{vmatrix} = 0 \text{ for } y = 7 \text{ or } -21$$

because for  $y = 7$  all rows are the same, and for  $y = -21$  the sum of all rows is a row of zeros. This determinant is a special case of Example 2.

### Example 5

$$\begin{vmatrix} x & x & x \\ 4 & 3 & -9 \\ -3 & -2 & 10 \end{vmatrix} = 0 \text{ for any } x,$$

because the first row, after factoring out  $x$ , equals the sum of the other two rows.

## 6. DIAGONAL EXPANSION

A matrix can always be expressed as the sum of two matrices one of which is a diagonal matrix, i.e., as  $(\mathbf{A} + \mathbf{D})$  where  $\mathbf{A} = \{a_{ij}\}$  for  $i, j = 1, 2, \dots, n$ , and  $\mathbf{D}$  is a diagonal matrix of order  $n$ . (Diagonal matrices are defined in Section 1.6.) The determinant of such a matrix can then be obtained as a polynomial of the elements of  $\mathbf{D}$ .

The minor of an element of a square matrix of order  $n$  is necessarily a determinant of order  $n - 1$ . But minors are not all of order  $n - 1$ . Deleting any  $r$  rows and any  $r$  columns from a square matrix of order  $n$  leaves a submatrix of order  $n - r$ ; the determinant of this submatrix is a *minor of order  $n - r$* , or an  $(n - r)$ -*order minor*. It is useful to introduce an abbreviated notation for minors of the determinant of  $\mathbf{A}$ ,

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}.$$

They will be denoted by just their diagonal elements; for example,  $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$  is written as  $|a_{11} \ a_{22}|$  and in similar fashion  $\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$  is written  $|a_{12} \ a_{23}|$ . Combined with the notation  $\mathbf{A} = \{a_{ij}\}$  no confusion can arise. For example,  $|a_{21} \ a_{32}|$  denotes the  $2 \times 2$  minor having  $a_{21}$  and  $a_{32}$  as diagonal elements, and from  $|\mathbf{A}|$  we see that the elements in the same rows and columns as these are  $a_{22}$  and  $a_{31}$ , so that

$$|a_{21} \ a_{32}| = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Similarly,

$$|a_{21} \ a_{33} \ a_{44}| = \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix}.$$

We will now consider the determinant  $|\mathbf{A} + \mathbf{D}|$ , initially for a  $2 \times 2$  case, denoting the diagonal elements of  $\mathbf{D}$  by  $x_1$  and  $x_2$ ; for example,

$$|\mathbf{A} + \mathbf{D}| = \begin{vmatrix} a_{11} + x_1 & a_{12} \\ a_{21} & a_{22} + x_2 \end{vmatrix}.$$

By direct expansion

$$|\mathbf{A} + \mathbf{D}| = (a_{11} + x_1)(a_{22} + x_2) - a_{12}a_{21}.$$

Written as a function of  $x_1$  and  $x_2$  this is

$$x_1x_2 + x_1a_{22} + x_2a_{11} + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

In similar fashion it can be shown that

$$\begin{vmatrix} a_{11} + x_1 & a_{12} & a_{13} \\ a_{21} & a_{22} + x_2 & a_{23} \\ a_{31} & a_{32} & a_{33} + x_3 \end{vmatrix} \\ = x_1x_2x_3 + x_1x_2a_{33} + x_1x_3a_{22} + x_2x_3a_{11} \\ + x_1 \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + x_2 \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + x_3 \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

which, using the abbreviated notation, can be written as

$$x_1x_2x_3 + x_1x_2a_{33} + x_1x_3a_{22} + x_2x_3a_{11} \\ + x_1|a_{22} \ a_{33}| + x_2|a_{11} \ a_{33}| + x_3|a_{11} \ a_{22}| \\ + |a_{11} \ a_{22} \ a_{33}|. \quad (18)$$

Considered as a polynomial in the  $x$ 's we see that the coefficient of the product of all the  $x$ 's is unity; the coefficients of the second-degree terms in the  $x$ 's are the diagonal elements of  $\mathbf{A}$ ; the coefficients of the first-degree terms in the  $x$ 's are the second-order minors of  $|\mathbf{A}|$  having diagonals that are part (or all) of the diagonal of  $|\mathbf{A}|$ ; and the term independent of the  $x$ 's is  $|\mathbf{A}|$  itself. The minors of  $|\mathbf{A}|$  in these coefficients, namely those whose diagonals are coincident with the diagonal of  $|\mathbf{A}|$ , are called the *principal minors* of  $|\mathbf{A}|$ .

This method of expansion is useful on many occasions because the determinantal form  $|\mathbf{A} + \mathbf{D}|$  occurs quite often, and when  $|\mathbf{A}|$  is such that many of its principal minors are zero the expansion of  $|\mathbf{A} + \mathbf{D}|$  by this method is greatly simplified.

**Example. If**

$$|\mathbf{X}| = \begin{vmatrix} 7 & 2 & 2 \\ 2 & 8 & 2 \\ 2 & 2 & 9 \end{vmatrix},$$

we have

$$|\mathbf{X}| = |\mathbf{A} + \mathbf{D}| = \begin{vmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{vmatrix} + \begin{vmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{vmatrix}.$$

Every element of  $\mathbf{A}$  is a 2 so that  $|\mathbf{A}|$  and all its  $2 \times 2$  minors are zero. Consequently  $|\mathbf{X}|$  evaluated by (18) consists of only the first four terms:

$$|\mathbf{X}| = 5(6)7 + 5(6)2 + 5(7)2 + 6(7)2 = 424.$$

Evaluating a determinant in this manner is also useful when all elements of the diagonal matrix  $\mathbf{D}$  are the same, i.e., when the  $x_i$ 's are equal. The expansion (7) then becomes

$$x^3 + x^2(a_{11} + a_{22} + a_{33}) + x(|a_{11} \ a_{22}| + |a_{11}a_{33}| + |a_{22}a_{33}|) + |\mathbf{A}|,$$

which is generally written as

$$x^3 + x^2 \text{tr}_1(\mathbf{A}) + x \text{tr}_2(\mathbf{A}) + |\mathbf{A}|$$

where  $\text{tr}_1(\mathbf{A})$  is the trace of  $\mathbf{A}$  (sum of diagonal elements; see Section 1.6) and where  $\text{tr}_2(\mathbf{A})$  is the sum of the principal minors of order 2 of  $|\mathbf{A}|$ . This method of expansion is known as expansion by *diagonal elements* or simply as *diagonal expansion*.

The general diagonal expansion of a determinant of order  $n$ ,

$$|\mathbf{A} + \mathbf{D}| = \begin{vmatrix} a_{11} + x_1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} + x_2 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} + x_3 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} + x_n \end{vmatrix},$$

consists of the sum of all possible products of the  $x_i$  taken  $r$  at a time for  $r = n, n-1, \dots, 2, 1, 0$ , each product being multiplied by its complementary principal minor of order  $n-r$  in  $|\mathbf{A}|$ . By *complementary principal*

*minor* in  $|\mathbf{A}|$  is meant the principal minor having diagonal elements other than those associated in  $|\mathbf{A} + \mathbf{D}|$  with the  $x$ 's of the particular product concerned; for example, the complementary principal minor associated with  $x_1 x_3 x_6$  is  $|a_{22} a_{44} a_{55} a_{77} a_{88} \cdots a_{nn}|$ . When the  $x$ 's are all equal the expression becomes  $\sum_{i=0}^n x^{n-i} \text{tr}_i(\mathbf{A})$  where  $\text{tr}_i(\mathbf{A})$  is the sum of the principal minors of order  $i$  of  $|\mathbf{A}|$  and, by definition,  $\text{tr}_0(\mathbf{A}) = 1$ . Note in passing that  $\text{tr}_n(\mathbf{A}) = |\mathbf{A}|$ .

**Example.** Diagonal expansion gives

$$|\mathbf{A} + \mathbf{D}| = \begin{vmatrix} a+b & a & a & a \\ a & a+b & a & a \\ a & a & a+b & a \\ a & a & a & a+b \end{vmatrix} \\ = b^4 + b^3 \text{tr}_1(\mathbf{A}) + b^2 \text{tr}_2(\mathbf{A}) + b \text{tr}_3(\mathbf{A}) + |\mathbf{A}|$$

where  $\mathbf{A}$  is the  $4 \times 4$  matrix whose every element is  $a$ . Thus  $|\mathbf{A}|$  and all minors of order 2 or more are zero. Hence

$$|\mathbf{A} + \mathbf{D}| = b^4 + b^3(4a) = (4a + b)b^3.$$

This is an extension of the example in Section 4 and similar to Example 2 in Section 3g.

## 7. THE LAPLACE EXPANSION

In the expansion of

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

the minor of  $a_{11}$  is  $|a_{22} a_{33} a_{44}|$ . An extension of this, easily verified, is that the coefficient of  $|a_{11} \ a_{22}|$  is  $|a_{33} \ a_{44}|$ ; namely, the coefficient in  $|\mathbf{A}|$  of

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \quad \text{is} \quad \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} = a_{33}a_{44} - a_{34}a_{43}.$$

Likewise the coefficient of  $|a_{11} \ a_{24}|$  is  $|a_{32} \ a_{43}|$ ; i.e., the coefficient of

$$|a_{11} \ a_{24}| = \begin{vmatrix} a_{11} & a_{14} \\ a_{21} & a_{24} \end{vmatrix} = a_{11}a_{24} - a_{21}a_{14}$$

in the expansion of  $|A|$  is

$$|a_{32} \ a_{43}| = \begin{vmatrix} a_{32} & a_{33} \\ a_{42} & a_{43} \end{vmatrix} = a_{32}a_{43} - a_{33}a_{42}.$$

Each determinant just described as the coefficient of a particular minor of  $|A|$  is the complementary minor in  $|A|$  of that particular minor: it is the determinant obtained from  $|A|$  by deleting from it all the rows and columns containing the particular minor. This is simply an extension of the procedure for finding the coefficient of an individual element in  $|A|$  as derived in the expansion by elements of a row or column discussed earlier. In that case the particular minor is a single element and its coefficient in  $|A|$  is  $|A|$  amended by deletion of the row and column containing the element concerned. A sign factor is also involved, namely  $(-1)^{i+j}$  for the coefficient of  $a_{ij}$  in  $|A|$ . In the extension to coefficients of minors the sign factor is minus one raised to the power of the sum of the subscripts of the diagonal elements of the chosen minor: for example, the sign factor for the coefficient of  $|a_{32} \ a_{43}|$  is  $(-1)^{3+2+4+3} = -1$ , as just given. The complementary minor multiplied by this sign factor can be appropriately referred to as the coefficient of the particular minor concerned. Furthermore, just as the expansion of a determinant is the sum of products of elements of a row (or column) with their coefficients, so also is it the sum of products of all minors of order  $m$  that can be derived from any set of  $m$  rows, each multiplied by its coefficient as just defined. This generalization of the method of expanding a determinant by elements of a row to expanding it by minors of a set of rows was first established by Laplace and so bears his name. Aitken (1948) and Ferrar (1941) are two books where proof of the procedure is given; we shall be satisfied here with a general statement of the method and an example illustrating its use.

The Laplace expansion of a determinant  $|A|$  of order  $n$  can be obtained as follows. (i) Consider any  $m$  rows of  $|A|$ . They contain  $n!/m!(n-m)!$  minors of order  $m$  (see footnote, Section 4.1). (ii) Multiply each of these minors,  $M$  say, by its complementary minor and by a sign factor, where (a) the complementary minor of  $M$  is the  $n-m$  order minor derived from  $|A|$  by deleting the  $m$  rows and columns containing  $M$ , and (b) the sign factor is  $(-1)^\mu$  where  $\mu$  is the sum of the subscripts of the diagonal elements of  $M$ ,  $A$  being defined as  $A = \{a_{ij}\}$ ,  $i, j = 1, 2, \dots, n$ . (iii) The sum of all such products is  $|A|$ .

**Example.** Interchanging the second and fourth rows of

$$|A| = \begin{vmatrix} 1 & 2 & 3 & 0 & 0 \\ 1 & 0 & 4 & 2 & 3 \\ 2 & 0 & 1 & 4 & 5 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & 2 & 1 & 2 & 3 \end{vmatrix} \quad \text{gives} \quad |A| = - \begin{vmatrix} 1 & 2 & 3 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 2 & 0 & 1 & 4 & 5 \\ 1 & 0 & 4 & 2 & 3 \\ 0 & 2 & 1 & 2 & 3 \end{vmatrix}.$$

In this form we will expand  $|A|$  using the Laplace expansion based on the first two rows, ( $m = 2$ ). There are ten minors of order 2 in these two rows; seven of them are

zero because they involve a column of zeros. Hence  $|A|$  can be expanded as the sum of three products involving the three  $2 \times 2$  nonzero minors in the first two rows, namely as  $-|A| =$

$$\begin{aligned} & (-1)^{1+1+2+2} \begin{vmatrix} 1 & 2 \\ -1 & 2 \end{vmatrix} \begin{vmatrix} 1 & 4 & 5 \\ 4 & 2 & 3 \\ 1 & 2 & 3 \end{vmatrix} + (-1)^{1+1+2+3} \begin{vmatrix} 1 & 3 \\ -1 & -1 \end{vmatrix} \begin{vmatrix} 0 & 4 & 5 \\ 0 & 2 & 3 \\ 2 & 2 & 3 \end{vmatrix} \\ & + (-1)^{1+2+2+3} \begin{vmatrix} 2 & 3 \\ 2 & -1 \end{vmatrix} \begin{vmatrix} 2 & 4 & 5 \\ 1 & 2 & 3 \\ 0 & 2 & 3 \end{vmatrix}. \end{aligned}$$

The sign factors in these terms have been derived by envisaging  $A$  as  $\{a_{ij}\}$ . Consequently the first  $2 \times 2$  minor,  $\begin{vmatrix} 1 & 2 \\ -1 & 2 \end{vmatrix}$ , is  $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ , leading to  $(-1)^{1+1+2+2}$  as its sign factor; likewise for the other terms. Simplification of the whole expression gives

$$-|A| = 4 \begin{vmatrix} 1 & 4 & 5 \\ 3 & 0 & 0 \\ 1 & 2 & 3 \end{vmatrix} - 2(2) \begin{vmatrix} 4 & 5 \\ 2 & 3 \end{vmatrix} + (-8) \begin{vmatrix} 2 & 4 & 5 \\ 1 & 0 & 0 \\ 0 & 2 & 3 \end{vmatrix} = -24 - 8 + 16$$

and hence  $|A| = 16$ . It will be found that expansion by a more direct method leads to the same result.

Numerous other methods of expanding determinants are based on extensions of the Laplace expansion, using it recurrently to expand a determinant not only by minors and their complementary minors but also to expand these minors themselves. Many of these expansions are identified by the names of their originators, for example, Cauchy, Binet-Cauchy and Jacoby. A good account of some of them is to be found in Aitken (1948) and Ferrar (1941).

## 8. SUMS AND DIFFERENCES OF DETERMINANTS

The sum of the determinants of each of two (or more) matrices generally does not equal the determinant of the sum. The simplest demonstration of this is

$$\begin{aligned} |A| + |B| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \\ &= a_{11}a_{22} - a_{12}a_{21} + b_{11}b_{22} - b_{12}b_{21} \neq |A + B|. \end{aligned}$$

The same applies to the difference,  $|A| - |B| \neq |A - B|$ .

We may note in passing that both  $|\mathbf{A}| + |\mathbf{B}|$  and  $|\mathbf{A}| - |\mathbf{B}|$  have meaning even when  $\mathbf{A}$  and  $\mathbf{B}$  are square matrices of different orders, because the value of a determinant is a scalar. This contrasts with  $\mathbf{A} + \mathbf{B}$  and  $\mathbf{A} - \mathbf{B}$ , which have meaning only when the matrices are conformable for addition (have the same order).

Another point of interest is that although  $|\mathbf{A}| + |\mathbf{B}|$  does not generally equal  $|\mathbf{A} + \mathbf{B}|$  the latter can be written as the sum of certain other determinants. For example,

$$\begin{aligned} |\mathbf{A} + \mathbf{B}| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = \begin{vmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{vmatrix} \\ &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{vmatrix} \\ &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix}. \end{aligned}$$

In general if  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$ ,  $|\mathbf{A} + \mathbf{B}|$  can be expanded as the sum of  $2^n$   $n$ -order determinants.

**Example.** The reader should verify that

$$\begin{vmatrix} a+b & a & a \\ a & a+b & a \\ a & a & a+b \end{vmatrix} = (3a+b)b^2,$$

by expanding it as  $|b\mathbf{I} + a\mathbf{J}|$ .

#### EXERCISES

1. Show that both

$$\begin{aligned} \text{(a)} \quad & \begin{vmatrix} 1 & 5 & -5 \\ 3 & 2 & -5 \\ 6 & -2 & -5 \end{vmatrix} \text{ and } \begin{vmatrix} -3 & 2 & -6 \\ -3 & 5 & -7 \\ -2 & 3 & -4 \end{vmatrix} \text{ equal } -5; \\ \text{(b)} \quad & \begin{vmatrix} 2 & 6 & 5 \\ -2 & 7 & -5 \\ 2 & -7 & 9 \end{vmatrix} \text{ and } \begin{vmatrix} 2 & -1 & 9 \\ -1 & 7 & 2 \\ 3 & -21 & 2 \end{vmatrix} \text{ equal } 104; \end{aligned}$$

#### EXERCISES

$$\text{(c)} \quad \begin{vmatrix} 1 & 6 & 4 & 3 \\ 2 & 8 & 5 & 4 \\ 3 & 8 & 7 & 5 \\ 4 & 9 & 7 & 7 \end{vmatrix} \text{ and } \begin{vmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{vmatrix} \text{ equal } -16;$$

$$\text{(d)} \quad \begin{vmatrix} 21 & 6 & 3 & 9 \\ 12 & 16 & 36 & 4 \\ 13 & 10 & 19 & 5 \\ 1 & 93 & 81 & 6 \end{vmatrix} \text{ and } \begin{vmatrix} 4 & 6 & 8 & 1 & 2 \\ -1 & -7 & 2 & 3 & 1 \\ 2 & -8 & 12 & 7 & 4 \\ 7 & 9 & 17 & 27 & -5 \\ 8 & 3 & 6 & 2 & 37 \end{vmatrix} \text{ are zero.}$$

$$2. \text{ Show that } \begin{vmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{vmatrix} = 18 \text{ but that } \begin{vmatrix} 0 & 0 & 1 \\ 0 & 3 & 2 \\ 6 & 5 & 4 \end{vmatrix} = -18,$$

$$\text{whereas } \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{vmatrix} = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 6 & 5 & 4 \\ 10 & 9 & 8 & 7 \end{vmatrix} = 180.$$

3. (a) Show that the determinant of  $x\mathbf{I}'$  is zero.  
 (b) If  $\mathbf{A}\mathbf{1} = \mathbf{0}$ , why does  $\mathbf{A}$  have a zero determinant?
4. For

$$\mathbf{A} = \begin{bmatrix} 0 & -a & b & -c \\ a & 0 & -d & e \\ -b & d & 0 & -f \\ c & -e & f & 0 \end{bmatrix}$$

- (a) Calculate  $|\mathbf{A}|$ .  
 (b) Prove  $|\mathbf{I} + \mathbf{A}| = 1 + (a^2 + b^2 + c^2 + d^2 + e^2 + f^2) + |\mathbf{A}|$ .  
 (c) Verify  $|\mathbf{I} + \mathbf{A}|$  by direct expansion.
5. Without expanding the determinants explain why

$$\begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ yz & xz & xy \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ x^2 & y^2 & z^2 \end{vmatrix}.$$

6. Show that

$$\begin{vmatrix} a+b+c & a+b & a & a \\ a+b & a+b+c & a & a \\ a & a & a+b+c & a+b \\ a & a & a+b & a+b+c \end{vmatrix} \\ = c^2(4a+2b+c)(2b+c).$$

7. Li (1955, p. 97) shows that in a random mating population the zygotic proportions of genotypes determined by only two pairs of genes can be expressed as

$$2 \begin{bmatrix} g_{11}^2 & 2g_{11}g_{13} & g_{13}^2 \\ g_{11}g_{31} & g_{11}g_{33} + g_{13}g_{31} & g_{13}g_{33} \\ g_{31}^2 & 2g_{31}g_{33} & g_{33}^2 \end{bmatrix}$$

where  $g_{11}$ ,  $g_{13}$ ,  $g_{31}$  and  $g_{33}$  are the initial gametic outputs. He states that the determinant of this matrix is

$$8 \begin{vmatrix} g_{11} & g_{13} \\ g_{31} & g_{33} \end{vmatrix}^3.$$

Prove this.

8. Explain why the determinant of a diagonal matrix is the product of its diagonal elements. Is the same true for a lower triangular matrix? For an upper triangular matrix? Give examples.
9. Without expanding the determinants, suggest values of  $x$  that satisfy the following equations:

$$(a) \begin{vmatrix} x & x & x \\ 2 & -1 & 0 \\ 7 & 4 & 5 \end{vmatrix} = 0; \quad (b) \begin{vmatrix} 1 & x & x^2 \\ 1 & 2 & 4 \\ 1 & -1 & 1 \end{vmatrix} = 0;$$

$$(c) \begin{vmatrix} 4 & x & x \\ x & 4 & x \\ x & x & 4 \end{vmatrix} = 0; \quad (d) \begin{vmatrix} x & 4 & 4 \\ 4 & x & 4 \\ 4 & 4 & x \end{vmatrix} = 0.$$

10. Explain why

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{vmatrix} + \begin{vmatrix} x & x & x \\ x & x & x \\ x & x & x \end{vmatrix} \\ = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ x & x & x \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ x & x & x \\ 5 & 7 & 9 \end{vmatrix} + \begin{vmatrix} x & x & x \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{vmatrix} = 0.$$

11. A typical transportation problem involves shipping goods from production plants to warehouses, with several plants supplying the needs of one warehouse. Suppose two plants with 10 and 25 units of product available have to supply three warehouses needing 15, 15 and 5 units, respectively. If  $x_{ij}$  represents the number of units shipped from plant  $i$  to warehouse  $j$ , the amount of product available at each plant necessitates the equations

$$x_{11} + x_{12} + x_{13} = 10 \quad \text{and} \quad x_{21} + x_{22} + x_{23} = 25$$

and warehouse requirements necessitate the equations

$$x_{11} + x_{21} = 15, \quad x_{12} + x_{22} = 15, \quad x_{13} + x_{23} = 5.$$

Define

$$\mathbf{x}' = [x_{11} \quad x_{12} \quad x_{13} \quad x_{21} \quad x_{22} \quad x_{23}]$$

and

$$\mathbf{y}' = [10 \quad 25 \quad 15 \quad 15 \quad 5].$$

- (a) Write the preceding equations as  $\mathbf{Ax} = \mathbf{y}$ , and so determine  $\mathbf{A}$ .
- (b) Take any five square submatrices of  $\mathbf{A}$ , one each of order 5, 4, 3, 2 and 1, and show that the determinant of each submatrix is  $+1$ ,  $-1$  or zero. (This is a general characteristic of transportation problems of this nature, which leads to their having integer solutions for the  $x$ 's.)

12. (a) When  $\mathbf{K}$  has 12 rows and  $\mathbf{K}^5 = 3\mathbf{K}$ , what is a numerical value for  $|\mathbf{K}|$ ? Is there more than one possibility? If so, give all possible values.  
 (b) When all row sums of  $\mathbf{A}$  are  $s$ , explain why  $s$  is a factor of  $|\mathbf{A}|$ .  
 (c) Prove that for the partitioned square matrix  $\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}$

$$|\mathbf{X}|^2 = |\mathbf{A}\mathbf{A}' + \mathbf{B}\mathbf{B}'| = \begin{vmatrix} \mathbf{A}'\mathbf{A} & \mathbf{A}'\mathbf{B} \\ \mathbf{B}'\mathbf{A} & \mathbf{B}'\mathbf{B} \end{vmatrix}.$$

Construct a simple numerical example of order 4 or more.

13. For  $\mathbf{J}$  square of order  $n$  with all elements unity and for  $\mathbf{D}\{x_i\}$  a diagonal matrix of order  $n$  with diagonal elements  $x_1, x_2, \dots, x_n$ , all nonzero, show that

$$|\mathbf{J} + \mathbf{D}\{x_i\}| = \left( \prod_{i=1}^n x_i \right) \left( 1 + \sum_{i=1}^n 1/x_i \right).$$

14. For  $\mathbf{a}'$  of order  $n$  show that  $|\lambda\mathbf{I} + \mathbf{1}\mathbf{a}'| = \lambda^{n-1}(\lambda + \mathbf{a}'\mathbf{1})$ .  
 15. Prove  $|a\mathbf{I}_n + b\mathbf{J}_n| = a^{n-1}(a + nb)$ .  
 16. Without expanding the determinants show that

$$2 \begin{vmatrix} a & b & c \\ x & y & z \\ p & q & r \end{vmatrix} = \begin{vmatrix} a+b & b+c & c+a \\ x+y & y+z & z+x \\ p+q & q+r & r+p \end{vmatrix}$$

and that

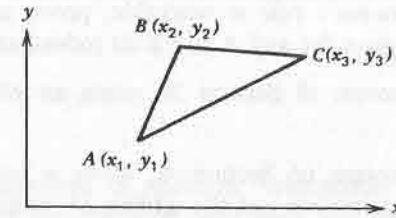
$$2 \begin{vmatrix} 0 & e & d \\ e & 0 & f \\ d & f & 0 \end{vmatrix} = \begin{vmatrix} d+e & f & f \\ d & e+f & d \\ e & e & f+d \end{vmatrix}.$$

17. Explain why  $\begin{vmatrix} \mathbf{P}' & \mathbf{X}' \\ \mathbf{0} & \mathbf{Q}' \end{vmatrix} = |\mathbf{P}'||\mathbf{Q}'|$ .

18. Suppose  $\mathbf{A}\mathbf{1} = \mathbf{0}$ , for square  $\mathbf{A}$ . Let  $\mathbf{c}_j$  be the vector of cofactors of elements of column  $j$  of  $|\mathbf{A}|$ . Prove that  $\mathbf{c}_j$  is the same for all  $j$ . [Hint: Partition  $\mathbf{A}$  to show the elements of its leading  $2 \times 2$  submatrix.]

19. Drop perpendiculars from  $A, B, C$  in the accompanying figure and by using the resulting trapezoids show that the area of the triangle  $ABC$  can be expressed as the absolute value of  $\frac{1}{2} |\mathbf{M}|$  where

$$\mathbf{M} = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}.$$



20. (a) The equation of a line in two co-ordinates can be written as  $1 + ax + by = 0$ . Show that the equation of a line passing through  $(x_1, y_1)$  and  $(x_2, y_2)$  is

$$\begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix} = 0.$$

- (b) The equation of a plane can be represented as  $1 + ax + by + cz = 0$ . Show that the equation of the plane through the three points  $(x_i, y_i, z_i)$  for  $i = 1, 2, 3$  is

$$\begin{vmatrix} 1 & x & y & z \\ 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \end{vmatrix} = 0.$$

21. Denote the equations

$$\left. \begin{aligned} x_1 + 2x_2 + 3x_3 &= 26 \\ 3x_1 + 7x_2 + 10x_3 &= 87 \\ 2x_1 + 11x_2 + 7x_3 &= 73 \end{aligned} \right\} \text{ by } \mathbf{Ax} = \mathbf{b}.$$

- (a) Solve the equation using successive elimination.  
 (b) Replace each column of  $\mathbf{A}$ , in turn, by  $\mathbf{b}$ . In replacing column  $j$ , call the resulting matrix  $\mathbf{A}_j$ . Verify that the solutions in (a) are

$$x_j = |\mathbf{A}_j| / |\mathbf{A}|.$$

This is known as Cramer's rule for solving linear equations.

- (c) What condition must  $|\mathbf{A}|$  satisfy for Cramer's rule to be workable?  
 (d) When Cramer's rule is workable, prove in general that the values it gives for any  $\mathbf{A}$  and  $\mathbf{b}$  do indeed satisfy  $\mathbf{Ax} = \mathbf{b}$ .
22. Prove the theorem of Section 3b using an elementary operator matrix.
23. Prove the theorem of Section 4c using a series of elementary operations that involve just the adding of multiples of a row to a row.

## CHAPTER 5

### INVERSE MATRICES

The arithmetic operations of addition, subtraction and multiplication in matrix algebra have already been dealt with, but division has not. Just as multiplication of matrices is more complicated than that of scalars, so is the reverse operation, division. In fact, in its usual sense, division does not exist in matrix algebra; the concept of "dividing" by  $\mathbf{A}$  is replaced by the concept of multiplying by a matrix called the inverse of  $\mathbf{A}$ .

#### 1. INTRODUCTION: SOLVING EQUATIONS

The inverse of a square matrix  $\mathbf{A}$  is a matrix whose product with  $\mathbf{A}$  is the identity matrix. It is denoted by the symbol  $\mathbf{A}^{-1}$ , and is read as " $\mathbf{A}$  to the (power of) minus one," as "the inverse of  $\mathbf{A}$ " or as " $\mathbf{A}$ -inverse." We introduce the idea of inverses by considering their use in solving linear equations.

**Illustration.** Suppose an entomologist counts 186 dead larvae in a laboratory experiment on the efficacy of a pesticide tested on 300 active larvae. He very easily calculates the fraction dead as  $186/300 = 0.62$ . Doing this is equivalent to solving the simple equation  $300x = 186$ , which can be done by multiplying both sides of it by  $1/300$  to get  $(1/300)300x = (1/300)186$ ; thus  $x = 186/300 = 0.62$ . And in general, where  $a$  and  $b$  are known values and  $a$  is not zero, the equation  $ax = b$  can be solved by multiplying both sides of it by  $1/a$  to give  $(1/a)ax = (1/a)b$  or

$$x = (1/a)b = a^{-1}b. \quad (1)$$

A more frequent problem is that of solving a set of several linear equations in several unknowns.



**Illustration.** Suppose  $m$  represents a genetic mean and  $a$  and  $d$  are measures of gene substitution and dominance, respectively. Equations for estimating these values from an experiment concerning the effect of a dwarfing gene on body weight in mice, based on Falconer (1981), are

$$\begin{aligned} m + a &= 14 \\ m + d &= 12 \\ m - a &= 6. \end{aligned} \quad (2)$$

Adding the first and last of these gives  $2m = 20$  or  $m = 10$ , so that from the first equation  $a = 14 - 10 = 4$  and from the second  $d = 12 - 10 = 2$ . This manipulation of individual equations in (2) cannot directly be represented in the same form as (1). But with the aid of matrix algebra the solution can be put in this form;  $x$  becomes a column vector of solutions,  $a^{-1}$  becomes what is known as the inverse of a matrix, for which the notation  $A^{-1}$  is used, and  $b$  becomes a column vector. The equations themselves are written as  $Ax = b$  and solved as  $x = A^{-1}b$ ; this form applies whether we are solving two simultaneous linear equations in two unknowns or 100 equations in 100 unknowns. No matter how many such equations there are, the symbolic form of their solution is always the same. The advantages so bestowed by matrix algebra are clearly manifold.

Observe that (2) can be written as a matrix equation

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 14 \\ 12 \\ 6 \end{bmatrix} \quad (3)$$

embodying the principles of multiplication and equality given in Chapter 2. Let us premultiply both sides of this equation by the matrix

$$\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix},$$

and get

$$\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 14 \\ 12 \\ 6 \end{bmatrix}.$$

Carrying out the multiplications gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 10 \\ 4 \\ 2 \end{bmatrix},$$

which is

$$\mathbf{I}_3 \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 10 \\ 4 \\ 2 \end{bmatrix}, \quad \text{or simply} \quad \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 10 \\ 4 \\ 2 \end{bmatrix}.$$

This is the solution  $m = 10$ ,  $a = 4$  and  $d = 2$  obtained previously. Hence we have found that the equations

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 14 \\ 12 \\ 6 \end{bmatrix}$$

have the solution

$$\begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 14 \\ 12 \\ 6 \end{bmatrix} = \begin{bmatrix} 10 \\ 4 \\ 2 \end{bmatrix}$$

where the  $3 \times 3$  matrix introduced on the right-hand side is so related to the  $3 \times 3$  matrix of the original equations that the product of the two matrices is an identity matrix:

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

Let us now write these matrices and vectors in symbols:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} m \\ a \\ d \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 14 \\ 12 \\ 10 \end{bmatrix}$$

and

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix}.$$

We have then solved the equation

$$\mathbf{Ax} = \mathbf{b} \quad (4)$$

as

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \quad (5)$$

where

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}. \quad (6)$$

Since  $\mathbf{I}$  is the "1" of matrix algebra, the matrix  $\mathbf{A}^{-1}$  is called the *reciprocal* of  $\mathbf{A}$ , or *inverse* of  $\mathbf{A}$ . We define it for the moment by (6), as a matrix which, when postmultiplied by  $\mathbf{A}$ , gives the identity matrix  $\mathbf{I}$ . With  $\mathbf{A}^{-1}$  so defined the solution for  $\mathbf{x}$  is given by (5).

Notice that this solution has not been derived from (4) by dividing both sides of (4) by  $\mathbf{A}$  but by premultiplying both sides by the matrix  $\mathbf{A}^{-1}$  so that

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b};$$

by using the definition  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$  this becomes  $\mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b}$  and hence  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ . At no stage has there been division by  $\mathbf{A}$ . The solution is obtained through premultiplication by a matrix  $\mathbf{A}^{-1}$  defined such that  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ . As stated at the outset, division in any ordinary sense is undefined in matrix algebra and has no meaning; i.e., the expression  $\mathbf{b}/\mathbf{A}$  has no meaning when  $\mathbf{A}$  is a matrix. Instead, we deal with inverse matrices and use them as multipliers, e.g.,  $\mathbf{A}^{-1}\mathbf{b}$ , where the product of a matrix and its inverse is an identity matrix.

Before considering the wider implications of the definition of  $\mathbf{A}^{-1}$  as so far derived, we give another example of its use in solving linear equations.

**Example.** It is suggested that the reader verify that when the equations

$$\begin{aligned} 2u - 2v - w &= 5 \\ u + v - 2w &= 1 \\ u - w &= 4 \end{aligned}$$

are written in matrix form, the solution obtained by premultiplying them by

$$\begin{bmatrix} -1 & -2 & 5 \\ -1 & -1 & 3 \\ -1 & -2 & 4 \end{bmatrix} \text{ is } \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 13 \\ 6 \\ 9 \end{bmatrix}.$$

If the matrix of coefficients of  $u$ ,  $v$  and  $w$  in the equations is denoted by  $\mathbf{A}$ ,

the matrix used as the premultiplier is the inverse  $\mathbf{A}^{-1}$ :

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & -1 \\ 1 & 1 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{bmatrix} -1 & -2 & 5 \\ -1 & -1 & 3 \\ -1 & -2 & 4 \end{bmatrix}.$$

It will be found that  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$  and if the equations are written as  $\mathbf{Ax} = \mathbf{b}$ , the solution is  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ .

The remainder of this chapter is devoted first to sharpening the definition of  $\mathbf{A}^{-1}$  as given by the equation  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$  and then to specifying how  $\mathbf{A}^{-1}$  is obtained from  $\mathbf{A}$ . When this is known, any set of simultaneous linear equations that have a unique solution can be put in the form  $\mathbf{Ax} = \mathbf{b}$  and solved as  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ , no matter how many such equations there are. Computational difficulties may sometimes arise, but the form of the solution and the procedure for obtaining it remain the same. Also, once  $\mathbf{A}^{-1}$  is understood, it is used in matrix algebra as a matrix in its own right. Indeed, this use is more widespread than in solving linear equations.

## 2. PRODUCTS EQUAL TO I

The concept of a matrix inverse has been established in the context of solving simultaneous linear equations. This is appropriate because it is a context where the inverse of a matrix is widely used. However, from a wider viewpoint there is a more basic question: "Is there a matrix  $\mathbf{L}$  whose products with the matrix  $\mathbf{A}$  are  $\mathbf{I}$ ?" There are three answers to this, depending on the characteristics of  $\mathbf{A}$ : (i) in some cases  $\mathbf{L}$  exists and is unique for a given  $\mathbf{A}$ ; (ii) sometimes numerous matrices  $\mathbf{L}$  exist for a particular  $\mathbf{A}$ , i.e.,  $\mathbf{L}$  exists but is not unique; and (iii) in some instances  $\mathbf{L}$  does not exist at all. Examples of these three situations are now given.

(i) If

$$\mathbf{A}_1 = \begin{bmatrix} 2 & 5 \\ 3 & 8 \end{bmatrix} \quad \text{and} \quad \mathbf{L}_1 = \begin{bmatrix} 8 & -5 \\ -3 & 2 \end{bmatrix}$$

$\mathbf{L}_1\mathbf{A}_1 = \mathbf{I}$ , and  $\mathbf{L}_1$  is the unique matrix  $\mathbf{L}$  for this particular  $\mathbf{A}$  such that  $\mathbf{LA} = \mathbf{I}$ . Furthermore,  $\mathbf{A}_1\mathbf{L}_1 = \mathbf{I}$ .

(ii) For

$$\mathbf{A}_2 = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 3 & -1 \end{bmatrix}$$

there is an infinite number of matrices  $L$  such that  $LA_2 = I$ ; e.g.,

$$L_2 = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 5 & 1 \end{bmatrix} \text{ and } L_2^* = \begin{bmatrix} 4 & 15 & 4 \\ 7 & 25 & 6 \end{bmatrix}.$$

(iii) When

$$A_3 = \begin{bmatrix} 0 & 3 & 7 \\ 0 & 2 & 5 \end{bmatrix}$$

there is no matrix  $L$  such that  $LA_3 = I$ , for the leading element of  $LA_3$  will always be zero and so the product matrix can never equal  $I$ .

One might notice a further characteristic of these three situations:

- (i)  $A_1L_1 = I$  as well as  $L_1A_1 = I$ .
- (ii)  $A_2L_2 \neq I$  even though  $L_2A_2 = I$ .
- (iii) There are matrices  $R$  such that  $A_3R = I$  even though there is no matrix  $L$  such that  $LA_3 = I$ ; e.g.,

$$R = \begin{bmatrix} 4 & 8 \\ 5 & -7 \\ -2 & 3 \end{bmatrix}.$$

The  $A^{-1}$  established in each example of solving linear equations was a matrix such that  $A^{-1}A = I$ . But we have just illustrated how  $LA = I$  does not necessarily yield a matrix  $L$  that is unique for a given  $A$ , nor one for which  $AL$  necessarily equals  $I$  even if  $LA$  does. Only in the case of  $A_1$  did  $L_1$  have both these properties; and they seem reasonable properties to ask of an inverse—that it be unique for a given  $A$ , and that both premultiplication and postmultiplication by  $A$  lead to the identity matrix. Analogous to the familiar reciprocal of a number in scalar algebra, these properties seem appropriate for the inverse (reciprocal) of a matrix. Accordingly, when possible, we will derive a matrix from  $A$  which has these properties. The derived matrix will be the inverse of  $A$ , and instead of satisfying only  $A^{-1}A = I$  it will have the two properties

$$A^{-1}A = AA^{-1} = I \quad (a)$$

and

$$A^{-1} \text{ unique for given } A. \quad (b)$$

The equality of  $A^{-1}A$  and  $AA^{-1}$  in (a) requires that both products exist and have the same order. As explained in Section 2.7d, this occurs only

when  $A$  and  $A^{-1}$  are both square, of the same order. This means that an  $A^{-1}$  such that (a) is true exists only when  $A$  is square. What we are about to describe as the inverse of a matrix can therefore exist only if that matrix is square (and its inverse will also be square, of the same order). In contrast, rectangular matrices do not have inverses, although some of them have what is called either a left or right inverse (see Section 10) and all of them have a generalized inverse (see Chapter 8).

Having established that the inverse  $A^{-1}$  can exist only when  $A$  is square, observe that it can therefore exist only when  $A$  has a determinant. In consequence of this the inverse is now derived in a manner that utilizes the determinant. This form of derivation has been chosen because it is particularly suitable for describing how the elements of the inverse of a matrix are related to, and can be obtained directly from, those of the matrix itself. However, it is tedious in operation in practically all instances except those of small-sized matrices and is therefore seldom used for calculating the inverse of a matrix in practical situations. Nevertheless, it is a method that presents with clarity the underlying form of the elements of an inverse; furthermore, it readily leads to demonstrating properties (a) and (b) just discussed. These and others are dealt with following the derivation. Brief discussion is also given to some arithmetic problems in calculating inverses.

### 3. COFACTORS OF A DETERMINANT

The cofactors of a determinant are defined in Section 4.3c. They are essential to understanding the functional relationship between the elements of  $A$  and those of  $A^{-1}$ . Equations (5) and (6) of Section 4.3c concern cofactors of elements of a row of  $A$ . Since the same results hold true for  $A'$  and  $|A| = |A'|$ , they can also be written in terms of columns of  $A$  and so we have

$$|A| = \sum_{i=1}^n a_{ij}c_{ij} \text{ for any } j \quad (7)$$

and

$$0 = \sum_{i=1}^n a_{ij}c_{ih} \text{ for } j \neq h. \quad (8)$$

### 4. DERIVATION OF THE INVERSE

The inverse of a square matrix is now derived from properties of the cofactors of elements of its determinant. The derivation is in terms of the

numerical example and involves the cofactors of all the elements of the matrix—column by column, rather than row by row.

Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}.$$

First we derive the cofactors of each column in turn and demonstrate the properties just discussed. The cofactors of the elements of the first column are, respectively,

$$\begin{aligned} & (-1)^{1+1} \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}, \quad (-1)^{2+1} \begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} \quad \text{and} \quad (+1)^{3+1} \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix}, \\ & = 2, 4 \text{ and } -3. \end{aligned} \quad (9)$$

Adding the products of these cofactors with the elements of each column of  $A$  gives, in turn:

$$2(1) + 4(4) - 3(7) = -3 = |A|, \quad (10)$$

$$2(2) + 4(5) - 3(8) = 0 \quad (11)$$

$$2(3) + 4(6) - 3(10) = 0. \quad (12)$$

Similarly, the cofactors of elements of the second column are

$$\begin{aligned} & (-1) \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}, \quad (+1) \begin{vmatrix} 1 & 3 \\ 7 & 10 \end{vmatrix} \quad \text{and} \quad (-1) \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}, \\ & = 2, -11 \text{ and } 6. \end{aligned} \quad (13)$$

and the corresponding sums of products with the elements of the three columns are:

$$2(1) - 11(4) + 6(7) = 0, \quad (14)$$

$$2(2) - 11(5) + 6(8) = -3 = |A|, \quad (15)$$

$$2(3) - 11(6) + 6(10) = 0. \quad (16)$$

Finally, the cofactors of the elements of the third column are:

$$\begin{aligned} & (+1) \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}, \quad (-1) \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} \quad \text{and} \quad (+1) \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}, \\ & = -3, 6 \text{ and } -3. \end{aligned} \quad (17)$$

Sums of products of these with the elements of the columns of  $A$  are:

$$-3(1) + 6(4) - 3(7) = 0, \quad (18)$$

$$-3(2) + 6(5) - 3(8) = 0, \quad (19)$$

$$-3(3) + 6(6) - 3(10) = -3 = |A|. \quad (20)$$

Now consider the matrix obtained by replacing the elements of  $A$  by their cofactors—i.e., from

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \quad \text{obtaining} \quad \begin{bmatrix} 2 & 2 & -3 \\ 4 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}.$$

The columns of the new matrix are the values in (9), (13) and (17). Let us transpose this newly formed matrix and multiply it by the scalar  $1/-3 = 1/|A|$ . This gives

$$\frac{1}{-3} \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}. \quad (21)$$

Now premultiply  $A$  by the matrix we have just derived:

$$\frac{1}{-3} \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}.$$

The elements in the rows of this product are the left-hand sides of equations (10), (11), (12); (14), (15), (16); and (18), (19), (20), respectively. Therefore the product is

$$\frac{1}{-3} \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In this instance, when  $\mathbf{A}$  is premultiplied by the matrix derived in (21), the product is the identity matrix  $\mathbf{I}$ . Thus (21) is the kind of matrix just discussed, namely  $\mathbf{A}^{-1}$  such that  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ . So far this is the sole property of  $\mathbf{A}^{-1}$ . We have not shown that for  $\mathbf{A}^{-1}$  so obtained  $\mathbf{A}\mathbf{A}^{-1}$  also equals  $\mathbf{I}$ , nor have we shown that it is the unique matrix for which  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ . These are properties (a) and (b) mentioned earlier. We shall discuss them shortly, but first let us reconstruct in general terms the derivation of  $\mathbf{A}^{-1}$  as obtained in (21).

Initially we had  $\mathbf{A}$  as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

and formed a new matrix by replacing each element of  $\mathbf{A}$  by its cofactor:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}.$$

This was transposed, giving

$$\begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \quad (22)$$

and multiplied by the scalar  $1/|\mathbf{A}|$  to produce

$$\frac{1}{|\mathbf{A}|} \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} = \mathbf{A}^{-1},$$

the inverse of  $\mathbf{A}$ . Premultiplying  $\mathbf{A}$  by this matrix gives the identity matrix because in the product each diagonal element is a sum of products of elements of a column of  $\mathbf{A}$  with their cofactors in  $|\mathbf{A}|$ , and therefore equal to  $|\mathbf{A}|$  by (7); the scalar  $1/|\mathbf{A}|$  reduces them to 1. And the off-diagonal elements of the product matrix are sums of products of elements of columns of  $\mathbf{A}$  with cofactors in  $|\mathbf{A}|$  of elements of other columns, and therefore equal to zero by (8). Hence the product is  $\mathbf{I}$  and so  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ . (The product  $\mathbf{A}\mathbf{A}^{-1}$  is considered later.)

The matrix (22), namely  $\mathbf{A}$  with its elements replaced by their cofactors and then transposed, is called the *adjugate*, or sometimes the *adjoint*, matrix of  $\mathbf{A}$ . Thus the inverse,  $\mathbf{A}^{-1}$ , can be described as the adjugate of  $\mathbf{A}$

multiplied by the scalar  $1/|\mathbf{A}|$ :

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \text{adj } \mathbf{A},$$

where  $\text{adj } \mathbf{A}$  is the adjugate (or adjoint) matrix of  $\mathbf{A}$ . Hence

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \left[ \begin{array}{l} \mathbf{A} \text{ with every element} \\ \text{replaced by its cofactor} \end{array} \right]^{\text{transposed}}$$

Note that the adjugate of  $\mathbf{A}$  is the transpose of the matrix of cofactors:

$$\text{adj } \mathbf{A} = \{c_{ij}\}' = \{c_{ji}\}.$$

It exists whether  $|\mathbf{A}| = 0$  or not, and always satisfies

$$(\text{adj } \mathbf{A})\mathbf{A} = \mathbf{A}(\text{adj } \mathbf{A}) = |\mathbf{A}|\mathbf{I};$$

when  $|\mathbf{A}| \neq 0$  and  $\mathbf{A}^{-1}$  exists

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

**Example.** The determinant of

$$\mathbf{A} = \begin{bmatrix} 2 & 5 \\ 3 & 9 \end{bmatrix} \quad \text{is} \quad |\mathbf{A}| = \begin{vmatrix} 2 & 5 \\ 3 & 9 \end{vmatrix} = 18 - 15 = 3.$$

The adjugate matrix is  $\begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix}$ , and so the inverse is

$$\mathbf{A}^{-1} = \frac{1}{3} \begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix}.$$

## 5. CONDITIONS FOR EXISTENCE OF THE INVERSE

The inverse  $\mathbf{A}^{-1}$  of a matrix  $\mathbf{A}$ , such that  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ , has been developed in the preceding paragraph as the adjugate matrix of  $\mathbf{A}$  multiplied by the scalar  $1/|\mathbf{A}|$ . Two conditions imposed on  $\mathbf{A}$  in order for this to exist are:

- (i)  $\mathbf{A}^{-1}$  can exist only when  $\mathbf{A}$  is square, as discussed in Section 2.
- (ii)  $\mathbf{A}^{-1}$  does exist only if  $|\mathbf{A}|$  is nonzero. (If  $|\mathbf{A}|$  is zero the scalar factor  $1/|\mathbf{A}|$  in  $\mathbf{A}^{-1}$  is infinite and  $\mathbf{A}^{-1}$  does not exist.  $|\mathbf{A}|$  must therefore be nonzero for  $\mathbf{A}^{-1}$  to exist.)

A square matrix is said to be *singular* when its determinant is zero and *nonsingular* when its determinant is nonzero. Singularity is therefore a property of square matrices only, not of rectangular matrices, and it is only nonsingular matrices that have inverses. Just as conformability is necessary for the existence of a matrix product, so is nonsingularity necessary for the existence of a matrix inverse. In both instances the necessary condition is sometimes not mentioned explicitly but it must always be satisfied.

## 6. PROPERTIES OF THE INVERSE

If  $\mathbf{A}$  is a square, nonsingular matrix its inverse,  $\mathbf{A}^{-1}$ , has the following properties.

- (i) The inverse commutes with  $\mathbf{A}$ , both products being the identity matrix:  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ .
- (ii) The inverse of  $\mathbf{A}$  is unique.
- (iii) The determinant of the inverse of  $\mathbf{A}$  is the reciprocal of the determinant of  $\mathbf{A}$ :  $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$ .
- (iv) The inverse matrix is nonsingular.
- (v) The inverse of  $\mathbf{A}^{-1}$  is  $\mathbf{A}$ :  $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ .
- (vi) The inverse of the transpose is the transpose of the inverse:

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

- (vii) If  $\mathbf{A}$  is symmetric, so is its inverse: if  $\mathbf{A}' = \mathbf{A}$ , then  $(\mathbf{A}^{-1})' = \mathbf{A}^{-1}$ .
- (viii) The inverse of a product is the product of the inverses taken in reverse order, provided the inverses exist: if  $\mathbf{A}^{-1}$  and  $\mathbf{B}^{-1}$  exist,

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

The first two of the foregoing have been referred to earlier as (a) and (b). Proofs of these properties are now indicated; examples are left to the reader.

(i) It has been shown that for  $\mathbf{A}^{-1}$  as developed,  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ . This is because the elements of the product  $\mathbf{A}^{-1}\mathbf{A}$  are sums of products of elements of columns of  $\mathbf{A}$  multiplied by cofactors of elements of the same and different columns of  $\mathbf{A}$ . In like fashion the product  $\mathbf{A}\mathbf{A}^{-1}$  has elements that are sums of products of elements of rows of  $\mathbf{A}$  multiplied by cofactors of elements of the same and different rows of  $\mathbf{A}$ . Hence  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ .

An alternative proof is to assume there is some matrix  $\mathbf{P}$ , possibly different from  $\mathbf{A}^{-1}$ , such that  $\mathbf{AP} = \mathbf{I}$ . Premultiplying by  $\mathbf{A}^{-1}$  gives  $\mathbf{A}^{-1}\mathbf{AP}$

$= \mathbf{A}^{-1}$  and so, because  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ , we have  $\mathbf{IP} = \mathbf{A}^{-1}$ , or  $\mathbf{P} = \mathbf{A}^{-1}$ ; i.e.,  $\mathbf{P}$  does equal  $\mathbf{A}^{-1}$ . Therefore  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ .

(ii) Suppose that  $\mathbf{A}^{-1}$  is not a unique inverse of  $\mathbf{A}$  and that  $\mathbf{S}$  is another inverse possibly different from  $\mathbf{A}^{-1}$ , such that  $\mathbf{SA} = \mathbf{I}$ . Then postmultiplying  $\mathbf{SA} = \mathbf{I}$  by  $\mathbf{A}^{-1}$  gives  $\mathbf{SAA}^{-1} = \mathbf{IA}^{-1} = \mathbf{A}^{-1}$  and because, from (i),  $\mathbf{AA}^{-1} = \mathbf{I}$  this means  $\mathbf{S} = \mathbf{A}^{-1}$ . Hence  $\mathbf{A}^{-1}$  is the unique inverse of  $\mathbf{A}$ .

(iii and iv) It was shown in Section 4.3e(iii) that if  $\mathbf{A}$  and  $\mathbf{B}$  are two square matrices of the same order,  $|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}|$ . Therefore

$$|\mathbf{A}||\mathbf{A}^{-1}| = |\mathbf{AA}^{-1}| = |\mathbf{I}| = 1, \quad \text{and so} \quad |\mathbf{A}^{-1}| = 1/|\mathbf{A}|.$$

The nonsingularity of  $\mathbf{A}^{-1}$  follows from this.

(v, vi and vii) Consider the identity  $\mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$ . Premultiplying it by  $(\mathbf{A}^{-1})^{-1}$  leads to the result  $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ ; transposing it leads to  $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$ . For  $\mathbf{A}' = \mathbf{A}$  the latter becomes  $\mathbf{A}^{-1} = (\mathbf{A}^{-1})'$ .

(viii) Provided  $\mathbf{A}$  and  $\mathbf{B}$  are both square, of the same order and nonsingular, it is clear that

$$\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{IB} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I},$$

and so, by postmultiplying by  $(\mathbf{AB})^{-1}$ , we have  $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ . The reversal rule that applies to transposing a product therefore applies also to inverting it.

## 7. SOME SIMPLE SPECIAL CASES

Certain matrices have inverses that are related to the matrices themselves in a simple manner. The following examples are particularly useful.

### a. Inverses of order 2

The inverse of a general  $2 \times 2$  matrix is easy:

$$\mathbf{A} = \begin{bmatrix} a & x \\ y & b \end{bmatrix} \quad \text{has} \quad \mathbf{A}^{-1} = \frac{1}{ab - xy} \begin{bmatrix} b & -x \\ -y & a \end{bmatrix} \quad \text{for} \quad ab - xy \neq 0. \quad (23)$$

As well as dividing by the determinant  $|\mathbf{A}| = ab - xy$ , which must not be zero, one simply interchanges the diagonal elements and changes the sign of the off-diagonal elements.

**b. Diagonal matrices**

The inverse of a diagonal matrix is also easy (providing all diagonal elements are nonzero):

$$(\mathbf{D}\{x_i\})^{-1} = \mathbf{D}\{1/x_i\} \quad \text{when } x_i \neq 0. \quad (24)$$

For example,

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}.$$

**c. I and J matrices**

Special cases of (24) are identity matrices, for which  $\mathbf{I}^{-1} = \mathbf{I}$ . On the other hand,  $\mathbf{J}$  matrices have  $|\mathbf{J}| = 0$ , and so have no inverses, but for  $a \neq 0$  and  $a + nb \neq 0$

$$(a\mathbf{I}_n + b\mathbf{J}_n)^{-1} = \frac{1}{a} \left( \mathbf{I}_n - \frac{b}{a + nb} \mathbf{J}_n \right). \quad (25)$$

**d. Orthogonal matrices**

In Section 3.4 we described how a matrix  $\mathbf{P}$  having orthonormal rows has  $\mathbf{P}\mathbf{P}' = \mathbf{I}$ ; and if it is an orthogonal matrix it is also square. Further, at the end of Section 4.3 we showed that when  $\mathbf{P}$  is orthogonal,  $|\mathbf{P}| \neq 0$ . Hence orthogonal  $\mathbf{P}$  is square and has nonzero determinant. Therefore  $\mathbf{P}^{-1}$  exists. This being so, premultiplying  $\mathbf{P}\mathbf{P}' = \mathbf{I}$  by  $\mathbf{P}^{-1}$  gives

$$\mathbf{P}^{-1} = \mathbf{P}' \quad (26)$$

and then postmultiplying (26) by  $\mathbf{P}$  gives  $\mathbf{P}'\mathbf{P} = \mathbf{I}$ ; i.e., the columns of an orthogonal matrix are orthonormal. Thus an orthogonal matrix  $\mathbf{P}$  has the following properties: (i) it is square; (ii)  $|\mathbf{P}| = \pm 1$ ; (iii) its rows are orthonormal, meaning that  $\mathbf{P}\mathbf{P}' = \mathbf{I}$ ; and (iv) its columns are orthonormal,  $\mathbf{P}'\mathbf{P} = \mathbf{I}$ . Any combination of (i) or (ii) with (iii) or (iv) is necessary and sufficient for the other two of those four to hold, and for  $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$  to hold also.

**Example.** The reader can demonstrate these properties for

$$\mathbf{P} = \frac{1}{13} \begin{bmatrix} 5 & -14 & 2 \\ -10 & -5 & -10 \\ 10 & 2 & -11 \end{bmatrix}.$$

**e. Idempotent matrices**

The only idempotent matrices that are nonsingular are identity matrices. But idempotent matrices can involve nonsingular matrices; e.g.,  $\mathbf{P}(\mathbf{Q}\mathbf{P})^{-1}\mathbf{Q}$  is idempotent when  $(\mathbf{Q}\mathbf{P})^{-1}$  exists. (Verification is left to the reader.)

## 8. EQUATIONS AND ALGEBRA

Two widespread uses of inverse matrices are in the solution of linear equations and, more generally, in the simplification of algebraic expressions.

**a. Solving linear equations**

The concept of an inverse matrix is introduced at the beginning of this chapter in terms of solving equations  $\mathbf{A}\mathbf{x} = \mathbf{b}$  as  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$  when  $\mathbf{A}^{-1}$  exists. This is one of the most important uses for inverse matrices. We offer two illustrations.

(i) *Age distributions in wild populations.* A problem that occurs in studying the dynamics of wild populations is that of assessing age distributions. Suppose for some animal population being studied, deer say, that there are  $n$  age groupings and the proportion of the population in age group  $i$  is  $r_i$  for  $i = 1, \dots, n$ . For a population of wild deer, these values can never be known exactly because not all deer can be located, nor can their ages be known with exactitude. But the ages of animals that can be located may be estimated, from which estimates  $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_n$  of the population proportions can be calculated. Deer whose exact ages are known might also be available, either in the form of a colony born and reared in captivity, or of deer born in captivity, tagged with their birth date, released and subsequently recaptured. From such animals one can assess the efficiency of personnel whose normal task it is to estimate the age of deer captured from the wild population, achieving this by having such personnel estimate the ages of a number of deer whose true ages are available. In this way it is possible to put a value on  $p_{ij}$ , the probability that an animal estimated as of age  $i$  is truly of age  $j$ . Then, for example, the estimated fraction in age 1 in the population,  $\hat{r}_1$ , is the sum of the true fractions  $r_1, r_2, \dots, r_n$  each multiplied by the probability that such an animal is estimated as of age 1. Thus

$$\hat{r}_1 = p_{11}r_1 + p_{12}r_2 + \dots + p_{1n}r_n,$$

and for all age groups together

$$\begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \vdots \\ \hat{r}_n \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}.$$

Hence if  $\hat{r}$  denotes the vector of estimated proportions,  $r$  the vector of true proportions and  $P$  the matrix of probabilities,  $\hat{r} = Pr$ . Improved estimates of the proportions can be obtained by solving this equation for  $r$  using the inverse of  $P$ , namely  $r = P^{-1}\hat{r}$ .

(ii) *Input-output analysis in economics.* Input-output analysis is a method of analyzing interdependencies of various sectors of an economy. Consider an economy with  $n$  sectors, each producing a single commodity. Let  $x_i$  = gross output of commodity  $i$ ,  $a_{ij}$  = amount of commodity  $i$  used in producing one unit of commodity  $j$ , and  $y_i$  = final demand for commodity  $i$ . The gross product for each commodity must equal the sum of the amounts of that commodity used in producing all the commodities, plus the final demand for the commodity itself:

$$x_i = \sum_{j=1}^n a_{ij}x_j + y_i \quad \text{for } i = 1, 2, \dots, n.$$

In matrix notation,  $x = Ax + y$ , where  $x = \{x_i\}$ ,  $y = \{y_i\}$  and  $A = \{a_{ij}\}$  for  $i, j = 1, 2, \dots, n$ .  $A$  is called the input-output coefficient matrix; its elements are non-negative, as are those of  $x$  and  $y$ , since the equation in  $x$  can be rearranged as  $(I - A)x = y$ . Its solution, provided  $(I - A)$  is nonsingular, is

$$x = (I - A)^{-1}y. \quad (27)$$

**Illustration.** Consider a simple four-sector economy with two industries (agriculture and manufacturing), one primary factor of production (labor) and a government sector that consumes all net output from both industries and also utilizes labor; i.e., the government produces nothing for the economy and its consumption from the industry and labor sectors represents the final demand for the commodities produced by these sectors. Each industry uses output from the other industry as well as labor, and the labor force requires output from both industries as well as labor services to sustain itself. Labor can be freely imported and exported, so that there is never any unemployment or excess demand for labor, and capital stock and inventories are held constant through time.

Suppose the flow of goods among the sectors of the economy for a particular period is as shown in Table 5.1. Adding across each row gives the total output of each industry and the total labor force employed. Since output is in physical units, adding down columns is meaningless. Yet any column taken as a whole is meaningful; it shows the inputs a given sector requires in order to produce its total output. For example, the second

TABLE 5.1. INPUT-OUTPUT TABLE

Production Sector	Consumption Sector				Total
	Agriculture	Manufacturing	Labor	Final Demand by Government	
Agriculture (tons)	600	400	1400	600	3000
Manufacturing (machines)	1500	800	700	1000	4000
Labor (employees)	900	4800	700	600	7000

column shows that the manufacturing industry, in order to produce the 4000 machines shown as the total of the second row, requires 400 tons of agricultural output, 800 machines and 4800 employees. Therefore, on dividing each of these requirements by the total production, 4000, we get the elements  $a_{i2}$  of  $A$ , the amount of commodity  $i$  used in producing one unit of commodity 2. For example,  $a_{12} = 400/4000 = 0.1$ : it takes 0.1 ton of agricultural output to produce one machine. In this way the input-output coefficient matrix is

$$A = \begin{bmatrix} \frac{600}{3000} & \frac{400}{4000} & \frac{1400}{7000} \\ \frac{1500}{3000} & \frac{800}{4000} & \frac{700}{7000} \\ \frac{900}{3000} & \frac{4800}{4000} & \frac{700}{7000} \end{bmatrix} = \begin{bmatrix} .2 & .1 & .2 \\ .5 & .2 & .1 \\ .3 & 1.2 & .1 \end{bmatrix}$$

and

$$(I - A)^{-1} = \frac{1}{0.264} \begin{bmatrix} .60 & .33 & .17 \\ .48 & .66 & .18 \\ .84 & .99 & .59 \end{bmatrix}.$$

We can now check that the figures in Table 5.1 are consistent; with  $x' = [3000 \ 4000 \ 7000]$  and  $y' = [600 \ 1000 \ 600]$  it will be found that (27) is satisfied:

$$x = (I - A)^{-1}y = \frac{1}{0.264} \begin{bmatrix} .60 & .33 & .17 \\ .48 & .66 & .18 \\ .84 & .99 & .59 \end{bmatrix} \begin{bmatrix} 600 \\ 1000 \\ 600 \end{bmatrix} = \begin{bmatrix} 3000 \\ 4000 \\ 7000 \end{bmatrix}.$$



Now suppose that in the same economy government demands become 1100 tons of agricultural produce, 1200 machines and 792 employees. Then  $y' = [1100 \ 1200 \ 792]$  and the corresponding outputs needed to satisfy these demands are given by the value of  $x$  corresponding to this  $y$ :

$$x = (\mathbf{I} - \mathbf{A})^{-1}y = \frac{1}{0.264} \begin{bmatrix} .60 & .33 & .17 \\ .48 & .66 & .18 \\ .84 & .99 & .59 \end{bmatrix} \begin{bmatrix} 1100 \\ 1200 \\ 792 \end{bmatrix} = \begin{bmatrix} 4510 \\ 5540 \\ 9770 \end{bmatrix}.$$

Thus to meet the new demands, 4510 tons of agricultural output and 5540 machines will have to be produced, and 9770 employees will be required.

Most input-output tables are expressed in monetary amounts. When physical units have been used (as in Table 5.1), they can be converted to a monetary basis by giving each a value, as in Exercise 19. A variation on this is to value each commodity differently for each user who consumes it. For example, in Table 5.1, agriculturalists and manufacturers may buy different machines at different prices. An illustration is given in Exercise 20.

(iii) *Least squares equations.* An important application of linear equations is that of least squares equations in statistics, typically  $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ . These are discussed in Chapters 9, 13, 14 and 15.

#### b. Algebraic simplifications

Division by a matrix never occurs. Multiplication by the corresponding inverse (when it exists) is used in its place. But, as usual, distinction must always be made between premultiplication and postmultiplication. For example, suppose  $\mathbf{A}^{-1}$  exists. Then if

$$\mathbf{AX} = \mathbf{B}, \quad (28)$$

premultiplying this equation by  $\mathbf{A}^{-1}$  gives

$$\mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{B}, \quad \text{i.e., } \mathbf{IX} = \mathbf{A}^{-1}\mathbf{B}, \quad \text{or } \mathbf{X} = \mathbf{A}^{-1}\mathbf{B}. \quad (29)$$

But postmultiplication by  $\mathbf{A}^{-1}$  (provided  $\mathbf{X}$  is square) gives just

$$\mathbf{AXA}^{-1} = \mathbf{BA}^{-1}, \quad (30)$$

which cannot be simplified (other than back to  $\mathbf{AX} = \mathbf{B}$ ). Unlike  $axa^{-1} = x$ , when  $a$  and  $x$  are scalars and so  $a$  and  $a^{-1}$  cancel out, in  $\mathbf{AXA}^{-1}$  the  $\mathbf{A}$  and  $\mathbf{A}^{-1}$  do not cancel and  $\mathbf{AXA}^{-1}$  is simply that, just  $\mathbf{AXA}^{-1}$ .

The nonexistence of division, and the existence of inverse matrices only in special cases, also means that in matrix algebra an equation such as  $\mathbf{PK} = \mathbf{QK}$  does not always imply that  $\mathbf{P}$  equals  $\mathbf{Q}$ . An example of this is

$$\begin{bmatrix} 7 & 5 & 4 \\ 3 & 3 & 10 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ -1 & 0 & 1 \\ 0 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 8 \\ 2 & 2 & 11 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ -1 & 0 & 1 \\ 0 & 2 & 3 \end{bmatrix} \quad (31)$$

The reason that the equality (31) does not imply equality of the two  $2 \times 3$  matrices (which would clearly be false) is that the  $3 \times 3$  matrix has no inverse. It is *only* the existence of  $\mathbf{K}^{-1}$  that permits the equation  $\mathbf{PK} = \mathbf{QK}$  to imply  $\mathbf{P} = \mathbf{Q}$ . This is so because when  $\mathbf{K}^{-1}$  exists, the equation

$$\mathbf{PK} = \mathbf{QK}$$

can be postmultiplied on both sides by  $\mathbf{K}^{-1}$  to yield

$$\mathbf{PKK}^{-1} = \mathbf{QKK}^{-1}, \quad \text{i.e., } \mathbf{PI} = \mathbf{QI}, \quad \text{or } \mathbf{P} = \mathbf{Q}. \quad (32)$$

But when  $\mathbf{K}^{-1}$  does not exist, as in (31), then (32) cannot be used.

The existence of matrix inverses often leads to algebraic equalities that might not otherwise be apparent. For example, it is easy to see that

$$\mathbf{R} + \mathbf{RST} = \mathbf{R}(\mathbf{I} + \mathbf{ST}),$$

but if  $\mathbf{T}^{-1}$  exists this can further be written as

$$\mathbf{R} + \mathbf{RST} = \mathbf{R}(\mathbf{T}^{-1} + \mathbf{S})\mathbf{T}. \quad (33)$$

On multiplying out the right-hand side of (33) one sees at once that (33) is correct. The useful thing to realize is that, in starting from the left-hand side of (33) one can, if  $\mathbf{T}^{-1}$  is known to exist, develop the right-hand side.

A further example of the algebra of inverses is a matrix analogue of the scalar result that for scalar  $x \neq 1$

$$1 + x + x^2 + \cdots + x^{n-1} = (x^n - 1)/(x - 1).$$

A matrix counterpart is that, provided  $(\mathbf{X} - \mathbf{I})^{-1}$  exists

$$\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1} = (\mathbf{X}^n - \mathbf{I})(\mathbf{X} - \mathbf{I})^{-1}. \quad (34)$$

This is established by noting that the product

$$\begin{aligned} & (\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1})(\mathbf{X} - \mathbf{I}) \\ &= \mathbf{X} + \mathbf{X}^2 + \mathbf{X}^3 + \cdots + \mathbf{X}^{n-1} + \mathbf{X}^n - (\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1}) \\ &= \mathbf{X}^n - \mathbf{I} \end{aligned} \quad (35)$$

and therefore if  $(\mathbf{X} - \mathbf{I})^{-1}$  exists, postmultiplication of both sides of (35) yields (34). The similar procedure for establishing

$$\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1} = (\mathbf{X} - \mathbf{I})^{-1}(\mathbf{X}^n - \mathbf{I})$$

is left as an exercise for the reader.

**Illustration.** Alling (1958) defines six states of pulmonary tuberculosis diagnosis relative to the disease being active or arrested. The transition probability matrix (Section 1.6) of the probabilities that an individual diagnosed in one state of the disease will have changed to another state at the next diagnosis he writes as the partitioned matrix (Section 2.2)

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{B} \end{bmatrix},$$

where  $\mathbf{C}$  and  $\mathbf{B}$  take particular forms based on description of the six states of the disease. The matrix of probabilities after two diagnoses (made at equal intervals of time) then depends on

$$\mathbf{P}^2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} + \mathbf{BC} & \mathbf{B}^2 \end{bmatrix},$$

and that after  $k$  diagnoses depends on

$$\mathbf{P}^k = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^k$$

which, by use of (28), is

$$\mathbf{P}^k = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^k)\mathbf{C} & \mathbf{B}^k \end{bmatrix}.$$

For large  $k$  this form of  $\mathbf{P}^k$  is more readily calculated than is  $\mathbf{P}^k$  directly. Also, it shows the relationship of  $\mathbf{P}^k$  to  $\mathbf{B}$  and  $\mathbf{C}$  and provides easy consideration of the limiting form of  $\mathbf{P}^k$  for infinitely large  $k$ .

The technique of establishing (34) through first deriving (35) and then postmultiplying by  $(\mathbf{X} - \mathbf{I})^{-1}$  is very powerful. Many results using inverses are established in this manner, particularly those of showing that some matrix inverse equals some other matrix. For example, suppose we want to show that  $\mathbf{A}^{-1}$  is some given matrix  $\mathbf{Q}$ . Provided we are told or can assume that  $\mathbf{A}^{-1}$  exists (i.e., that  $\mathbf{A}$  is square with nonzero determinant), we therefore anticipate  $\mathbf{AQ}$  to be an identity matrix. Hence on postmultiplying  $\mathbf{A}$  by  $\mathbf{Q}$ , we endeavor (hope) to simplify the product until it is an identity matrix. We will then have  $\mathbf{AQ} = \mathbf{I}$ , and so  $\mathbf{A}^{-1}\mathbf{AQ} = \mathbf{A}^{-1}$  and hence  $\mathbf{Q} = \mathbf{A}^{-1}$ .

**Example.** To prove that  $(\mathbf{I} + \mathbf{M}^{-1})^{-1}$  is  $\mathbf{M}(\mathbf{M} + \mathbf{I})^{-1}$  we hope to show that  $(\mathbf{I} + \mathbf{M}^{-1})[\mathbf{M}(\mathbf{M} + \mathbf{I})^{-1}]$  is an identity matrix; indeed it is, because

$$\begin{aligned} (\mathbf{I} + \mathbf{M}^{-1})[\mathbf{M}(\mathbf{M} + \mathbf{I})^{-1}] &= (\mathbf{I} + \mathbf{M}^{-1})\mathbf{M}(\mathbf{M} + \mathbf{I})^{-1} \\ &= (\mathbf{M} + \mathbf{M}^{-1}\mathbf{M})(\mathbf{M} + \mathbf{I})^{-1} \\ &= (\mathbf{M} + \mathbf{I})(\mathbf{M} + \mathbf{I})^{-1} = \mathbf{I}. \end{aligned}$$

Now, since we set out to prove something about the inverse of  $\mathbf{I} + \mathbf{M}^{-1}$ , we need to either prove or assume that it has a nonzero determinant. In the absence of evidence to prove this we assume it, and then make the step that is crucial to the logic of developing our desired result, of premultiplying the preceding equation by  $(\mathbf{I} + \mathbf{M}^{-1})^{-1}$  to get

$$(\mathbf{I} + \mathbf{M}^{-1})^{-1}(\mathbf{I} + \mathbf{M}^{-1})[\mathbf{M}(\mathbf{M} + \mathbf{I})^{-1}] = (\mathbf{I} + \mathbf{M}^{-1})^{-1}\mathbf{I}.$$

This simplifies to what we want:

$$\mathbf{M}(\mathbf{M} + \mathbf{I})^{-1} = (\mathbf{I} + \mathbf{M}^{-1})^{-1}.$$

## 9. COMPUTERS AND INVERSES

The inverse of a matrix has been developed in terms of determinants because they provide an exact connection between elements of a matrix and those of its inverse (when it exists). But using determinants is computationally tedious, as readers will soon discover on inverting matrices even of small order, say 4 or 5. Fortunately, more efficient methods are used in computer programs that invert matrices. It is beyond the scope of this book

to give details of these methods, for which the reader is referred to Stewart (1973). Instead, we outline a general basis for many of the methods, and give some simple illustrations of the effect of rounding error.

### a. The arithmetic of linear equations

Several computing procedures for inverting matrices are based on solving linear equations by successive elimination and backward substitution. As an example consider the equations

$$2x + 5y + 7z = a \quad (36a)$$

$$3x + 9y + 15z = b \quad (36b)$$

$$5x + 16y + 20z = c. \quad (36c)$$

For

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 7 \\ 3 & 9 & 15 \\ 5 & 16 & 20 \end{bmatrix}, \text{ their solution will be } \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad (37)$$

provided  $\mathbf{A}^{-1}$  exists. Hence if we solve the equations by some means other than direct calculation of  $\mathbf{A}^{-1}$  and get  $x$ ,  $y$  and  $z$  in terms of  $a$ ,  $b$  and  $c$ , then the coefficients of  $a$ ,  $b$  and  $c$  will be the elements of  $\mathbf{A}^{-1}$ .

Derivation of  $\mathbf{A}^{-1}$  from (37) is achieved through solving by successive elimination. To (36b) add  $-\frac{3}{2}$  times (36a), and to (36c) add  $-\frac{5}{2}$  times (36a):

$$2x + 5y + 7z = a \quad (38a)$$

$$\left(\frac{3}{2}\right)y + \left(\frac{9}{2}\right)z = \left(-\frac{3}{2}\right)a + b \quad (38b)$$

$$\left(\frac{7}{2}\right)y + \left(\frac{5}{2}\right)z = \left(-\frac{5}{2}\right)a + c. \quad (38c)$$

Then to (38c) add  $-\frac{7}{3}$  times (38b):

$$2x + 5y + 7z = a \quad (39a)$$

$$\left(\frac{3}{2}\right)y + \left(\frac{9}{2}\right)z = \left(-\frac{3}{2}\right)a + b \quad (39b)$$

$$-8z = a + \left(-\frac{7}{3}\right)b + c. \quad (39c)$$

Now multiply (39c) by  $-\frac{1}{8}$  and after doing so add  $\left(-\frac{9}{2}\right)$  and  $(-7)$  times

the result to (39b) and (39a), respectively:

$$2x + 5y = \left(\frac{15}{8}\right)a + \left(-\frac{49}{24}\right)b + \left(\frac{7}{8}\right)c \quad (40a)$$

$$\left(\frac{3}{2}\right)y = \left(-\frac{15}{16}\right)a + \left(-\frac{5}{16}\right)b + \left(\frac{9}{16}\right)c \quad (40b)$$

$$z = \left(-\frac{1}{8}\right)a + \left(\frac{7}{24}\right)b + \left(-\frac{1}{8}\right)c. \quad (40c)$$

Finally, multiply (40b) by  $\frac{2}{3}$ , add  $-5$  times the result to (40a) and multiply that result by  $\frac{1}{2}$ . This gives

$$x = \left(2\frac{1}{2}\right)a + \left(-\frac{1}{2}\right)b + \left(-\frac{1}{2}\right)c \quad (41a)$$

$$y = \left(-\frac{5}{8}\right)a + \left(-\frac{5}{24}\right)b + \left(\frac{3}{8}\right)c \quad (41b)$$

$$z = \left(-\frac{1}{8}\right)a + \left(\frac{7}{24}\right)b + \left(-\frac{1}{8}\right)c; \quad (41c)$$

i.e.,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{5}{8} & -\frac{5}{24} & \frac{3}{8} \\ -\frac{1}{8} & \frac{7}{24} & -\frac{1}{8} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

and so from (37)

$$\mathbf{A}^{-1} = \begin{bmatrix} 2\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{5}{8} & -\frac{5}{24} & \frac{3}{8} \\ -\frac{1}{8} & \frac{7}{24} & -\frac{1}{8} \end{bmatrix} = \frac{1}{24} \begin{bmatrix} 60 & -12 & -12 \\ -15 & -5 & 9 \\ -3 & 7 & -3 \end{bmatrix}.$$

It is, of course, quite easy to do all the arithmetic without using the symbols  $x$ ,  $y$ ,  $z$  and  $a$ ,  $b$ ,  $c$ . In terms of the symbols, the meaning of a calculated value can be maintained by its position in a  $3 \times 6$  partitioned matrix which begins as

$$\left[ \begin{array}{ccc|ccc} 2 & 5 & 7 & 1 & 0 & 0 \\ 3 & 9 & 15 & 0 & 1 & 0 \\ 5 & 16 & 20 & 0 & 0 & 1 \end{array} \right]. \quad (42)$$

This comes from (36), with the first three columns in (42) being the coefficients of  $x$ ,  $y$  and  $z$  in (36) and the second three being the coefficients

of  $a$ ,  $b$  and  $c$ . Then, because the operations that yield (36)–(41) are just the adding of multiples of equations to other equations, they are equivalent to the same operations on rows of (42). As such they are called *row operations* and are similar to those described in Section 4.4 for simplifying a determinant. For example, akin to the derivation of (38), the operations row 2 plus  $-\frac{3}{2}$  times row 1, and row 3 plus  $-\frac{5}{2}$  times row 1 carried out on (42) yield

$$\left[ \begin{array}{ccc|ccc} 2 & 5 & 7 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{9}{2} & -\frac{3}{2} & 1 & 0 \\ 0 & \frac{7}{2} & \frac{5}{2} & -\frac{5}{2} & 0 & 1 \end{array} \right]$$

which is just the numeric representation of (38); and so on, down to (41), represented as

$$\left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & 2\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & -\frac{5}{8} & -\frac{5}{24} & \frac{3}{8} \\ 0 & 0 & 1 & -\frac{1}{8} & \frac{7}{24} & -\frac{1}{8} \end{array} \right]$$

This is the end, when the left-hand submatrix has become  $\mathbf{I}$ , whereupon the right-hand matrix is  $\mathbf{A}^{-1}$ . Thus having started off in (42) with  $[\mathbf{A} \mid \mathbf{I}]$ , we now have  $[\mathbf{I} \mid \mathbf{A}^{-1}]$ —providing  $\mathbf{A}^{-1}$  exists.

The arithmetic procedure just illustrated is the foundation for several computing techniques familiarly known by such names as Gauss, Doolittle, Fox and Crout. They are essentially variants of each other, directed toward improving the efficiency of organizing the calculations rather than being distinctly different methods of calculation. The reader who follows the arithmetic just described will therefore have a basic understanding of the calculations involved.

### b. Rounding error

The gain in speed and accuracy of inverting matrices using computers is obvious to anyone who has done such arithmetic without them. Generally speaking, the time required is proportional to  $n^3$  for a matrix of order  $n$ , and absolute time depends on computer hardware and software (equipment and programs), but for  $n < 100$  the time is usually a matter of seconds. In contrast, using the electric desk calculators that were stock-in-trade thirty years ago, I know of one case where it took six weeks to invert a  $40 \times 40$  matrix.

Although computers are accurate, of themselves, they do involve the difficulty of rounding error: infinite decimals cannot be represented with

complete accuracy. No matter what finite number of 6's follow the decimal point in  $4.666\dots$ , that value is not  $4\frac{2}{3}$ . Nor is this inexactitude alleviated if the last 6 is rounded to 7; e.g., 4.6667 is not exactly  $4\frac{2}{3}$  any more than is 4.6666. This rounding error may appear slight, or even inconsequential, when considering just a single number, but sometimes it accumulates to be very serious in lengthy calculations such as those of inverting matrices. This is all the more difficult because situations in which such errors will be serious cannot be specified with certainty, and a mathematical discussion of the cumulative effects of rounding errors is very involved in anything but simple situations (see Stewart, 1973). At best one can usually say that trouble does not occur very often. At worst one must be cognizant of the possibility of trouble of this sort on any occasion that rounding or truncation is involved.

Critical analysis of rounding error generation is not within the scope of this book. Instead, we give three illustrations of its effect, with a view to alerting the reader to the kinds of things that can occur with numbers. For numerical simplification, the illustrations are based on two hypothetical computers of very limited capacity.

(i) *Addition.* Suppose that a computer handles only two-digit numbers and comes in two different versions, one that rounds and one that truncates. Because computer addition is performed by adding each number to a preceding total, the sequence in which numbers are added can influence the computer total. Table 5.2 illustrates this. Three different totals are possible.

TABLE 5.2. ROUNDING ERROR IN ADDITION

True Total	Computer Addition					
	Rounding at each sum			Truncating at each sum		
	Sequence			Sequence		
	(a)	(b)	(c)	(a)	(b)	(c)
8.5	6.8	6.8	7.4	6.8	6.8	7.4
7.4	7.4	8.5	8.5	7.4	8.5	8.5
6.8	14	15	16	14	15	15
	8.5	7.4	6.8	8.5	7.4	6.8
22.7	23	22	23	22	22	21

(ii) *Inverting a matrix.* Suppose we seek to invert

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} \quad (43)$$

using a computer. Clearly  $|\mathbf{A}| = 0$  and  $\mathbf{A}^{-1}$  does not exist. But suppose  $\mathbf{A}$  comes not in the form (43) but is generated within a computer as

$$\mathbf{A} = \begin{bmatrix} 6 \times (\frac{1}{3}) & 9 \times (\frac{1}{3}) \\ 66 \times (\frac{1}{11}) & 21 \times (\frac{1}{3}) \end{bmatrix}.$$

Furthermore, suppose that the computer carries only four decimal digits, so that numbers are rounded off in the fourth decimal place, a 5 in the fifth place being rounded up. Then the matrix computed as being  $\mathbf{A}$ , which we shall call  $\mathbf{A}_c$ , is

$$\mathbf{A}_c = \begin{bmatrix} 6 \times (.3333) & 9 \times (.1111) \\ 66 \times (.0909) & 21 \times (.1429) \end{bmatrix} = \begin{bmatrix} 1.9998 & .9999 \\ 5.9994 & 3.0009 \end{bmatrix}. \quad (44)$$

Its determinant is

$$|\mathbf{A}_c| = 1.9998(3.0009) - 5.9994(.9999) = 6.0012 - 5.9988 = .0024. \quad (45)$$

We note at once that although  $|\mathbf{A}| = 0$  with  $\mathbf{A}$  having no inverse,  $|\mathbf{A}_c| \neq 0$  and  $\mathbf{A}_c$  does have an inverse:

$$\mathbf{A}_c^{-1} = \frac{1}{.0024} \begin{bmatrix} 3.0009 & -.9999 \\ -5.9994 & 1.9998 \end{bmatrix} = \begin{bmatrix} 1250.3750 & -416.6250 \\ -2499.7500 & 833.2500 \end{bmatrix}. \quad (46)$$

Clearly  $\mathbf{A}_c^{-1}$  cannot be  $\mathbf{A}^{-1}$  because  $\mathbf{A}^{-1}$  does not exist. Existence of  $\mathbf{A}_c^{-1}$  in the face of nonexistence of  $\mathbf{A}^{-1}$  occurs because  $\mathbf{A}_c$  of (44) is not the same as  $\mathbf{A}$ , and whereas the determinant of  $\mathbf{A}$  is zero, that of  $\mathbf{A}_c$  has been computed as small and nonzero, 0.0024. This is a common consequence of rounding error, a zero value being computed as nonzero, and in this instance it leads to the existence of  $\mathbf{A}_c^{-1}$ .

The situation of knowing at the outset, for some matrix  $\mathbf{A}$ , that its inverse does not exist is almost self-evident for  $\mathbf{A}$  as simple as a  $2 \times 2$ , but this is seldom the case for matrices of large order encountered in real life.

Fortunately, erroneous inverses do not occur frequently; nevertheless, even the best of programming safeguards against their occasional appearance are not universally inviolable. It is therefore important to have an appreciation for the possible consequences of rounding error, some of which have been illustrated.

The important question is when should computer output (designated  $\mathbf{A}_c^{-1}$ ) be accepted as  $\mathbf{A}^{-1}$  and when should it not? Acceptance should occur only if some check calculations yield satisfactory results. These should include calculating at least some elements of  $\mathbf{A}_c^{-1}\mathbf{A}$  and of  $\mathbf{A}\mathbf{A}_c^{-1}$  to see that they are satisfactorily close to the ones and zeros of  $\mathbf{I}$  if indeed  $\mathbf{A}_c^{-1}$  is  $\mathbf{A}^{-1}$ . For our example this is found *not* to be so:

$$\mathbf{A}_c^{-1}\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \neq \mathbf{I} \neq \mathbf{A}\mathbf{A}_c^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}.$$

Calculations like these immediately suggest that  $\mathbf{A}_c^{-1}$  is suspect as a value for  $\mathbf{A}^{-1}$ ; and, of course, the existence of  $\mathbf{A}^{-1}$  is to be questioned, too. Although calculation of  $\mathbf{A}_c^{-1}\mathbf{A}$  and  $\mathbf{A}\mathbf{A}_c^{-1}$  on a desk calculator is too onerous when  $\mathbf{A}$  is of larger order, some elements of these products should always be so checked. In addition, the consequences of using  $\mathbf{A}_c^{-1}$  should always be carefully scrutinized for reasonableness within the context of its use.

(iii) *Solving linear equations.* A computer's capacity for significant digits can materially affect the usefulness of its output, as has already been illustrated. It can also affect the input of a desired calculation to such an extent as to materially alter the whole problem. We illustrate this in terms of linear equations in just two variables.

The equations

$$\begin{bmatrix} 2.04 & 2.49 \\ 2.49 & 3.04 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} .45 \\ .55 \end{bmatrix} \quad \text{with solution} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad (47)$$

will be called the real problem and the real solution, respectively. In contrast, for a computer that handles only two-digit numbers, rounded in the right-most position, the problem has to be formulated as the computer problem

$$\begin{bmatrix} 2.0 & 2.5 \\ 2.5 & 3.0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} .45 \\ .55 \end{bmatrix} \quad \text{with real solution} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} .1 \\ .1 \end{bmatrix}. \quad (48)$$

The alteration in these real solutions, between (47) and (48), is startling: a change in sign, and a change in an order of magnitude. And these changes

arise solely as a result of having to alter the real problem so as to computerize it. Small changes in the problem have yielded considerable changes in the real solution.

Now consider the computer solution to the computer problem: taking  $2.0(3.0) - (2.5)^2 = 6.0 - 6.25$  as  $-0.3$ , it is

$$-\frac{1}{.3} \begin{bmatrix} 3.0 & -2.5 \\ -2.5 & 2.0 \end{bmatrix} \begin{bmatrix} .45 \\ .55 \end{bmatrix} = \begin{bmatrix} -10 & 8.3 \\ 8.3 & -6.7 \end{bmatrix} \begin{bmatrix} .45 \\ .55 \end{bmatrix} = \begin{bmatrix} .1 \\ 0 \end{bmatrix}. \quad (49)$$

This differs from the real solution (48) of the same problem, which in turn is different from (47), the real solution to the real problem.

The reason for this apparent fickleness of numbers is that the  $2 \times 2$  matrix in (47) is what is called *ill-conditioned*. Its determinant, relative to its elements, is quite small—0.0015. As a result, the small changes made in the matrix in deriving (48) bring about a relatively large change in the determinant, which is now 0.25; and this in turn leads to even larger differences between the inverses:

$$\begin{bmatrix} 2.04 & 2.49 \\ 2.49 & 3.04 \end{bmatrix}^{-1} = \begin{bmatrix} 2026\frac{2}{3} & -1660 \\ -1660 & 1360 \end{bmatrix}$$

whereas 
$$\begin{bmatrix} 2.0 & 2.5 \\ 2.5 & 3.0 \end{bmatrix}^{-1} = \begin{bmatrix} -12 & 10 \\ 10 & -8 \end{bmatrix}.$$

Further demonstration of the effect of the relatively large numbers in the first of these inverses is to change the real problem of (47) to be

$$\begin{bmatrix} 2.04 & 2.49 \\ 2.49 & 3.04 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} .451 \\ .55 \end{bmatrix} \quad \text{with real solution} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.026 \\ -.66 \end{bmatrix}.$$

Simply adding .001 to one element of the right-hand side of the equations brings the substantive change of altering the sign of both elements of the solution.

These simple examples illustrate how computers can produce unsatisfactory output as a consequence of rounding error. Their simplicity might suggest that this occurs frequently, but it does not for several reasons. First, real computers handle many more than the three or four digits of those of the examples. Second, even rounding error accumulated through lengthy arithmetic will seldom affect more than two of the right-most digits. Third, good computing methods, such as the use of double-precision floating-point numbers, can minimize the effects of rounding error. Finally, real-life data

are not often as pathological as those of the examples that were designed to illustrate what *can* sometimes happen rather than what usually *does* happen. Consequently, computer routines designed for inverting matrices deliver results that are seldom spoiled by rounding error. But this can happen, and computer users need to be aware of such an eventuality, its possible causes and the consequences.

## 10. LEFT AND RIGHT INVERSES

Section 2 of this chapter begins with illustrating matrices **A** and **L** for which either

- (i)  $\mathbf{LA} = \mathbf{I} = \mathbf{AL}$ ,
- (ii)  $\mathbf{LA} = \mathbf{I}$ , but  $\mathbf{AL} \neq \mathbf{I}$ , with more than one such **L**,
- or (iii) There is no **L** for which  $\mathbf{LA} = \mathbf{I}$ , but there are matrices **R** for which  $\mathbf{AR} = \mathbf{I}$ .

There is also a fourth case:

- (iv) There are matrices **A** for which there is neither **L** satisfying  $\mathbf{LA} = \mathbf{I}$  nor **R** satisfying  $\mathbf{AR} = \mathbf{I}$ .

Case (i) is when  $\mathbf{L} = \mathbf{A}^{-1}$ , which has been fully discussed. In case (ii) the matrices **L** are known as *left inverses* of **A**, and in (iii) the matrices **R** are called *right inverses* of **A**. A simple theorem delineates when left (and conversely right) inverses may exist.

**Theorem.**  $\mathbf{A}_{r \times c}$  may have left inverses only when  $r \geq c$ .

**Proof.** For  $r \geq c$  partition **A** as  $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$  where **X** is square; and let  $\mathbf{L} = [\mathbf{M} \ \mathbf{N}]$  be conformable for the product  $\mathbf{LA}$ . Then  $\mathbf{LA} = \mathbf{MX} + \mathbf{NY}$ , so that whenever **M** and **N** are such that  $\mathbf{MX} + \mathbf{NY} = \mathbf{I}$ , the matrix **L** is a left inverse of **A**. An example, for nonsingular **X** is  $\mathbf{M} = \mathbf{X}^{-1}$  and  $\mathbf{N} = \mathbf{0}$ .

For  $r < c$  partition **A** as  $[\mathbf{X} \ \mathbf{Y}]$  and **L** as  $\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}$ , and assume **L** to be a left inverse of **A**. Then  $\mathbf{LA} = \mathbf{I}$ , i.e.,

$$\begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix} [\mathbf{X} \ \mathbf{Y}] = \mathbf{I}, \quad \text{or} \quad \begin{bmatrix} \mathbf{MX} & \mathbf{MY} \\ \mathbf{NX} & \mathbf{NY} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Therefore

$$\mathbf{MX} = \mathbf{I}, \quad \mathbf{NX} = \mathbf{0} \quad \text{and} \quad \mathbf{NY} = \mathbf{I}. \quad (50)$$

Because **M** and **X** are square, the first equation in (50) implies the existence of  $\mathbf{X}^{-1}$ . Hence from the second equation  $\mathbf{N} = \mathbf{0}$ , which contradicts the third equation in (50); i.e., the assumption that **L** is a left inverse is contradicted. QED

**Converse.**  $\mathbf{A}_{r \times c}$  may have right inverses only if  $r \leq c$ .

**Corollary.**  $A$  may have both a left and a right inverse only when  $r = c$ .

Note that the conditions of the theorem, its converse and corollary are only those under which left and right inverses may exist. Conditions under which they *do* exist are specified in Lemma 8 of Section 7.5.

**Summary.** Square matrices may have both a left and right inverse. Rectangular matrices may have left (right) inverses when they have more (fewer) rows than columns.

## EXERCISES

1. Show that if

$$(a) \mathbf{A} = \begin{bmatrix} 6 & 13 \\ 5 & 12 \end{bmatrix}, \text{ then } |\mathbf{A}| = 7, \quad \mathbf{A}^{-1} = \frac{1}{7} \begin{bmatrix} 12 & -13 \\ -5 & 6 \end{bmatrix}$$

$$\text{and } \mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I};$$

$$(b) \mathbf{B} = \begin{bmatrix} 3 & -4 \\ 7 & 14 \end{bmatrix}, \text{ then } |\mathbf{B}| = 70, \quad \mathbf{B}^{-1} = \frac{1}{70} \begin{bmatrix} 14 & 4 \\ -7 & 3 \end{bmatrix}$$

$$\text{and } \mathbf{B}\mathbf{B}^{-1} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}.$$

2. Demonstrate the reversal rule for the inverse of a product of two matrices, using  $\mathbf{A}$  and  $\mathbf{B}$  given in Exercise 1.

3. Show that if

$$(a) \mathbf{A} = \begin{bmatrix} -1 & 3 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 4 \end{bmatrix}, |\mathbf{A}| = -5, \quad \mathbf{A}^{-1} = \frac{1}{5} \begin{bmatrix} -8 & 12 & -3 \\ -1 & 4 & -1 \\ 2 & -3 & 2 \end{bmatrix}$$

$$\text{and } \mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I};$$

(b)

$$\mathbf{B} = \begin{bmatrix} 10 & 6 & -1 \\ 6 & 5 & 4 \\ -1 & 4 & 17 \end{bmatrix}, |\mathbf{B}| = 25, \quad \mathbf{B}^{-1} = \begin{bmatrix} 2.76 & -4.24 & 1.16 \\ -4.24 & 6.76 & -1.84 \\ 1.16 & -1.84 & 0.56 \end{bmatrix}$$

$$\text{and } \mathbf{B}^{-1}\mathbf{B} = \mathbf{B}\mathbf{B}^{-1} = \mathbf{I};$$

$$(c) \mathbf{C} = \frac{1}{10} \begin{bmatrix} 0 & -6 & 8 \\ -10 & 0 & 0 \\ 0 & -8 & -6 \end{bmatrix}, \text{ then } |\mathbf{C}| = 1, \quad \mathbf{C}^{-1} = \mathbf{C}'$$

$$\text{and } \mathbf{C}\mathbf{C}' = \mathbf{C}'\mathbf{C} = \mathbf{I}.$$

4. For each of the following matrices derive the determinant and inverse. Check each inverse by multiplication.

$$(a) \begin{bmatrix} 7 & 3 \\ 8 & 9 \end{bmatrix}; \quad (b) \begin{bmatrix} 6 & 31 \\ 8 & 29 \end{bmatrix}; \quad (c) \begin{bmatrix} -7 & -4 \\ 3 & 1 \end{bmatrix};$$

$$(d) \begin{bmatrix} 1 & 5 & -5 \\ 3 & 2 & -5 \\ 6 & -2 & -5 \end{bmatrix}; \quad (e) \begin{bmatrix} -3 & 2 & -6 \\ -3 & 5 & -7 \\ -2 & 3 & -4 \end{bmatrix}; \quad (f) \begin{bmatrix} 2 & 1 & 3 \\ -5 & 1 & 0 \\ 1 & 4 & -2 \end{bmatrix}$$

$$(g) \begin{bmatrix} -1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & -1 \end{bmatrix}; \quad (h) \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix};$$

$$(i) \begin{bmatrix} 7 & 4 & -1 \\ 4 & 7 & -1 \\ -4 & -4 & 4 \end{bmatrix}$$

5. Find the inverse of  $\begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 7 & 0 \\ 0 & 4 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{bmatrix}$ .

6. Calculate the inverse of  $\begin{bmatrix} 1 & 0 & 6 & 8 \\ 0 & 1 & 5 & 4 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$  and of  $\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$ .

7. Let  $\mathbf{A} = \begin{bmatrix} 6 & -1 & 4 \\ 2 & 5 & -3 \\ 1 & 1 & 2 \end{bmatrix}$ .

(a) Calculate the transpose of  $\mathbf{A}^{-1}$  and the inverse of  $\mathbf{A}'$ .

(b) Calculate the inverse of  $\mathbf{A}^{-1}$ .

8. Write each of the following sets of equations in matrix and vector form. Solve the equations, after obtaining the necessary inverse matrices (if they exist). If an inverse does not exist, explain why.

$$(a) \begin{cases} 3x + 4y - 2z = 4, \\ -x - y + 3z = 6, \\ x - 7y + z = -2. \end{cases} \quad (b) \begin{cases} -3x + 4y = 2, \\ x + 7z = -8, \\ -2x - 3y - 8z = -11. \end{cases}$$

$$(c) \begin{cases} 2w + 8y - z = 0, \\ 8x - 3z = -4, \\ 4y + \frac{3}{2}z = 42, \\ \frac{3}{8}z = 7\frac{1}{2}. \end{cases} \quad (d) \begin{cases} x + 3y - z = 7, \\ 4x - y + 2z = 8\frac{1}{2}, \\ 5x + 2y + z = 10. \end{cases}$$

9. Calculate the inverse of the following matrices. Check each inverse by calculating a product that is an identity matrix.

$$(a) \begin{bmatrix} 1 & -4 & 1 \\ 2 & 7 & 0 \\ 0 & 1 & 0 \end{bmatrix}; \quad (b) \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & 0 \\ 4 & 0 & 1 \end{bmatrix}; \quad (c) \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

10. In Exercise 9 denote each matrix in turn by  $\mathbf{A}$  and solve the equations  $\mathbf{Ax} = \mathbf{1}$ .
11. Prove that the only nonsingular idempotent matrices are identity matrices.
12. Verify

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_1 & b_1 & c_1 & d_1 \\ a_1 & b_1 & c_1 & d_1 \\ a_1 & b_1 & c_1 & d_1 \end{bmatrix}^{-1} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{bmatrix} = a_1.$$

13. When  $\mathbf{A}$  is skew-symmetric, prove the following.
- (a) When  $\mathbf{A}^{-1}$  exists, it is skew-symmetric.
- (b)  $(\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A})^{-1}$  is orthogonal.  
[Hint:  $\mathbf{I} - \mathbf{A}$  and  $\mathbf{I} + \mathbf{A}$  commute in multiplication.]
14. For  $\mathbf{A} = \{a_{ij}\}$ , explain why
- (a)  $\mathbf{B} = \mathbf{A} - \mathbf{A}(\mathbf{I}'\mathbf{A}\mathbf{I})^{-1}\mathbf{I}'\mathbf{A} = \{a_{ij} - a_i a_{.j}/a_{.i}\}$ ;
- (b)  $\mathbf{B}$  has row sums that are zero.
15. Prove the following statements.
- (a) For  $\mathbf{A}$  and  $\mathbf{B}$  symmetric,  $[(\mathbf{AB})']^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$ .
- (b)  $\mathbf{C} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  is symmetric and idempotent.
- (c) The transpose and inverse of an orthogonal matrix are equal.
- (d) All powers of a symmetric orthogonal matrix are the matrix itself, or an identity matrix.
- (e) When  $\mathbf{A} = \lambda\mathbf{B}$ , of order  $n$  and for scalar  $\lambda$ , then  $|\mathbf{A}| = \lambda^n |\mathbf{B}|$  and  $\mathbf{A}^{-1} = (1/\lambda)\mathbf{B}^{-1}$ . When is the latter not true?
- (f) Prove that there is only one form of matrix that is orthogonal and idempotent, and find that form.
- (g) If  $\mathbf{A}$  is a square matrix of order  $n$ ,  $|\text{adj}(\mathbf{A})| = |\mathbf{A}|^{n-1}$ .

- (h) The product of two matrices being a null matrix does not necessarily imply that one of them is null.
- (i) When  $\mathbf{A}$  is positive definite, so is  $\mathbf{A}^{-1}$ .
16. Assuming that each of the stated inverses exist, prove the following statements.
- (a)  $(\mathbf{I} + \mathbf{A}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{I})^{-1}$ .
- (b)  $(\mathbf{A} + \mathbf{BB}')^{-1}\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}$ .
- (c)  $(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}$ .
- (d)  $\mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} = \mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$ .
- (e)  $\mathbf{A}^{-1} + \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{A} + \mathbf{B})\mathbf{B}^{-1}$ .
- (f)  $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1}$  implies  $\mathbf{AB}^{-1}\mathbf{A} = \mathbf{BA}^{-1}\mathbf{B}$ .
- (g)  $(\mathbf{I} + \mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}\mathbf{B}$ .
- (h)  $(\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}$ .
17. For  $\mathbf{H} = \mathbf{I} - 2\mathbf{w}\mathbf{w}'$  with  $\mathbf{w}'\mathbf{w} = 1$ , prove that  $\mathbf{H} = \mathbf{H}^{-1}$ .
18. Suppose  $\mathbf{KA} = \mathbf{0}$  and  $\mathbf{K}$  is idempotent. Define  $\mathbf{G} = (\mathbf{A} - \mathbf{K})^{-1}$ . Prove that (a)  $\mathbf{AG} = \mathbf{I} - \mathbf{K}$ , (b)  $\mathbf{AGA} = \mathbf{A}$ , and (c)  $\mathbf{AGK} = \mathbf{0}$ .
19. In Table 5.1 of Section 8a take the value of agricultural output to be \$1000/ton, that of a machine to be \$2000 and that of an employee to be \$4000. Rework Table 5.1 in terms of dollar values, check it using (27) and find the output required to satisfy government demands of \$600,000 worth of agricultural produce, \$2,000,000 worth of machines and \$2,400,000 for labor.
20. Suppose the three production sectors charge each other and the government variable prices as in the following table.

Production Sector	Price Charged			
	Agriculture	Manuf.	Labor	Govt.
Agriculture (per ton)	\$ 700	\$ 900	\$ 800	\$1,000
Manufacturing (per machine)	2,000	1,537½	1,800	2,010
Labor (per employee)	3,840	3,700	2,530	3,355

- (a) Using these prices, rework Table 5.1 of Section 8a and check it.
- (b) How much agricultural and manufacturing output is needed (tons of produce and numbers of machines) and how many



employees are needed to satisfy a government demand of \$800,000, \$3,200,000 and \$2,400,000 worth of agricultural produce, machines and employees, respectively?

Hint: Make use of

$$\begin{bmatrix} .8320 & -.048 & -.04480 \\ -1.2000 & .836 & -.05040 \\ -1.3824 & -2.368 & .92916 \end{bmatrix}^{-1} = \begin{bmatrix} 2.1137 & .4855 & .1282 \\ 3.8087 & 2.2863 & .3076 \\ 12.8514 & 6.5475 & 2.0510 \end{bmatrix}.$$

21. Suppose an input-output table for  $n$  production sectors has the form of Table 5.1 in Section 8a. Define the  $n \times n$  matrix for the production sectors as  $\mathbf{M}$ , the final demand vector as  $\mathbf{y}$  and the vector of production totals as  $\mathbf{x}$ . Then the row sums of the partitioned matrix  $[\mathbf{M} \ \mathbf{y}]$  are the elements of  $\mathbf{x}$ .

- (a) State this relationship between  $\mathbf{M}$ ,  $\mathbf{y}$  and  $\mathbf{x}$  as an equation.  
 (b) Define a diagonal matrix  $\mathbf{D}$  such that  $\mathbf{MD}^{-1}$  is the input-output coefficient matrix in (27).  
 (c) Show that (27) always holds for such a table.

22. For  $p$  means  $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$  for  $i = 1, 2, \dots, p$ , define  $p$  linear combinations

$$c_k = \sum_{i=1}^p \lambda_{ki} \bar{x}_i.$$

with

$$\sum_{i=1}^p \frac{\lambda_{ki}^2}{n_i} = \alpha_k \quad \text{for any } \alpha_k, \quad \text{and} \quad \sum_{i=1}^p \frac{\lambda_{ki} \lambda_{mi}}{n_i} = 0$$

for  $k \neq m = 1, 2, \dots, p$ . Develop appropriate matrices to provide a proof of the equality

$$\sum_{k=1}^p \frac{c_k^2}{\alpha_k} = \sum_{i=1}^p n_i \bar{x}_i^2.$$

23. For  $\mathbf{I}$  and  $\mathbf{J}$  of order  $n$ , prove the following.

- (a) For  $p \neq 0$  and  $p + nq \neq 0$ ,

$$(p\mathbf{I} + q\mathbf{J})^{-1} = \frac{1}{p} \left( \mathbf{I} - \frac{q}{p + nq} \mathbf{J} \right). \quad (\text{See Exercise 3.11c.})$$

- (b) For  $\mathbf{C}$  such that  $\mathbf{JC} = \mathbf{0} = \mathbf{CJ}$ ,

$$\mathbf{J}(\mathbf{C} + a\mathbf{J})^{-1} = (\mathbf{C} + a\mathbf{J})^{-1}\mathbf{J} = \frac{1}{an}\mathbf{J}$$

and hence

$$(\mathbf{C} + a\mathbf{J})^{-1} - (\mathbf{C} + b\mathbf{J})^{-1} = \frac{b-a}{abn^2}\mathbf{J}.$$

24. For orthogonal  $\mathbf{T}$ , and  $\mathbf{L}$  being a left inverse of  $\mathbf{A}$ , what is a left inverse of  $\mathbf{TA}$  and of  $\mathbf{AT}$ ?
25. When  $\mathbf{G}$  represents either a left or a right inverse (whichever exists) of  $\mathbf{A}$ , show that  $\mathbf{AGA} = \mathbf{A}$  and  $\mathbf{GAG} = \mathbf{G}$ , and that  $\mathbf{GA}$  is symmetric in one case and  $\mathbf{AG}$  is in the other.
26. (a) For square  $\mathbf{L}$  and  $\mathbf{LL}'$  nonsingular, prove that  $\mathbf{L}$  is nonsingular.  
 (b) For symmetric  $\mathbf{V}$  prove that  $\mathbf{K} = \mathbf{X}'(\mathbf{XV}^{-1}\mathbf{X}')^{-1}\mathbf{X}$  is symmetric and that the transpose of  $\mathbf{D} = \mathbf{KV}^{-1}$  is  $\mathbf{V}^{-1}\mathbf{DV}$ . Assume  $\mathbf{X}$  is rectangular.
27. (a) Prove  $(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBV}(\mathbf{I} + \mathbf{A}^{-1}\mathbf{UBV})^{-1}\mathbf{A}^{-1}$ . [Henderson and Searle (1981a) deal with this inverse in detail.]  
 (b) Derive the inverse of  $\begin{bmatrix} \mathbf{M} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}$ .
28. For  $\mathbf{w}' = [\mathbf{y}' \ \mathbf{0}]$  where  $\mathbf{y}$  is a vector of order  $N$ , and with  $\mathbf{Z}' = [\mathbf{0}_{k \times N} \ \mathbf{I}_{k \times k}]$ , prove that

$$\begin{aligned} \mathbf{w}' \left( \mathbf{Z} - \frac{1}{N+k} \mathbf{JZ} \right) \left( \mathbf{Z}'\mathbf{Z} - \frac{1}{N+k} \mathbf{Z}'\mathbf{JZ} \right)^{-1} \left( \mathbf{Z} - \frac{1}{N+k} \mathbf{JZ} \right)' \mathbf{w} \\ = \frac{Nk\bar{y}^2}{N+k} \quad \text{for } \bar{y} \text{ being the average of the elements of } \mathbf{y}. \end{aligned}$$

29. Suppose  $\mathbf{J}_{ij}$  is a  $\mathbf{J}$ -matrix of order  $n_i \times n_j$ . Define, for  $i, j = 1, \dots, c$

$$\mathbf{A} = \{a_{ij}\}, \quad \mathbf{D} = \text{diag}\{n_i\} \quad \text{and} \quad \mathbf{B} = \{b_{ij}\} = (\mathbf{AD} + \mathbf{I})^{-1}\mathbf{A}.$$

Prove that

$$\mathbf{T}^{-1} = [\mathbf{I} + \{a_{ij}\mathbf{J}_{ij}\}]^{-1} = \mathbf{I} - \{b_{ij}\mathbf{J}_{ij}\}.$$

30. In Section 3.2 an expression is given for the vector of total expected rewards after  $n$  transitions in a Markov chain:

$$v_n = (\mathbf{P} \cdot \mathbf{R})\mathbf{1} + \mathbf{P}v_{n-1}.$$

- (a) Show that for  $v_0 = \mathbf{0}$ ,

$$v_n = (\mathbf{I} + \mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^{n-1})(\mathbf{P} \cdot \mathbf{R})\mathbf{1}.$$

- (b) Explain why (34) and (35) cannot be used to simplify  $v_n$ .

## CHAPTER 6

### RANK

Equations  $\mathbf{Ax} = \mathbf{b}$  can be solved as  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$  only if the inverse,  $\mathbf{A}^{-1}$ , exists. And  $\mathbf{A}^{-1}$  does exist only if  $|\mathbf{A}|$ , the determinant of  $\mathbf{A}$ , is nonzero. Therefore, whenever wanting to solve equations  $\mathbf{Ax} = \mathbf{b}$  as  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ , we apparently must first ascertain whether  $|\mathbf{A}|$  is zero or not. This can be as lengthy a task as solving the equations, and although  $|\mathbf{A}|$  plays an important role in  $\mathbf{A}^{-1}$  its explicit value is not really required for calculating either  $\mathbf{A}^{-1}$  (see, for example, Section 5.9a), or  $\mathbf{A}^{-1}\mathbf{b}$ . But before attempting either of these calculations, we should know that  $|\mathbf{A}|$  is nonzero or, equivalently, that  $\mathbf{A}^{-1}$  does exist. This chapter introduces concepts that, among other things, permit us to ascertain whether or not  $|\mathbf{A}|$  is zero, without having to tediously expand  $|\mathbf{A}|$  in full, as described in Chapter 4. The concepts involved also have far wider ramifications throughout matrix theory than just the calculation of determinants. They apply to matrices of all kinds and are the foundation of many important results in matrix algebra.

The starting point is the idea of a linear combination of vectors.

#### 1. LINEAR COMBINATIONS OF VECTORS

When  $x$  and  $y$  represent the axes in two-dimensional Cartesian coordinates, the equation of a straight line can be represented as  $ax + by = c$  for constants  $a$ ,  $b$  and  $c$ . The idea of linearity extends to vectors by calling  $a\mathbf{x} + b\mathbf{y}$  a *linear combination* of the vectors  $\mathbf{x}$  and  $\mathbf{y}$ . More generally, if we have  $n$  vectors all of the same order, which we call a set of  $n$  vectors,  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , then for scalars  $a_1, a_2, \dots, a_n$ ,

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n \quad (1)$$

is called a *linear combination of the set of  $n$  vectors*. For example, suppose we have the set of three vectors:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 6 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 11 \\ 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 7 \end{bmatrix}. \quad (2)$$

A linear combination of these is  $2\mathbf{x}_1 + 3\mathbf{x}_2 - 4\mathbf{x}_3$ . Its value is

$$\begin{aligned} 2\mathbf{x}_1 + 3\mathbf{x}_2 - 4\mathbf{x}_3 &= 2 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 6 \end{bmatrix} + 3 \begin{bmatrix} -2 \\ 11 \\ 3 \\ 5 \end{bmatrix} - 4 \begin{bmatrix} 0 \\ 1 \\ 2 \\ 7 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 0 \\ -2 \\ 12 \end{bmatrix} + \begin{bmatrix} -6 \\ 33 \\ 9 \\ 15 \end{bmatrix} + \begin{bmatrix} 0 \\ -4 \\ -8 \\ -28 \end{bmatrix} = \begin{bmatrix} -4 \\ 29 \\ -1 \\ -1 \end{bmatrix}. \end{aligned} \quad (3)$$

A linear combination of vectors is always a vector. For example, using (2)

$$\begin{aligned} a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 &= a_1 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 6 \end{bmatrix} + a_2 \begin{bmatrix} -2 \\ 11 \\ 3 \\ 5 \end{bmatrix} + a_3 \begin{bmatrix} 0 \\ 1 \\ 2 \\ 7 \end{bmatrix} \\ &= \begin{bmatrix} a_1 \\ 0 \\ -a_1 \\ 6a_1 \end{bmatrix} + \begin{bmatrix} -2a_2 \\ 11a_2 \\ 3a_2 \\ 5a_2 \end{bmatrix} + \begin{bmatrix} 0 \\ a_3 \\ 2a_3 \\ 7a_3 \end{bmatrix} = \begin{bmatrix} a_1 - 2a_2 + 0a_3 \\ 0a_1 + 11a_2 + a_3 \\ -a_1 + 3a_2 + 2a_3 \\ 6a_1 + 5a_2 + 7a_3 \end{bmatrix} \end{aligned} \quad (4)$$

is a vector for any scalar values  $a_1$ ,  $a_2$  and  $a_3$ . The  $a$ 's must be scalar, either positive or negative (real) numbers or even mathematical functions, so long as they are scalars. In particular, some or all of the  $a$ 's can be negative, as in (3).

Notice that the final result in (4) can also be written as

$$\begin{bmatrix} a_1 - 2a_2 + 0a_3 \\ 0a_1 + 11a_2 + a_3 \\ -a_1 + 3a_2 + 2a_3 \\ 6a_1 + 5a_2 + 7a_3 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 11 & 1 \\ -1 & 3 & 2 \\ 6 & 5 & 7 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \mathbf{X}\mathbf{a}.$$

on defining  $\mathbf{X}$  as the matrix having columns  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$ , and  $\mathbf{a}$  as the vector of  $a$ 's. This is true in general: define

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \quad \text{and} \quad \mathbf{a}' = [a_1 \quad a_2 \quad \cdots \quad a_n]. \quad (5)$$

Then the linear combination (1) can be written in the alternative forms as

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \sum_{i=1}^n a_i\mathbf{x}_i = \mathbf{X}\mathbf{a}. \quad (6)$$

Conversely, given a matrix  $\mathbf{X}$  and a vector  $\mathbf{a}$  for which the vector  $\mathbf{X}\mathbf{a}$  exists, we see from (5) and (6) that

$$\mathbf{X}\mathbf{a} \text{ is a column vector, a linear combination of the columns of } \mathbf{X}. \quad (7)$$

Similarly,

$$\mathbf{b}'\mathbf{X} \text{ is a row vector: it is a linear combination of the rows of } \mathbf{X},$$

and

$\mathbf{A}\mathbf{B}$  is a matrix:

its rows are linear combinations of the rows of  $\mathbf{B}$ , and

its columns are linear combinations of the columns of  $\mathbf{A}$ . (8)

These descriptions are a direct outcome of the definitions of a linear combination of vectors and of matrix multiplication. But they have many important uses in the development of matrix algebra, especially (7).

## 2. LINEAR TRANSFORMATIONS

The vector  $\mathbf{X}\mathbf{a}$  is sometimes called the *linear transformation* of the vector  $\mathbf{a}$  to the vector  $\mathbf{X}\mathbf{a}$ , with  $\mathbf{X}$  being the matrix of the transformation. This idea is particularly pertinent when  $a_1$ ,  $a_2$  and  $a_3$  say, are three variables that are transformed, with some matrix  $\mathbf{X}$ , into other variables  $b_1$ ,  $b_2$  and  $b_3$  by the linear transformation

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b} = \mathbf{X}\mathbf{a} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

Each new variable  $b_i$  is then a linear combination of the three  $a_i$ 's; e.g.,  $b_2 = x_{21}a_1 + x_{22}a_2 + x_{23}a_3$ .

**Illustration.** In genetics, the three possible genotypes concerning a single locus on a chromosome at which there are only two alleles, G and g, are GG, Gg and gg. Denoting these by  $g_1$ ,  $g_2$  and  $g_3$ , respectively, gene effects relative to these genotypes can be defined (Anderson and Kempthorne, 1954, for example) in terms of a mean  $\mu$ , a measure of gene substitution  $\alpha$ , and a measure of dominance  $\delta$ , such that

$$\begin{aligned}\mu &= \frac{1}{4}g_1 + \frac{1}{2}g_2 + \frac{1}{4}g_3 \\ \alpha &= \frac{1}{4}g_1 \quad -\frac{1}{4}g_3 \\ \delta &= -\frac{1}{4}g_1 + \frac{1}{2}g_2 - \frac{1}{4}g_3,\end{aligned}$$

i.e.,

$$\begin{bmatrix} \mu \\ \alpha \\ \delta \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \quad (9)$$

This represents a linear transformation of the vector  $[g_1 \ g_2 \ g_3]'$  to  $[\mu \ \alpha \ \delta]'$ .

An alternative definition of gene effects  $m$ ,  $a$  and  $d$ , given by Hayman and Mather (1955), is

$$\begin{aligned}g_1 &= m + a \\ g_2 &= m + d, \quad \text{i.e.,} \\ g_3 &= m - a\end{aligned} \quad \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} \quad (10)$$

The combined effect of these two transformations yields the relationship between the two definitions. Substituting (10) into (9) gives

$$\begin{aligned}\begin{bmatrix} \mu \\ \alpha \\ \delta \end{bmatrix} &= \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} m + \frac{1}{2}d \\ \frac{1}{2}a \\ \frac{1}{2}d \end{bmatrix} \quad (11)\end{aligned}$$

In this way (11) represents the linear transformation of  $m$ ,  $a$  and  $d$  to  $\mu$ ,  $\alpha$  and  $\delta$ :  $\mu = m + \frac{1}{2}d$ ,  $\alpha = \frac{1}{2}a$  and  $\delta = \frac{1}{2}d$ .

Notice that whereas  $y = Ax$  represents the linear transformation of  $x$  to  $y$ , this also implies the linear transformation of  $y$  to  $x$  in the form  $x = A^{-1}y$ ,

provided  $A^{-1}$  exists. Thus from (11)

$$\begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}^{-1} \begin{bmatrix} \mu \\ \alpha \\ \delta \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \delta \end{bmatrix} \quad (12)$$

Also, derivation of both (11) and (12) involves only matrix arithmetic, multiplication for (11) and inversion for (12). There is no need for manipulation of equations, which can be tedious when many variables are involved, especially compared to using matrix arithmetic, which, when computerized, is as equally feasible for matrices of large order as it is for those of order 3 illustrated here.

**Illustration.** An application of the foregoing occurs in the illustration at the end of Section 2.7b concerning inbreeding. There we had two vectors  $f^{(0)}$  and  $f^{(1)}$  related by the equation  $f^{(1)} = Af^{(0)}$ . By the same arguments the equation  $f^{(2)} = Af^{(1)}$  also holds true, and therefore  $f^{(2)} = AAf^{(0)} = A^2f^{(0)}$ ; and in general  $f^{(n)} = A^n f^{(0)}$ ,  $f^{(n)}$  being the vector of frequencies of different types of mating after  $n$  generations of inbreeding.

Two practical problems arise here: that of computing  $A^n$ , the  $n$ th power of  $A$ , and that of finding the limiting value of  $A^n$ , if any, for very large values of  $n$ . We return to these problems in Section 11.5b.

### 3. LINEAR DEPENDENCE AND INDEPENDENCE

#### a. Definitions

The product  $Xa$  is a vector, and it is a linear combination of the column vectors in  $X$  as in (6):

$$Xa = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

**Definition: linearly dependent vectors.** If there exists a vector  $a \neq 0$ , such that  $a_1x_1 + \cdots + a_nx_n = 0$ , then provided none of  $x_1, x_2, \dots, x_n$  is null, those vectors are said to be *linearly dependent vectors*.

An alternative statement of the definition is: if  $Xa = 0$  for some non-null  $a$ , then the columns of  $X$  are linearly dependent vectors, provided none is null.

**Example.** Consider the vectors

$$x_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad x_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} \quad \text{and} \quad x_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}.$$

It is clear that

$$2\mathbf{x}_1 + \mathbf{x}_4 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}; \quad (13)$$

i.e.,

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_4 = \mathbf{0} \quad \text{for} \quad \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

which is non-null. Therefore  $\mathbf{x}_1$  and  $\mathbf{x}_4$  are linearly dependent. So also are  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$  because

$$2\mathbf{x}_1 + 3\mathbf{x}_2 - 3\mathbf{x}_3 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} 0 \\ 15 \\ -15 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}. \quad (14)$$

In contrast to (13) and (14), consider

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = \begin{bmatrix} 3a_1 \\ -6a_1 \\ 9a_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 5a_2 \\ -5a_2 \end{bmatrix} = \begin{bmatrix} 3a_1 \\ -6a_1 + 5a_2 \\ 9a_1 - 5a_2 \end{bmatrix}. \quad (15)$$

There are no values  $a_1$  and  $a_2$  which make (15) a null vector other than  $a_1 = 0 = a_2$ . Therefore  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are not linearly dependent. They are said to be linearly independent.

**Definition: linearly independent vectors.** If  $\mathbf{a} = \mathbf{0}$  is the only vector for which  $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$ , then provided none of  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  is null, those vectors are said to be *linearly independent vectors*.

An alternative statement of this is: if  $\mathbf{Xa} = \mathbf{0}$  only for  $\mathbf{a} = \mathbf{0}$ , then the columns of  $\mathbf{X}$  are linearly independent vectors.

**Example (continued).** Since (15) is null only for  $a_1 = 0 = a_2$ , i.e.,

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = \mathbf{0} \quad \text{only for} \quad \mathbf{a} = \mathbf{0}, \quad (16)$$

the vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are linearly independent.

A difficulty with assimilating these definitions is that they are not as succinct as is the definition, for example, of  $\mathbf{A}$  being idempotent:  $\mathbf{A}^2 = \mathbf{A}$ . That definition is summarized by the simple statement  $\mathbf{A}^2 = \mathbf{A}$ . There is no such simplicity for the definitions of linear dependence and independence, because they involve not just an equation but statements about conditions for an equation being true:  $\mathbf{Xa} = \mathbf{0}$  being true for some  $\mathbf{a} \neq \mathbf{0}$  means the columns of  $\mathbf{X}$  are linearly dependent, whereas it being true only for  $\mathbf{a} = \mathbf{0}$  means they are linearly independent.

### b. General characteristics

The linear dependence or linear independence of vectors is a characteristic pertaining to a set of vectors of the same order. It is not a characteristic of individual vectors.

(i) *Sets of vectors.* The characteristic of vectors being linearly dependent or independent belongs to the whole set of vectors being considered in the equation  $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$  and not just to subsets of them. Thus although (14) shows that  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$  are linearly dependent, there is no confusion of this with the consequence of (16), that  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are linearly independent—even though  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$  include  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . The characteristic of linear dependence applies to the set of vectors  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$  and that of linear independence applies to the set  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .

(ii) *Some a's zero.* The definitions of linear dependence and of linear independence turn on whether  $\mathbf{Xa} = \mathbf{0}$  for some  $\mathbf{a} \neq \mathbf{0}$  or only for  $\mathbf{a} = \mathbf{0}$ . As a result, the definitions rely on whether none (or only some) of the  $a_i$ 's in  $\mathbf{a}$  are zero, or on whether all of them are. Thus, because in (13)  $2\mathbf{x}_1 + \mathbf{x}_4 = \mathbf{0}$ , we say that  $\mathbf{x}_1$  and  $\mathbf{x}_4$  are linearly dependent; and because  $2\mathbf{x}_1 + 0\mathbf{x}_2 + \mathbf{x}_4 = \mathbf{0}$ , so also are  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_4$ ; in this case  $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_4 = \mathbf{0}$  with some (but not all) the  $a$ 's being zero. Hence the conditions are satisfied for  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_4$  being linearly dependent. If the sole value of  $\mathbf{a}$  for which  $\mathbf{Xa} = \mathbf{0}$  is true is  $\mathbf{a} = \mathbf{0}$  (i.e., all  $a_i$ 's zero), only then do we say that the columns of  $\mathbf{X}$  are linearly independent.

(iii) *Existence and non-uniqueness of nonzero a's.* Discussion of examples relies heavily on actual values of  $a$ 's. In practice, it is not individual values of  $a$ 's that are important but whether or not  $\mathbf{Xa} = \mathbf{0}$  is true only for all the  $a_i$ 's being zero. If it is, then the vectors are linearly independent. If not, if nonzero  $a$ 's exist such that  $\mathbf{Xa} = \mathbf{0}$  is true, then the vectors are linearly dependent and the important thing is not the individual values of those  $a$ 's but the fact that they do exist; i.e., it is the *existence* of nonzero  $a$ 's that is important. It is their existence and not their actual value that is utilized.

One reason for not being concerned with individual values of nonzero  $a$ 's when they exist is that they are not unique; for example, with (13) being true, so also is  $4\mathbf{x}_1 + 2\mathbf{x}_4 = \mathbf{0}$ . In general, if  $\mathbf{Xa} = \mathbf{0}$  is true for some  $\mathbf{a}$ , then so is  $\mathbf{X}\lambda\mathbf{a} = \mathbf{0}$  for scalar  $\lambda$ .

(iv) *Null vectors.* Suppose  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are any vectors of order 3. Then

$$a_1\mathbf{y}_1 + a_2\mathbf{y}_2 + a_3 \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}$$

is true for  $a_1 = 0 = a_2$  and for any nonzero  $a_3$ . Therefore  $\mathbf{y}_1$ ,  $\mathbf{y}_2$  and the null vector (of order 3) are linearly dependent. This is true generally: vectors that

include a null vector are always linearly dependent. This is why null vectors are excluded from the definitions of linearly dependent and linearly independent vectors. Henceforth, therefore, whenever linearly dependent or linearly independent vectors are mentioned, they never include a null vector.

**Notation.** The abbreviation LIN shall be used to mean "linear independence" or "linearly independent," depending on context.

#### 4. LINEARLY DEPENDENT VECTORS

##### a. At least two $a$ 's are nonzero

Vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$  are linearly dependent when

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p = \mathbf{0} \text{ for not all the } a\text{'s being zero.} \quad (17)$$

Suppose only one  $a$  is nonzero— $a_2$ , say. Then (17) is  $a_2\mathbf{x}_2 = \mathbf{0}$ , and because null vectors are excluded,  $\mathbf{x}_2$  is not null and so  $a_2 = 0$ . This contradicts the supposition. Therefore more than one  $a$  is nonzero, i.e., at least two of them are. Also, of course, the definition of linearly dependent vectors demands that there be at least two of them, and so at least two  $a$ 's are nonzero.

##### b. Vectors are linear combinations of others

Suppose in (17) that  $a_1$  and  $a_2$  are nonzero. (There is no loss of generality in this identifying of any two nonzero  $a$ 's as  $a_1$  and  $a_2$ .) Then from (17)

$$\mathbf{x}_1 + (a_2/a_1)\mathbf{x}_2 + \dots + (a_p/a_1)\mathbf{x}_p = \mathbf{0} \quad (18)$$

so that

$$\mathbf{x}_1 = -(a_2/a_1)\mathbf{x}_2 - (a_3/a_1)\mathbf{x}_3 - \dots - (a_p/a_1)\mathbf{x}_p;$$

i.e.,  $\mathbf{x}_1$  can be expressed as a linear combination of the other  $\mathbf{x}$ 's. And this can always be done because at least two  $a$ 's are always nonzero. For example, in (14) we have  $2\mathbf{x}_1 + 3\mathbf{x}_2 - 3\mathbf{x}_3 = \mathbf{0}$ , so that

$$\mathbf{x}_1 = -1\frac{1}{2}\mathbf{x}_2 + 1\frac{1}{2}\mathbf{x}_3. \quad (19)$$

And, of course, this can also be stated in other equivalent ways such as

$$\mathbf{x}_3 = \frac{2}{3}\mathbf{x}_1 + \mathbf{x}_2. \quad (20)$$

Similarly, (13) gives  $\mathbf{x}_1 = -\frac{1}{2}\mathbf{x}_4$  and

$$\mathbf{x}_4 = -2\mathbf{x}_1. \quad (21)$$

In Section 3b(iii) the existence rather than the individual values of nonzero  $a$ 's was emphasized. This emphasis carries over here to the existence (among dependent vectors) of linear combinations of some of them being equal to others: it is the existence of such combinations that is important rather than the specific scalars that multiply the vectors in those combinations.

##### c. Partitioning matrices

As a result of (20) and (21) we can write, for the example,

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3 \ \mathbf{x}_4] = [\mathbf{x}_1 \ \mathbf{x}_2 \ (\frac{2}{3}\mathbf{x}_1 + \mathbf{x}_2) \ (-2\mathbf{x}_1 + 0\mathbf{x}_2)]. \quad (22)$$

With the last two columns being linear combinations of the first two,  $\mathbf{X}$  can be rewritten further as

$$\begin{aligned} \mathbf{X} &= \left[ \mathbf{x}_1 \ \mathbf{x}_2 \ (\mathbf{x}_1 \ \mathbf{x}_2) \begin{pmatrix} \frac{2}{3} \\ 1 \end{pmatrix} \ (\mathbf{x}_1 \ \mathbf{x}_2) \begin{pmatrix} -2 \\ 0 \end{pmatrix} \right] \\ &= \left[ (\mathbf{x}_1 \ \mathbf{x}_2) \ (\mathbf{x}_1 \ \mathbf{x}_2) \begin{pmatrix} \frac{2}{3} & -2 \\ 1 & 0 \end{pmatrix} \right] \\ &= [\mathbf{X}_1 \ \mathbf{X}_1\mathbf{K}], \text{ for } \mathbf{X}_1 = [\mathbf{x}_1 \ \mathbf{x}_2] \text{ and } \mathbf{K} = \begin{bmatrix} \frac{2}{3} & -2 \\ 1 & 0 \end{bmatrix} \quad (23) \\ &= \mathbf{X}_1[\mathbf{I} \ \mathbf{K}]. \quad (24) \end{aligned}$$

Thus from (22), in which the last two columns of  $\mathbf{X}$  are linear combinations of the first two, we have in (23) partitioned  $\mathbf{X}$  into two matrices:  $\mathbf{X}_1$ , the first two columns of  $\mathbf{X}$ , and  $\mathbf{X}_1\mathbf{K}$ , the matrix  $\mathbf{X}_1$  postmultiplied by a matrix  $\mathbf{K}$  which expresses columns 3 and 4 of  $\mathbf{X}$  as linear combinations of columns 1 and 2, i.e., of columns of  $\mathbf{X}_1$ . And in (24) this is further expressed as the product of  $\mathbf{X}_1$  and the partitioned matrix  $[\mathbf{I} \ \mathbf{K}]$ .

This partitioning can be carried out for any  $\mathbf{X}$  having columns 1, 2, ...,  $p$  represented by  $\mathbf{X}_1$  where its remaining columns are linear combinations of those in  $\mathbf{X}_1$ . Then

$$\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_1\mathbf{K}] = \mathbf{X}_1[\mathbf{I} \ \mathbf{K}] \quad (25)$$

for some matrix  $\mathbf{K}$ . Again, interest does not lie in the specific value of  $\mathbf{K}$  but in the fact that such a  $\mathbf{K}$  always exists. Applications of (25) can also be made to cases where those columns of  $\mathbf{X}$  which are linear combinations of others occur in  $\mathbf{X}$  not just as the last (right-hand) columns. The last paragraph of Section 9 illustrates this.

**d. Zero determinants**

Suppose  $p$  linearly dependent vectors of order  $p$  are used as columns of a matrix. We have just seen that linear dependence of vectors implies that one vector (column of the matrix) can always be expressed as a linear combination of the others. Therefore subtracting from that column the linear combination of the other columns which it equals gives a zero column, and so the determinant is zero.

**Example**

$$|x_1 \ x_2 \ x_3| = \begin{vmatrix} 3 & 0 & 2 \\ -6 & 5 & 1 \\ 9 & -5 & 1 \end{vmatrix}$$

Since subtracting  $(-1\frac{1}{2}x_2 + 1\frac{1}{2}x_3)$  from  $x_1$ , i.e.,  $x_1 - (-1\frac{1}{2}x_2 + 1\frac{1}{2}x_3)$  is by (19) just subtracting  $x_1$  from itself, carrying out this subtraction in the determinant gives it a column of zeros, and so it is zero:

$$|x_1 \ x_2 \ x_3| = \begin{vmatrix} 0 & 0 & 2 \\ 0 & 5 & 1 \\ 0 & -5 & 1 \end{vmatrix} = 0.$$

**e. Inverse matrices**

The preceding result applies to any set of  $p$  linearly dependent vectors of order  $p$ : the determinant of the matrix having those vectors as columns is always zero. (The same is true, of course, if the vectors used are rows.) Consequently, the matrix so formed has no inverse. Hence when the columns (or rows) of a square matrix are linearly dependent, that matrix has no inverse. It is singular.

**f. Testing for dependence (simple cases)**

A simple test for linear dependence among  $p$  vectors of order  $p$  is to evaluate the determinant of the matrix formed from using the vectors as columns. A zero determinant means the vectors are linearly dependent, otherwise they are LIN.

**Example**

$$|x_1 \ x_2 \ x_3| = \begin{vmatrix} 3 & 0 & 2 \\ -6 & 5 & 1 \\ 9 & -5 & 1 \end{vmatrix} = \begin{vmatrix} 3 & 0 & 2 \\ -6 & 5 & 1 \\ 3 & 0 & 2 \end{vmatrix} = 0$$

and so  $x_1$ ,  $x_2$  and  $x_3$  are linearly dependent, whereas

$$|x_1 \ x_2 \ x_5| = \begin{vmatrix} 3 & 0 & 2 \\ -6 & 5 & -3 \\ 9 & -5 & 3 \end{vmatrix} = \begin{vmatrix} 3 & 0 & 2 \\ -6 & 5 & -3 \\ 3 & 0 & 0 \end{vmatrix} = -30 \neq 0$$

and so  $x_1$ ,  $x_2$  and  $x_5$  are LIN.

The definitions of linear dependence and independence of vectors represented as columns of a matrix  $\mathbf{X}$  are based on whether the equations  $\mathbf{Xa} = \mathbf{0}$  are satisfied for some non-null  $\mathbf{a}$  (dependence) or are satisfied only for  $\mathbf{a} = \mathbf{0}$  (independence). Given a set of vectors, their dependence or independence can therefore be ascertained by attempting to solve  $\mathbf{Xa} = \mathbf{0}$ . If a solution can be found other than  $\mathbf{a} = \mathbf{0}$  (which is always a solution), then it will be a non-null solution, and the vectors will be dependent. Otherwise they are LIN.

**Examples.** For  $\mathbf{X} = [x_1 \ x_2 \ x_3]$  equations  $\mathbf{Xa} = \mathbf{0}$  are

$$\begin{aligned} 3a_1 + 2a_3 &= 0 \\ -6a_1 + 5a_2 + a_3 &= 0 \\ 9a_1 - 5a_2 + a_3 &= 0 \end{aligned} \quad \text{with solution } \mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ -3 \end{bmatrix}.$$

Since  $\mathbf{a} \neq \mathbf{0}$ , the vectors  $x_1$ ,  $x_2$  and  $x_3$  are linearly dependent. But for  $\mathbf{X} = [x_1 \ x_2 \ x_5]$ , equations  $\mathbf{Xa} = \mathbf{0}$  are

$$\begin{aligned} 3a_1 + 2a_3 &= 0 \\ -6a_1 + 5a_2 - 3a_3 &= 0 \\ 9a_1 - 5a_2 + 3a_3 &= 0 \end{aligned}$$

Adding the last two of these equations gives  $3a_1 = 0$ , i.e.,  $a_1 = 0$ . Hence from the first,  $a_3 = 0$  and so in the second,  $a_2 = 0$ . Thus  $\mathbf{a} = [0 \ 0 \ 0]^T = \mathbf{0}$  is the only solution and so  $x_1$ ,  $x_2$  and  $x_5$  are LIN.

This principle can also be used when  $\mathbf{X}$  is rectangular: for  $\mathbf{X} = [x_1 \ x_2 \ x_3 \ x_4]$  equations  $\mathbf{Xa} = \mathbf{0}$  are

$$\begin{aligned} 3a_1 + 2a_3 - 6a_4 &= 0 \\ -6a_1 + 5a_2 + a_3 + 12a_4 &= 0 \\ 9a_1 - 5a_2 + a_3 - 18a_4 &= 0 \end{aligned} \quad \text{with solution } \mathbf{a} = \begin{bmatrix} a_4 \\ -1\frac{1}{2}a_4 \\ 1\frac{1}{2}a_4 \\ a_4 \end{bmatrix},$$

which is clearly non-null for any nonzero  $a_4$ . Hence  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ ,  $\mathbf{x}_3$  and  $\mathbf{x}_4$  are linearly dependent. Similarly, for  $\mathbf{x}_1$  and  $\mathbf{x}_3$  we have  $\mathbf{X}\mathbf{a} = \mathbf{0}$  as

$$\begin{aligned} 3a_1 + 2a_2 &= 0 \\ -6a_1 + a_2 &= 0 \\ 9a_1 + a_2 &= 0 \end{aligned}$$

From the first two equations  $5a_2 = 0$ , i.e.,  $a_2 = 0$  and so  $a_1 = 0$ . Hence  $\mathbf{a} = [0 \ 0]' = \mathbf{0}$  is the only solution and so  $\mathbf{x}_1$  and  $\mathbf{x}_3$  are LIN. But for  $\mathbf{x}_1$  and  $\mathbf{x}_4$  the equations to be solved are

$$\begin{aligned} 3a_1 - 6a_2 &= 0 \\ -6a_1 + 12a_2 &= 0; \\ 9a_1 - 18a_2 &= 0 \end{aligned}$$

$a_1 = 2$  and  $a_2 = 1$  are solutions for the first two equations, and they also satisfy the third. Thus  $\mathbf{a} = [2 \ 1]' \neq \mathbf{0}$  is a solution, and so  $\mathbf{x}_1$  and  $\mathbf{x}_4$  are linearly dependent.

Solving equations is easy when only a few vectors of small order are involved, as in these examples. But for more extensive situations the arithmetic can be tedious. Furthermore, since easier procedures are available for ascertaining whether or not  $\mathbf{X}\mathbf{a} = \mathbf{0}$  has a non-null solution other than solving the equations explicitly, we proceed towards developing those procedures. Before doing so, notice the equivalence of the following statements for square  $\mathbf{X}$  that has no null columns:

- (i) Columns of  $\mathbf{X}$  are linearly dependent.
- (ii)  $\mathbf{X}\mathbf{a} = \mathbf{0}$  can be satisfied for a non-null  $\mathbf{a}$ .
- (iii)  $\mathbf{X}$  is singular, i.e.,  $\mathbf{X}^{-1}$  does not exist.

The definition of linear dependence assures the equivalence of (i) and (ii), and subsection e shows the equivalence of these to (iii).

## 5. LINEARLY INDEPENDENT (LIN) VECTORS

### a. Nonzero determinants and inverse matrices

The only value of  $\mathbf{a}$  for which  $\mathbf{X}\mathbf{a} = \mathbf{0}$  when the columns of  $\mathbf{X}$  are LIN is  $\mathbf{a} = \mathbf{0}$  in which case no equation like (18) exists. Therefore no column in  $\mathbf{X}$  can be expressed as a linear combination of others in  $\mathbf{X}$ . Hence when  $\mathbf{X}$  has  $p$  columns of order  $p$ , i.e., when  $\mathbf{X}$  is square, its determinant is nonzero, and so its inverse does exist. Hence for square  $\mathbf{X}$  the following three statements

are equivalent:

- (i) Columns of  $\mathbf{X}$  are LIN.
- (ii)  $\mathbf{X}\mathbf{a} = \mathbf{0}$  only for  $\mathbf{a} = \mathbf{0}$ .
- (iii)  $\mathbf{X}$  is nonsingular, i.e.,  $\mathbf{X}^{-1}$  exists.

### b. Linear combinations of LIN vectors

When  $p$  vectors of order  $p$  are LIN, any other vector of order  $p$  can be expressed as a linear combination of those  $p$  vectors. Suppose  $\mathbf{X}$  is the matrix of the  $p$  vectors used as columns, and let  $\mathbf{v}$  be some other vector of the same order. Then this result states that there exists a vector  $\mathbf{m}$  for which  $\mathbf{X}\mathbf{m}$ , which is a linear combination of columns of  $\mathbf{X}$  [see (7)], has the property  $\mathbf{X}\mathbf{m} = \mathbf{v}$ . This is so because with  $\mathbf{X}$  being square and having LIN columns,  $\mathbf{X}^{-1}$  exists and so  $\mathbf{X}\mathbf{m} = \mathbf{v}$  implies  $\mathbf{m} = \mathbf{X}^{-1}\mathbf{v}$ .

**Example.** Any vector  $\mathbf{v}$  of order 2 can be expressed as a linear combination of the two LIN vectors  $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$  and  $\mathbf{x}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ . Let  $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$  and  $\mathbf{X} = \begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix}$ . Then

$$\mathbf{m} = \mathbf{X}^{-1}\mathbf{v} = (1/7) \begin{bmatrix} 5 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (5a - 3b)/7 \\ (-a + 2b)/7 \end{bmatrix};$$

$$\begin{aligned} \text{i.e.,} \quad \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} &= \mathbf{X}\mathbf{m} = \begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} (5a - 3b)/7 \\ (-a + 2b)/7 \end{bmatrix} \\ &= \frac{5a - 3b}{7} \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{-a + 2b}{7} \begin{bmatrix} 3 \\ 5 \end{bmatrix}. \end{aligned} \quad (26)$$

The reader should verify in a similar manner that

$$\begin{aligned} \begin{bmatrix} a \\ b \\ c \end{bmatrix} &= \frac{9a + 48b - 15c}{6} \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} + \frac{3a + 30b - 9c}{6} \begin{bmatrix} -1 \\ 3 \\ 9 \end{bmatrix} \\ &+ \frac{-3a - 22b + 7c}{6} \begin{bmatrix} 3 \\ 6 \\ 21 \end{bmatrix}. \end{aligned} \quad (27)$$

### c. A maximum number of LIN vectors

A question that might be asked about independent vectors is "How many independent vectors are there?" The answer is contained in the following theorem.



**Theorem.** A set of LIN vectors of order  $n$  cannot contain more than  $n$  such vectors.

**Proof.** Let  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  be  $n$  LIN vectors of order  $n$ . Let  $\mathbf{u}_{n+1}$  be any other non-null vector of order  $n$ . We show that it and  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  are linearly dependent.

Since  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$  has LIN columns,  $|\mathbf{U}| \neq 0$  and  $\mathbf{U}^{-1}$  exists. Let  $\mathbf{q} = -\mathbf{U}^{-1}\mathbf{u}_{n+1} \neq \mathbf{0}$ , because  $\mathbf{u}_{n+1} \neq \mathbf{0}$ ; i.e., not all elements of  $\mathbf{q}$  are zero. Then  $\mathbf{U}\mathbf{q} + \mathbf{u}_{n+1} = \mathbf{0}$ , which can be rewritten as

$$q_1\mathbf{u}_1 + q_2\mathbf{u}_2 + \dots + q_n\mathbf{u}_n + \mathbf{u}_{n+1} = \mathbf{0} \quad (28)$$

with not all the  $q$ 's being zero. This means  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n+1}$  are linearly dependent; i.e., with  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  being LIN it is impossible to put another vector  $\mathbf{u}_{n+1}$  with them and have all  $n+1$  vectors being LIN. QED

**Corollary.** When  $p$  vectors of order  $n$  are LIN then  $p \leq n$ .

It is important to note that this theorem is *not* stating that there is only a single set of  $n$  vectors of order  $n$  that are LIN. What it is saying is that if we do have a set of  $n$  LIN vectors of order  $n$ , then there is no larger set of LIN vectors of order  $n$ ; i.e., there are no sets of  $n+1, n+2, n+3, \dots$  vectors of order  $n$  that are LIN. Although there are many sets of  $n$  LIN vectors of order  $n$ , an infinite number of them in fact, for each of them it is impossible to put another vector (or vectors) with them and have the set, which then contains more than  $n$  vectors, still be LIN.

#### Example

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

are a set of two LIN vectors of order 2. Put any other vector of order 2 with these two vectors and the set will be linearly dependent, e.g.,

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix} \text{ and } \begin{bmatrix} 7 \\ 8 \end{bmatrix} \text{ are linearly dependent;}$$

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix} \text{ and } \begin{bmatrix} a \\ b \end{bmatrix} \text{ are linearly dependent.}$$

This latter result is true of course because, as (26) shows,  $\begin{bmatrix} a \\ b \end{bmatrix}$  is a linear combination of  $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$  and  $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ . But the theorem is not saying that  $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$  and  $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$  is the only set of two LIN vectors of order 2. For example,  $\begin{bmatrix} 5 \\ 7 \end{bmatrix}$  and

$\begin{bmatrix} 9 \\ 13 \end{bmatrix}$  also form such a set; and any other vector of order 2 put with them forms a linearly dependent set of three vectors. This is so because that third vector is a linear combination of these two; e.g.,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{2}(13a - 9b)\begin{bmatrix} 5 \\ 7 \end{bmatrix} + \frac{1}{2}(-7a + 5b)\begin{bmatrix} 9 \\ 13 \end{bmatrix}.$$

No matter what the values of  $a$  and  $b$  are, this expression holds true—i.e., every second-order vector can be expressed as a linear combination of a set of two LIN vectors of order 2. In general, every  $n$ th-order vector can be expressed as a linear combination of any set of  $n$  independent vectors of order  $n$ . The maximum number of non-null vectors in a set of independent vectors is therefore  $n$ , the order of the vectors. This is simply a restatement of the theorem.

## 6. THE NUMBER OF LIN ROWS AND COLUMNS IN A MATRIX

Explanation has been given of how a determinant is zero when any of its rows (or columns) are linear combinations of other rows (or columns). In other words, a determinant is zero when its rows (or columns) do not form a set of LIN vectors. Evidently, therefore, a determinant cannot have both its rows forming a dependent set and its columns an independent set, a statement which prompts the more general question of the relationship between the number of LIN rows of a matrix and the number of LIN columns. The relationship is simple.

**Theorem.** The number of LIN rows in a matrix is the same as the number of LIN columns.

Before proving this, notice that independence of rows (columns) is a property of rows that is unrelated to their sequence within a matrix. For example, because the rows of

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 9 & 14 \\ 3 & 0 & 1 \end{bmatrix} \text{ are LIN, so also are the rows of } \begin{bmatrix} 6 & 9 & 14 \\ 3 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

Therefore, insofar as general discussion of independence properties is concerned, there is no loss of generality for a matrix that has  $k$  LIN rows in assuming that they are the first  $k$  rows. This assumption is therefore often made when discussing properties and consequences of independence of rows (columns) of a matrix.

We now prove the theorem.

**Proof.** Let  $A_{p \times q}$  have  $k$  LIN rows and  $m$  LIN columns. We show that  $k = m$ . Assume that the first  $k$  rows of  $A$  are LIN and similarly the first  $m$  columns. Then partition  $A$  as

$$A_{p \times q} = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} = \begin{bmatrix} X_{k \times m} & Y_{k \times (q-m)} \\ Z_{(p-k) \times m} & W_{(p-k) \times (q-m)} \end{bmatrix} \leftarrow k \text{ LIN rows.} \quad (29)$$

$\uparrow$   
 $m$  LIN columns

Thus the  $k$  rows of  $A$  through  $X$  and  $Y$  are LIN (as are the  $m$  columns of  $A$  through  $X$  and  $Z$ ). Since  $A$  has only  $k$  LIN rows the other rows of  $A$  (those through  $Z$  and  $W$ ) are linear combinations of the first  $k$  rows. In particular, the rows of  $Z$  are linear combinations of the rows of  $X$ . Hence by (8) these rows can be expressed as  $Z = TX$  for some matrix  $T$ . Now assume that the columns of  $X$  are linearly dependent, i.e., that

$$Xa = 0 \quad \text{for some vector } a \neq 0. \quad (30)$$

Then  $Za = TXa = 0$  and so

$$\begin{bmatrix} X \\ Z \end{bmatrix} a = 0 \quad \text{for that same } a \neq 0.$$

But this is a statement of the linear dependence of the columns of  $\begin{bmatrix} X \\ Z \end{bmatrix}$ , i.e., of the first  $m$  columns of  $A$ . These columns, however, have been taken in (29) as being LIN. This is a contradiction, and so assumption (30), from which it is derived, is false; i.e., the columns of  $X$  are not dependent. Hence they must be LIN.

Having shown the columns of  $X$  to be LIN, observe from (29) that there are  $m$  of them and they are of order  $k$ . Hence by the theorem in the preceding section,  $m \leq k$ . A similar argument based on the rows of  $X$  and  $Y$ , rather than the columns of  $X$  and  $Z$ , shows that  $k \leq m$ . Hence  $m = k$ . QED

It is important to notice that this theorem says nothing about which rows (columns) of a matrix are LIN—it is concerned solely with how many of them are LIN. This means, for example, that if there are two LIN rows in a matrix of order  $5 \times 4$  then once two rows are ascertained as being LIN the other three rows can be expressed as linear combinations of those two. And there may be (there usually is) more than one set of two rows that are LIN. For example, in

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & -1 & 3 & 2 \\ 3 & 0 & 6 & 5 \\ 2 & 1 & 3 & 3 \\ 6 & 0 & 12 & 10 \end{bmatrix}$$

rows 1 and 2 are LIN, and each of rows 3, 4 and 5 is a linear combination of those first two. Rows 1 and 3 are also LIN, and the remaining rows are linear combinations of them; but not all pairs of rows are LIN. For example, rows 3 and 5 are not LIN.

### 7. THE RANK OF A MATRIX

The theorem in Section 6 shows that every matrix has the same number of LIN rows as it does LIN columns.

**Definition.** The *rank* of a matrix is the number of linearly independent rows (and columns) in the matrix.

**Notation.** The rank of  $A$  will be denoted equivalently by  $r_A$  or  $r(A)$ . Thus if  $r_A \equiv r(A) = k$ , then  $A$  has  $k$  LIN rows and  $k$  LIN columns. The symbol  $r$  is often used for rank, i.e.,  $r_A \equiv r$ .

Notice again that no specific set of LIN rows (columns) is identified by knowing the rank of a matrix. Rank indicates only how many are LIN and not where they are located in the matrix. The following properties and consequences of rank are important.

- (i)  $r_A$  is a positive integer, except that  $r_0$  is defined as  $r_0 = 0$ .
- (ii)  $r(A_{p \times q}) \leq p$  and  $\leq q$ : the rank of a matrix equals or is less than the smaller of its number of rows or columns.
- (iii)  $r(A_{n \times n}) \leq n$ : a square matrix has rank not exceeding its order.
- (iv) When  $r_A = r \neq 0$  there is at least one square submatrix of  $A$  having order  $r$  that is nonsingular. Equation (29) with  $k = m = r$  is

$$A_{p \times q} = \begin{bmatrix} X_{r \times r} & Y_{r \times (q-r)} \\ Z_{(p-r) \times r} & W_{(p-r) \times (q-r)} \end{bmatrix} \quad (31)$$

and  $X_{r \times r}$ , the intersection of  $r$  LIN rows and  $r$  LIN columns, is nonsingular—as is evident from the proof of the theorem in Section 3. All square submatrices of order greater than  $r$  are singular.

- (v) When  $r(A_{n \times n}) = n$  then by (iv)  $A$  is nonsingular, i.e.,  $A^{-1}$  exists. [In (31)  $A \equiv X$ .]
- (vi) When  $r(A_{n \times n}) < n$  then  $A$  is singular and  $A^{-1}$  does not exist.
- (vii) When  $r(A_{p \times q}) = p < q$ ,  $A$  is said to have *full row rank*, or to be of full row rank. Its rank equals its number of rows.

- (viii) When  $r(\mathbf{A}_{p \times q}) = q < p$ ,  $\mathbf{A}$  is said to have *full column rank*, or to be of full column rank. Its rank equals its number of columns.
- (ix) When  $r(\mathbf{A}_{n \times n}) = n$ ,  $\mathbf{A}$  is said to have *full rank*, or to be of full rank. Its rank equals its order, it is nonsingular, its inverse exists, and it is said to be *invertible*.

Rank is one of the most useful and important characteristics of any matrix. It occurs again and again in this book and plays a vital role throughout all aspects of matrix algebra. For example, from items (vi) and (ix) we see at once that ascertaining whether  $|\mathbf{A}|$  is zero or not for determining the existence of  $\mathbf{A}^{-1}$  can be replaced by ascertaining whether  $r_{\mathbf{A}} < n$  or  $r_{\mathbf{A}} = n$ . And almost always it is far easier to work with rank than determinants.

### 8. RANK AND INVERSE MATRICES

A square matrix has an inverse if and only if its rank equals its order. This and other equivalent statements, including those at the end of Sections 4f and 5a, are summarized in Table 6.1. In each half of the table any one of the statements implies all the others: the first and second statements are basically definitional and the last six are equivalences. Hence whenever assurance is needed for the existence of  $\mathbf{A}^{-1}$ , we need only establish any one of the last six statements in the first column of the table. The easiest of these is usually that concerning rank: when  $r_{\mathbf{A}} = n$  then  $\mathbf{A}^{-1}$  exists, and when  $r_{\mathbf{A}} < n$  then  $\mathbf{A}^{-1}$  does not exist. The problem of ascertaining the existence of an inverse is therefore equivalent to ascertaining if the rank of a square matrix is less than its order.

More generally, there are often occasions when the rank of a matrix is needed exactly. Although locating  $r_{\mathbf{A}}$  LIN rows in  $\mathbf{A}$  may not always be easy,

TABLE 6.1. EQUIVALENT STATEMENTS FOR THE EXISTENCE OF  $\mathbf{A}^{-1}$  OF ORDER  $n$

Inverse existing	Inverse not existing
$\mathbf{A}^{-1}$ exists	$\mathbf{A}^{-1}$ does not exist
$\mathbf{A}$ is nonsingular	$\mathbf{A}$ is singular
$ \mathbf{A}  \neq 0$	$ \mathbf{A}  = 0$
$\mathbf{A}$ has full rank	$\mathbf{A}$ has less than full rank
$r_{\mathbf{A}} = n$	$r_{\mathbf{A}} < n$
$\mathbf{A}$ has $n$ LIN rows	$\mathbf{A}$ has fewer than $n$ LIN rows
$\mathbf{A}$ has $n$ LIN columns	$\mathbf{A}$ has fewer than $n$ LIN columns
$\mathbf{A}\mathbf{x} = \mathbf{0}$ has sole solution, $\mathbf{x} = \mathbf{0}$	$\mathbf{A}\mathbf{x} = \mathbf{0}$ has many solutions $\mathbf{x} \neq \mathbf{0}$

deriving  $r_{\mathbf{A}}$  itself is conceptually not difficult. General procedures for doing this are described in the next chapter.

### 9. PERMUTATION MATRICES

Proof of the theorem in Section 6 begins by assuming that all the LIN rows come first in a matrix, and the LIN columns likewise. But this is not always so. For example, in

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 4 & 4 & 12 \\ 2 & 2 & 5 \end{bmatrix}$$

there are two LIN rows and two LIN columns—but not the first two. Now consider  $\mathbf{E}_{24}\mathbf{M}$  for

$$\mathbf{E}_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix};$$

$$\mathbf{E}_{24}\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 4 & 4 & 12 \\ 2 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 4 & 12 \\ 1 & 1 & 3 \end{bmatrix}. \quad (32)$$

Note that  $\mathbf{E}_{24}$  is an identity matrix with its second and fourth rows interchanged, and  $\mathbf{E}_{24}\mathbf{M}$  is  $\mathbf{M}$  with those same two rows interchanged.

$\mathbf{E}_{24}$  of (32) exemplifies what shall be called an *elementary permutation matrix*: in general,  $\mathbf{E}_{rs}$  is an identity matrix with its  $r$ th and  $s$ th rows interchanged. And  $\mathbf{E}_{rs}\mathbf{M}$  is  $\mathbf{M}$  with those same two rows interchanged. (The order of  $\mathbf{E}_{rs}$  is determined by the product it is used in.) By virtue of its definition,  $\mathbf{E}_{rs}$  is always symmetric. It is also orthogonal. Thus  $\mathbf{E}_{rs}\mathbf{E}'_{rs} = \mathbf{E}_{rs}\mathbf{E}_{rs} = \mathbf{I}$ , the second equality being true because  $\mathbf{E}_{rs}\mathbf{E}_{rs}$  is  $\mathbf{E}_{rs}$  with its  $r$ th and  $s$ th rows interchanged and so is  $\mathbf{I}$ , i.e.,  $\mathbf{E}_{rs}\mathbf{E}_{rs} = \mathbf{I}$ .

In the same way that premultiplication of  $\mathbf{M}$  by  $\mathbf{E}_{rs}$  interchanges rows  $r$  and  $s$  of  $\mathbf{M}$ , so does postmultiplication interchange columns. Thus using (32)

$$\mathbf{E}_{24}\mathbf{M}\mathbf{E}_{23} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 4 & 12 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 5 & 2 \\ 4 & 12 & 4 \\ 1 & 3 & 1 \end{bmatrix} \quad (33)$$

is  $\mathbf{E}_{24}\mathbf{M}$  with its second and third columns interchanged.

Note in (33) that  $E_{24}ME_{23}$  has its first two rows LIN and its first two columns LIN also. Hence by premultiplying  $M$  by one elementary permutation matrix and postmultiplying it by another we get a matrix to which the argument of the proof of the theorem applies. And so, because the theorem applies only to the number of LIN rows and LIN columns, whatever holds for (33) in this connection also holds for  $M$ .

This use of permutation matrices extends to cases involving more than just a single interchange of rows and/or columns. Consider

$$A = \begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 3 & 2 \\ 3 & 3 & 9 & 6 \\ 2 & 2 & 5 & 4 \\ 1 & 1 & 7 & 8 \end{bmatrix}. \quad (34)$$

In premultiplying  $A$  first by  $E_{34}$  (to interchange rows 3 and 4) and then by  $E_{25}$  (to interchange rows 2 and 5), define  $P$  as  $P = E_{25}E_{34}$ . Then

$$PA = E_{25}E_{34}A = E_{25} \begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 3 & 2 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 9 & 6 \\ 1 & 1 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 7 & 8 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 9 & 6 \\ 1 & 1 & 3 & 2 \end{bmatrix}$$

where

$$P = E_{25}E_{34} = E_{25} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Then, for  $Q = E_{24}$ ,

$$PAQ = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 8 & 7 & 1 \\ 2 & 4 & 5 & 2 \\ 3 & 8 & 9 & 3 \\ 1 & 2 & 3 & 1 \end{bmatrix}$$

and this is a matrix with its first three rows LIN and its first three columns likewise.

In this case  $P$  is a product of two  $E$ -matrices. It is an example of a *permutation matrix* which, in general, is an identity matrix with its rows resequenced in some order. Because such a matrix is always a product of

elementary permutation matrices (the  $E$ -matrices),  $P$  is not necessarily symmetric, but it is always orthogonal—because it is a product of orthogonal  $E$ -matrices (see Exercise 3.13). Furthermore,  $P^{-1} = P'$  is also a permutation matrix, because although  $P$  is defined as an identity matrix with its rows resequenced, it is also an identity matrix with its columns resequenced. Therefore  $P'$  is an identity matrix with its rows resequenced and so  $P^{-1} = P'$  is a permutation matrix, too.

One of the great uses of permutation matrices is that for situations like that of the theorem in Section 6, permutation matrices provide a mechanism for resequencing rows and columns in a matrix so that a matrix having  $k$  LIN rows can be resequenced into one having its first  $k$  rows and its first  $k$  columns LIN. Properties of the permutation matrices then allow many properties concerning linear independence of the resequenced matrices to also apply to the original matrix.

## 10. FULL-RANK FACTORIZATION

An immediate consequence of the notion of rank is that a  $p \times q$  matrix of rank  $r \neq 0$  can be partitioned into a group of  $r$  independent rows and a group of  $p - r$  rows that are linear combinations of the first group. This leads to a useful factorization.

### a. Basic development

Equation (31) shows the partitioning of  $A_{p \times q}$  of rank  $r$  when the first  $r$  rows and first  $r$  columns are assumed to be LIN, and where  $X$  is nonsingular. The first  $r$  rows in (31) are those of  $[X \ Y]$  and are LIN, and so the rows of  $[Z \ W]$  are linear combinations of those of  $[X \ Y]$ ; hence for some matrix  $F$ ,

$$[Z \ W] = F[X \ Y]. \quad (35)$$

Similar reasoning applied to the columns of  $A$  gives, for some matrix  $H$ ,

$$\begin{bmatrix} Y \\ W \end{bmatrix} = \begin{bmatrix} X \\ Z \end{bmatrix} H. \quad (36)$$

But (35) has  $Z = FX$  and  $W = FY$ , and (36) has  $Y = XH$ . Hence  $W = FY = FXH$ , and so therefore

$$A = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} = \begin{bmatrix} X & XH \\ FX & FXH \end{bmatrix} = \begin{bmatrix} I \\ F \end{bmatrix} X [I \ H]. \quad (37)$$

Furthermore, because  $\mathbf{X}$  in (31) is nonsingular, its inverse  $\mathbf{X}^{-1}$  exists, so that  $\mathbf{Z} = \mathbf{FX}$  and  $\mathbf{Y} = \mathbf{XH}$  give  $\mathbf{F} = \mathbf{ZX}^{-1}$  and  $\mathbf{H} = \mathbf{X}^{-1}\mathbf{Y}$  and then  $\mathbf{W} = \mathbf{FY} = \mathbf{ZX}^{-1}\mathbf{Y}$ . Hence (37) can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{ZX}^{-1}\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{ZX}^{-1} \end{bmatrix} \mathbf{X}[\mathbf{I} \quad \mathbf{X}^{-1}\mathbf{Y}]. \quad (38)$$

Equation (37) can be further rewritten in two equivalent ways as

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} \\ \mathbf{FX} \end{bmatrix} [\mathbf{I} \quad \mathbf{H}] = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \end{bmatrix} [\mathbf{X} \quad \mathbf{XH}], \quad (39)$$

each of which is of the form

$$\mathbf{A}_{p \times q} = \mathbf{K}_{p \times r} \mathbf{L}_{r \times q} \quad (40)$$

where  $\mathbf{K}$  has full column rank  $r = r_{\mathbf{A}}$ , and  $\mathbf{L}$  has full row rank  $r$ . We call the matrix product in (40) the full-rank factorization, after Ben-Israel and Greville (1974, p. 22). It has also been called the full rank decomposition by Marsaglia and Styan (1974a, p. 271). Whatever its name, it can always be done (see subsection b, which follows) and it has many uses in matrix algebra.

**Example.** Equation (38) is illustrated by

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \\ 4 & 9 & 3 & 24 \end{bmatrix} \\ &= \begin{bmatrix} & & & \mathbf{I}_2 \\ (4 \ 9) & \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 5 \\ 1 & 14 \end{pmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & -3 \\ 0 & 1 & -1 & 4 \end{bmatrix} \quad (41) \end{aligned}$$

and from (39) we have the two equivalent forms of (41):

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & -3 \\ 0 & 1 & -1 & 4 \end{bmatrix}. \quad (42) \end{aligned}$$

A particular case of interest is when  $r_{\mathbf{A}} = 1$ , whereupon  $\mathbf{A} = \mathbf{xy}'$ ; e.g.,

$$\begin{bmatrix} 2 & -2 & 4 \\ 3 & -3 & 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} [1 \quad -1 \quad 2].$$

### b. The general case

Development of (35)–(39) rests upon the first  $r$  rows (and  $r$  columns) of  $\mathbf{A}$  in (31) being LIN. But suppose this is not the case, as in (34). It is in just such a situation that permutation matrices play their part. Let  $\mathbf{M} = \mathbf{PAQ}$  where  $\mathbf{P}$  and  $\mathbf{Q}$  are permutation matrices, as discussed in Section 9, and where  $\mathbf{M}$  has the form (31). Then  $\mathbf{M}$  can be expressed as  $\mathbf{M} = \mathbf{KL}$  as in (40). Therefore, on using the orthogonality of  $\mathbf{P}$  and  $\mathbf{Q}$  (e.g.,  $\mathbf{P}^{-1} = \mathbf{P}'$ , as at the end of Section 9),

$$\mathbf{A} = \mathbf{P}^{-1} \mathbf{M} \mathbf{Q}^{-1} = \mathbf{P}' \mathbf{K} \mathbf{L} \mathbf{Q}' = (\mathbf{P}' \mathbf{K})(\mathbf{L} \mathbf{Q}') \quad (43)$$

where  $\mathbf{P}' \mathbf{K}$  and  $\mathbf{L} \mathbf{Q}'$  here play the same roles as  $\mathbf{K}$  and  $\mathbf{L}$  do in (40). Hence (40) can be derived quite generally, for any matrix.

### c. Matrices of full row (column) rank

When  $\mathbf{A}_{p \times q}$  has full row rank, there will be no  $\mathbf{Z}$  and  $\mathbf{W}$  in (31); and  $\mathbf{K}$  of (40) will be an identity matrix. Then (31) is  $\mathbf{A} = [\mathbf{X} \quad \mathbf{Y}]$ , and if  $\mathbf{X}$  is singular, postmultiplication of  $\mathbf{A}$  by a permutation matrix  $\mathbf{Q}$  can lead to the partitioning

$$\mathbf{A} \mathbf{Q} = [\mathbf{M} \quad \mathbf{L}]$$

where  $\mathbf{M}$  is nonsingular. We then have the following lemma.

**Lemma.** A matrix of full row rank can always be written as a product of matrices one of which has the partitioned form  $[\mathbf{I} \quad \mathbf{H}]$  for some matrix  $\mathbf{H}$ .

**Proof.** From the preceding equations  $\mathbf{A} = [\mathbf{M} \quad \mathbf{L}] \mathbf{Q}^{-1} = \mathbf{M} [\mathbf{I} \quad \mathbf{M}^{-1} \mathbf{L}] \mathbf{Q}' = \mathbf{M} [\mathbf{I} \quad \mathbf{H}] \mathbf{Q}'$  for  $\mathbf{H} = \mathbf{M}^{-1} \mathbf{L}$ , and we have the desired result. ( $\mathbf{M}^{-1}$  exists because  $\mathbf{M}$  is nonsingular, and  $\mathbf{Q}^{-1} = \mathbf{Q}'$  because  $\mathbf{Q}$  is a permutation matrix.) QED

## 11. VECTOR SPACES

This section introduces the reader to ideas that are used extensively in many books on linear and matrix algebra. They involve extensions of geometry which, although difficult for some people, are important in current mathematical literature,

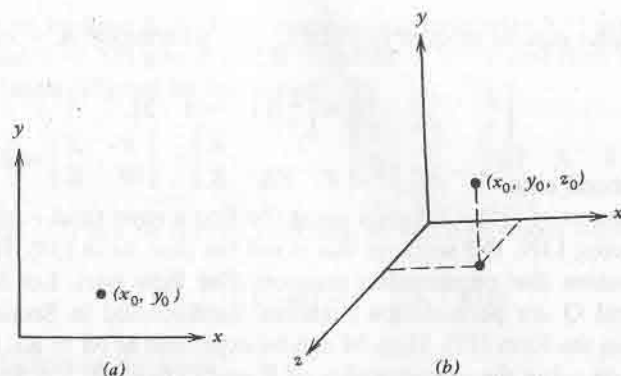


Figure 6.1

and so are described here as a convenient introductory reference for the reader. We confine ourselves, as usual, to real numbers and to vectors whose elements are real numbers.

#### a. Euclidean space

A vector  $[x_0 \ y_0]$  of order 2 can be thought of as representing a point on a plane using familiar Cartesian  $x, y$  coordinates, as in Figure 6.1a. Similarly, a vector  $[x_0 \ y_0 \ z_0]$  of order 3 can represent a point in three-dimensional space (Figure 6.1b). In similar manner a vector of order  $n$  can be said to represent a point in what is called  $n$ -dimensional space, or  $n$ -space. This appeal to geometry invites conceptualization of space as we ordinarily know it to more than three dimensions, a conceptualization that is not easy for some people. However, inasmuch as the set of all vectors of order 3 can be thought of as defining (all points in) 3-space, we use the word  $n$ -space to mean a representation of the set of all vectors of order  $n$ . Since it is a natural extension of familiar lines (1-space), planes (2-space) and the three-dimensional space in which we live, it is called Euclidean  $n$ -space. It will be denoted by  $R^n$ . Thus  $R^3$  is the space of the world we live in.

#### b. Vector spaces

$R^n$  is a special case of a wider concept called a vector space, which we now define.

Suppose  $x_i$  and  $x_j$  are two vectors in a set of vectors to be called  $S$ :

$$x_i \text{ and } x_j \text{ are in } S. \quad (44)$$

If, for every such pair  $x_i$  and  $x_j$  we also find that

$$x_i + x_j \text{ is in } S \quad (45)$$

and

$$ax_i \text{ is in } S, \text{ for all real scalars } a, \quad (46)$$

then  $S$  is said to be a *vector space*. When the vectors have order  $n$ ,  $S$  is a vector space of order  $n$ .

**Examples.** Euclidean  $n$ -space is a vector space of order  $n$ : for every pair of vectors in  $R^n$  their sum is in  $R^n$ , and so is the product of a scalar and any vector that is in  $R^n$ .

Viewed as a set of vectors,  $\mathbf{0}$  is a set with but a single vector in it and so (45) and (46) need be checked for the only possible choices,  $x_i = x_j = \mathbf{0}$ ; obviously they are satisfied. Hence  $\mathbf{0}$  constitutes a vector space. More than that,  $\mathbf{0}$  belongs to every vector space in  $R^n$ .

#### c. Spanning sets and bases

The defining conditions (44), (45) and (46) of a vector space ensure that all linear combinations of vectors that are in a vector space are also in the same vector space. Consider  $x_1, x_2$  and  $x_3$  in some vector space  $S$ . Then by (46), for real values  $a_1, a_2$  and  $a_3$ , we know that  $a_1x_1, a_2x_2$  and  $a_3x_3$  are in  $S$ , as is  $a_1x_1 + a_2x_2$  from (45); and similarly  $(a_1x_1 + a_2x_2) + a_3x_3 = a_1x_1 + a_2x_2 + a_3x_3$  is also in  $S$ .

Suppose that every vector in a vector space  $S$  can be expressed as a linear combination of the  $t$  vectors  $x_1, x_2, \dots, x_t$ . Then this set of vectors is said to *span* or *generate* the vector space  $S$ . It is called a *spanning set* for  $S$ . When the vectors of a spanning set are also LIN then the set is said to be a *basis* for  $S$ . It is a basis and not *the* basis because for any  $S$  there are many bases. All bases of a space  $S$  have the same number of vectors (see Exercise 7.17) and that number is called the *dimension* of  $S$ , or  $\dim(S)$ .

#### Example

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad x_3 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

are all in  $R^3$ . Any two of them form a basis for the vector space whose typical vector is  $[\alpha \ \beta \ 0]$  for  $\alpha$  and  $\beta$  real. The dimension of the space is 2. (The space in this case is, of course, a subspace of  $R^3$ .)

A basis for any vector space  $S$  consists of LIN vectors,  $\dim(S)$  of them in number; and all vectors in  $S$  are linear combinations of those vectors in the basis.

#### d. Many spaces of order $n$

It is important to realize that although there is only one Euclidean space of order  $n$ , there are many vector spaces of order  $n$ .

**Example.** Consider the set  $S_1$  of vectors  $x = [\alpha \ \beta \ \gamma \ 0 \ 0]$  for all real values  $\alpha, \beta$  and  $\gamma$ . They satisfy (44), (45) and (46) for  $S_1$ . Hence  $S_1$  is a vector space of order 5.

Clearly every vector in  $S_1$  is also in  $R^5$ ; but there are also vectors in  $R^5$  that are not in  $S_1$ , e.g.,  $[2 \ 1 \ 7 \ 3 \ -4]$ . Thus  $S_1$  and  $R^5$  are different vector spaces, although both of order 5.

## e. Subspaces

Because all vectors of  $S_1$  are also in  $R^5$ , and  $R^5$  has vectors that are not in  $S_1$ , the set of vectors in  $S_1$  is clearly a subset of those in  $R^5$ . Since  $S_1$  is also a vector space it is therefore called a *vector subspace* of  $R^5$ . All vector spaces of order  $n$  (involving vectors of real numbers) are subspaces of  $R^n$ , and some of them are also subspaces of others.

**Example.** Let  $S_2$ ,  $S_3$  and  $S_4$  be the sets of vectors having the forms  $x$ ,  $y$  and  $z$ , respectively:

$$x = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} 0 \\ 0 \\ \beta \end{bmatrix} \quad \text{and} \quad z = \begin{bmatrix} \gamma \\ 0 \\ \delta \end{bmatrix}, \quad \text{for real } \alpha, \beta, \gamma \text{ and } \delta.$$

Then  $S_2$ ,  $S_3$  and  $S_4$  each define a vector space, and they are all subspaces of  $R^3$ . Furthermore,  $S_2$  and  $S_3$  are each a subspace of  $S_4$ .

## f. The range and null space of a matrix

A matrix  $A$  of rank  $r_A$  has  $r_A$  LIN columns. All vectors that are linear combinations of those independent columns form a vector space. It is known as the *column space* of  $A$ , *range* of  $A$  or *manifold* of  $A$ , and is often denoted by  $\mathcal{R}(A)$ . It has dimension  $r_A$ :

$$r_A = \dim[\mathcal{R}(A)].$$

Any set of  $r_A$  LIN columns of  $A$  is a basis for  $\mathcal{R}(A)$ .  
The equations

$$\begin{bmatrix} 2 & 1 & 3 & -1 \\ 3 & 5 & 1 & 2 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

have many solutions for  $\mathbf{x}' = [x_1 \ x_2 \ x_3 \ x_4]$ ; for example,  $[-1 \ 0 \ 1 \ 1]$  and  $[-3 \ 1 \ 2 \ 1]$  are solutions. More generally, there are numerous situations when  $A\mathbf{x} = \mathbf{0}$  has many solutions for  $\mathbf{x}$ . When this is so, the space defined by the vectors  $\mathbf{x}$  for which  $A\mathbf{x} = \mathbf{0}$  is called the *null space* of  $A$ , denoted by  $\mathcal{N}(A)$ . Its dimension is called the *nullity* of  $A$ :

$$\text{nullity}(A) \equiv \dim[\mathcal{N}(A)].$$

Suppose  $\mathbf{X}'\mathbf{y} = \mathbf{0}$ , i.e.,  $\mathbf{y}'\mathbf{X} = \mathbf{0}'$ . Then  $\mathbf{y}'\mathbf{x} = 0$  for every column  $\mathbf{x}$  of  $\mathbf{X}$ ; i.e.,  $\mathbf{y}'$  and  $\mathbf{x}$  are orthogonal. Let  $\mathbf{Y}$  be the matrix of all vectors  $\mathbf{y}$  satisfying  $\mathbf{X}'\mathbf{y} = \mathbf{0}$ . Then for every  $\mathbf{x}$  in  $\mathcal{R}(\mathbf{X})$  and every  $\mathbf{y}$  in  $\mathcal{R}(\mathbf{Y})$  the vectors  $\mathbf{x}$  and  $\mathbf{y}$  are orthogonal and we say that  $\mathcal{R}(\mathbf{X})$  is orthogonal to  $\mathcal{R}(\mathbf{Y})$ . This is sometimes written as  $\mathcal{R}(\mathbf{X}) \perp \mathcal{R}(\mathbf{Y})$ . Furthermore, because columns of  $\mathbf{X}$  (i.e., rows of  $\mathbf{X}'$ ) and columns of  $\mathbf{Y}$  are

orthogonal; they are also LIN (see Exercise 10). Therefore, if these columns have  $n$  elements, the theorem of Section 5c indicates that  $r_Y = n - r_X$ ; i.e.,  $\dim(\mathbf{Y}) = n - r_X$ . Hence

$$\text{nullity}(\mathbf{X}) = n - r_X.$$

## EXERCISES

1. Correct the following sentences by adding the prefix "in" to "\_\_\_ dependent" whenever necessary.

Vectors  $v_1, v_2, \dots, v_n$  are said to be linearly \_\_\_ dependent if  $\sum_{i=1}^n k_i v_i = \mathbf{0}$  is true only for  $k_i = 0$  for  $i = 1, 2, \dots, n$ , except that if one of them is null they are linearly \_\_\_ dependent. When they are \_\_\_ dependent at least two of the  $k$ 's are nonzero and, because they are linearly \_\_\_ dependent, at least one of the vectors can be expressed as a linear combination of the other vectors.

It is possible to have fourteen non-null vectors of order thirteen linearly \_\_\_ dependent but only thirteen of them \_\_\_ dependent; in which case the matrix of those thirteen vectors is nonsingular. And indeed it is impossible to have a set of fourteen linearly \_\_\_ dependent non-null vectors of order thirteen.

Consider five vectors (of order four), two of which are linear combinations of the other three. The five vectors then form a linearly \_\_\_ dependent set. And if the five vectors are used as the columns of a matrix the four rows of that matrix will be linearly \_\_\_ dependent.

In a set of non-null vectors of order  $p$  there cannot be more than  $p$  vectors if they are to be linearly \_\_\_ dependent.

2. For  $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ ,  $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}$ ,  $\mathbf{x}_3 = \begin{bmatrix} -13 \\ -1 \\ 2 \end{bmatrix}$  and  $\mathbf{x}_4 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ , show the following.
- $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$  are linearly dependent, and find a linear relationship among them.
  - $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_4$  are LIN, and find the linear combination of them that equals  $[a \ b \ c]'$ .
3. Give two reasons why the columns of  $\mathbf{I}_4$  are LIN, and show that every vector of order 4 is a linear combination of those vectors. (Note that similar results hold for all identity matrices.)

4. With the aid of permutation matrices, calculate (40) for

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

5. For a square matrix  $A$ , suppose there is an  $x \neq 0$  such that  $Ax = 0$ . Explain why  $A$  is singular.
6. Details of proving  $m \leq k$  in the theorem of Section 6 are shown in the text. Give details of proving  $k \leq m$ .
7. For  $A$  square of order  $n$  and  $\mathbf{1}$  of order  $n$ , suppose that  $x \neq 0$  exists such that  $Ax = 0$  and  $x'\mathbf{1} = 0$ . Prove (a) that  $A + \lambda\mathbf{1}\mathbf{1}'$  is singular for any scalar  $\lambda$  and (b) that  $A + \mathbf{f}\mathbf{f}'$  is also singular for any vector  $\mathbf{f}$ .
8. When  $A$  and  $B$  are idempotent with columns of  $B$  being linear combinations of columns of  $A$ , prove that  $AB = B$ .
9. Prove that  $\mathbf{v}\mathbf{v}' - \mathbf{v}'\mathbf{v}$  is singular.
10. When two vectors are orthogonal, prove that they are also LIN.
11. If  $A$  is singular and has nonzero row sums that are the same for every row, prove (a) that  $A + \lambda\mathbf{1}\mathbf{1}'$  is singular and (b) that  $A + \mathbf{1}\mathbf{f}'$  is also singular for any  $\mathbf{f}$ .
12. If  $A$  is singular and  $A\mathbf{v} = \lambda\mathbf{v}$ , prove that  $A + \lambda\mathbf{v}\mathbf{v}'$  and  $A + \mathbf{v}\mathbf{f}'$  for any  $\mathbf{f}$  are singular.
13. When  $X$  has full column rank and  $T$  has full row rank, prove that  $\mathbf{TX} = \mathbf{0}$  implies that the rows of  $T$  are LIN of the rows of  $X'$ .
14.  $M = \{m_{ij}\}$  for  $i, j = 1, 2, \dots, t$  has the following properties:

(i)  $m_{ij} = t^{-1/2}$  for  $j = 1, 2, \dots, t$ .

(ii)  $\sum_{j=1}^t m_{ij} = 0$  for  $i = 2, 3, \dots, t$ .

(iii)  $\sum_{j=1}^t m_{ij}m_{i'j} = 0$  for  $i \neq i'$  and  $i, i' = 2, 3, \dots, t$ .

(iv) Rows of  $M$  are LIN.

On defining  $\lambda_i = \sum_{k=1}^t m_{ik}^2 \neq 0$ , prove

(a)  $\sum_{i=2}^t m_{ij}^2/\lambda_i = 1 - 1/t$  for  $j = 1, 2, \dots, t$ ;

(b)  $\sum_{i=2}^t m_{ij}m_{i'j}/\lambda_i = -1/t$  for  $j \neq j' = 1, 2, \dots, t$ ;

(c)  $\sum_{i=2}^t \left( \sum_{j=1}^t m_{ij}x_j \right)^2 / \lambda_i = \sum_{j=1}^t x_j^2 - \left( \sum_{j=1}^t x_j \right)^2 / t$ .



# CHAPTER 7

## CANONICAL FORMS

### 1. ELEMENTARY OPERATORS

The three elementary operator matrices (also known as elementary operators and as elementary matrices)  $E_{ij}$ ,  $R_{ii}(\lambda)$  and  $P_{ij}(\lambda)$ , introduced in Section 4.4 play important roles in matrix theory, especially in connection with the rank of a matrix. Each elementary operator is a square matrix derived from an identity matrix. When used in a product with some other matrix it leads to the same manipulation of rows (or columns) of that matrix as is used in simplifying determinants. Furthermore, the product matrix has the same rank as the original matrix. As a result, elementary operators facilitate ascertaining the rank of a matrix.

#### a. Row operations

Premultiplication by elementary operators leads to manipulation of rows. Thus  $E_{ij}A$  interchanges the  $i$ th row and  $j$ th row of  $A$ :

$$E_{12}A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix}.$$

$R_{ii}(\lambda)$  multiplies the  $i$ th row of  $A$  by  $\lambda$ :

$$[R_{22}(4)]A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 8 & 8 & 8 \\ 3 & 3 & 3 \end{bmatrix},$$

and  $P_{ij}(\lambda)A$  adds  $\lambda$  times the  $j$ th row of  $A$  to its  $i$ th row:

$$[P_{12}(\lambda)]A = \begin{bmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1+2\lambda & 1+2\lambda & 1+2\lambda \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}.$$

#### b. Transposes

The following transposes are easily verified:

$$[E_{ij}]' = E_{ij}, \quad [R_{ii}(\lambda)]' = R_{ii}(\lambda) \quad \text{and} \quad [P_{ij}(\lambda)]' = P_{ji}(\lambda).$$

Thus an elementary operator and its transpose are of the same form;  $E$ - and  $R$ -type operators are symmetric, and although a  $P$ -type operator is not symmetric, its transpose is a  $P$ -type operator.

#### c. Column operations

Let  $T$  be any one of the elementary operator matrices. Then, because

$$AT = (T'A) \tag{1}$$

and  $T'$  is an elementary operator,  $T'A$  represents row operations on the rows of  $A'$ , which are columns of  $A$ . Hence the right-hand side of (1) is just elementary operations carried out on columns of  $A$ . Thus postmultiplication by elementary operators performs similar manipulations on the columns of  $A$  as are performed on the rows of  $A$  by premultiplication. With  $E$ - and  $R$ -type operators these operations are precisely the same, because  $E_{ij}$  and  $R_{ii}(\lambda)$  are symmetric. But with  $P$ -type operators

$$AP_{ji}(\lambda) = [P_{ji}'(\lambda)A']' = [P_{ij}(\lambda)A']', \tag{2}$$

so that  $AP_{ji}(\lambda)$  does to columns of  $A$  what  $P_{ij}(\lambda)A$  does to rows of  $A$ .

**Examples.** In

$$A[P_{21}(\lambda)] = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+\lambda & 1 & 1 \\ 2+2\lambda & 2 & 2 \\ 3+3\lambda & 3 & 3 \end{bmatrix},$$

$\lambda \times$  (column 2 of  $A$ ) is added to column 1, which is the same operation as  $[P_{12}(\lambda)]A$  did to rows. In contrast, in

$$A[P_{12}(\lambda)] = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \lambda+1 & 1 \\ 2 & 2\lambda+2 & 2 \\ 3 & 3\lambda+3 & 3 \end{bmatrix},$$

$\lambda \times$  (column 1 of  $A$ ) is added to column 2.

#### d. Inverses

As indicated in Section 4.4, all three of the elementary operator matrices have nonzero determinants. Hence all three have inverses. They are as

follows:

$$E_{ij}^{-1} = E_{ij}, \quad [R_{ii}(\lambda)]^{-1} = R_{ii}(1/\lambda) \quad \text{and} \quad [P_{ij}(\lambda)]^{-1} = P_{ij}(-\lambda).$$

Thus an **E**-type operator is the same as its inverse, and the effect of inverting **R**- and **P**-type operators is simply that of changing the constants, replacing  $\lambda$  by  $1/\lambda$  in the **R**-type and  $\lambda$  by  $-\lambda$  in the **P**-type.

In summary, the underlying form of any elementary operator is unchanged when it is either inverted or transposed.

## 2. RANK AND THE ELEMENTARY OPERATORS

### a. Rank

The rank of a matrix is unaffected when it is multiplied by an elementary operator; i.e., if the matrix **A** is multiplied by an **E**-, **P**- or **R**-type operator the rank of the product is the same as the rank of **A**. We discuss the validity of this statement for each operator in turn. If **A** is premultiplied by an **E**-type operator the product **EA** is **A** with two rows interchanged. Hence **EA** is the same as **A** except with the rows in a different sequence. The number of rows that are linearly independent will be unaltered. Therefore  $r(\mathbf{EA}) = r(\mathbf{A})$ .

In the product of **A** and an **R**-type operator every element of some row (or column) of **A** is multiplied by a constant, and the product of a **P**-type operator with **A** is the same as **A** except it has a multiple of one row (column) added to another. In both cases the independence of rows (columns) is unaffected and the same number will be linearly independent after making the product as before. Thus multiplication of any matrix by an elementary operator does not alter rank.

### b. Products of elementary operators

We now use **E** to represent any elementary operator matrix (be it of the **E**-, **P**- or **R**-type) and **P** and **Q** to represent products of elementary operators; e.g.,  $\mathbf{P} = \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1$  is a product of three of them. Then, because multiplication by an elementary operator does not affect rank we know that for any **A**

$$r(\mathbf{A}) = r(\mathbf{E}_1\mathbf{A}) = r[\mathbf{E}_2(\mathbf{E}_1\mathbf{A})] = r[\mathbf{E}_3(\mathbf{E}_2\mathbf{E}_1\mathbf{A})] = r(\mathbf{PA}).$$

This result is used for finding the rank of any matrix **A**. It is done by using the operators of elementary operators to change **A** until its rank is obvious.

### c. Equivalence

When **A** is multiplied by elementary operator matrices, the product is said to be *equivalent* to **A**. Whenever  $\mathbf{B} = \mathbf{PAQ}$ , for **P** and **Q** each being products of elementary operators, then **B** is equivalent to **A** and we write  $\mathbf{B} \cong \mathbf{A}$ . Since all elementary operator matrices have inverses,  $\mathbf{P}^{-1}$  and  $\mathbf{Q}^{-1}$  exist and so  $\mathbf{A} = \mathbf{P}^{-1}\mathbf{BQ}^{-1}$ ; furthermore, because **P** and **Q** are products of elementary operators whose inverses are themselves also elementary operators,  $\mathbf{P}^{-1}$  and  $\mathbf{Q}^{-1}$  are products of elementary operators too, and so  $\mathbf{A} = \mathbf{P}^{-1}\mathbf{BQ}^{-1}$  implies that **A** is equivalent to **B**, i.e.,  $\mathbf{A} \cong \mathbf{B}$ . Hence equivalence is a reflexive operation. Furthermore,  $r_{\mathbf{A}} = r_{\mathbf{B}}$ .

## 3. CALCULATING THE RANK OF A MATRIX

### a. Some special LIN vectors

A set of vectors such as

$$\mathbf{x}_1 = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ 2 \\ 7 \\ 0 \\ 0 \end{bmatrix}$$

is linearly independent because

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 = \begin{bmatrix} 4a_1 + 2a_2 - a_3 \\ 3a_2 + 2a_3 \\ 7a_3 \\ 0 \\ 0 \end{bmatrix} \quad (3)$$

can equal **0** only if all the  $a$ 's are zero. Furthermore, any vector of order 5 having its only nonzero elements in its first three elements is a linear combination of  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$ . This is so because equating (3) to such a vector, e.g.,

$$\begin{bmatrix} 4a_1 + 2a_2 - a_3 \\ 3a_2 + 2a_3 \\ 7a_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 9 \\ 4 \\ 6 \\ 0 \\ 0 \end{bmatrix},$$

always yields solutions for the  $a$ 's that are not all zero. Hence, for example, in the matrix

$$\begin{bmatrix} 4 & 2 & -1 & 9 & 3 \\ 0 & 3 & 2 & 4 & 7 \\ 0 & 0 & 7 & 6 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4)$$

all columns after the first three are linear combinations of those three, and so the matrix has rank 3; and this rank is, by the nature of the occurrence of the zeros, the number of non-null rows. Matrices of the form (4), with ever-increasing numbers of zeros in their rows (until, maybe, all elements are zero), are said to be matrices of *row-echelon form*.

### b. Calculating rank

Matrices like (4) are the foundation for calculating the rank of any matrix  $A$ . By row operations on  $A$  similar to those used in evaluating determinants, we reduce elements of  $A$  to zero until  $A$  has a form like (4)—all elements below the “diagonal”  $a_{11}, a_{22}, a_{33}, \dots$  being zero. Because these row operations are identical to those represented by elementary operator matrices, the resulting matrix  $B$  is  $B = PA$  for  $P$  being a product of elementary operators. Therefore  $r_A$ , which we seek, equals  $r_B$ , which by the form of  $B$  is immediately obvious. Two examples illustrate the procedure.

#### Example 1

$$A = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix}$$

Carrying out the operations

$$\begin{array}{l} \text{row 2} - 3(\text{row 1}) \\ \text{row 3} - 5(\text{row 1}) \end{array} \text{ gives } A \cong \begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 0 & -14 & -20 & -22 \end{bmatrix} \quad (5)$$

and then on this matrix the operation

$$\text{row 3} - 2(\text{row 2}) \text{ gives } A \cong \begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B. \quad (6)$$

$B$  is of the desired form, its rank is 2 and that is the rank of  $A$ .

#### Example 2

$$A = \begin{bmatrix} 1 & 4 & 6 \\ 4 & -1 & 2 \\ -7 & 6 & 2 \\ 2 & 8 & 13 \end{bmatrix}$$

The operations

$$\begin{array}{l} \text{row 2} - 4(\text{row 1}) \\ \text{row 3} + 7(\text{row 1}) \\ \text{row 4} - 2(\text{row 1}) \end{array} \text{ give } A \cong \begin{bmatrix} 1 & 4 & 6 \\ 0 & -17 & -22 \\ 0 & 34 & 44 \\ 0 & 0 & 1 \end{bmatrix}, \quad (7)$$

and then

$$\text{row 3} + 2(\text{row 2}) \text{ yields } A \cong \begin{bmatrix} 1 & 4 & 6 \\ 0 & -17 & -22 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8)$$

Finally, in order to have all non-null rows preceding null rows,

$$\text{interchange rows 3 and 4 to get } A \cong \begin{bmatrix} 1 & 4 & 6 \\ 0 & -17 & -22 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = B. \quad (9)$$

All elements below the “diagonal” are null; this is the appropriate form, the number of non-null rows is 3 and so  $r_A = r_B = 3$ .

Finding the rank of a matrix by this means is very simple—merely manipulate the rows of the matrix by adding multiples of them to each other to reduce the subdiagonal elements to zero. A formal description follows.

### c. A general procedure

For rectangular and square matrices define the elements  $a_{11}, a_{22}, a_{33}, \dots$  of  $A = \{a_{ij}\}$  as the diagonal. The procedure is to use elementary row operations to reduce to zero all the sub-diagonal elements of the first column, basing the calculations on the first element of the first row. Operations (5) and (7) are examples. Then the elements below the diagonal in the second column are reduced to zeros using the second row, based on its diagonal term. Because the sub-diagonal elements of the first column have already been made zero in the first step, this second step does not affect their values. Operations (6) and (8) are examples of this second step. When all elements below the diagonal are zero, the process is complete. The rank is then the number of non-null rows, as illustrated by (6) and (9).

Sometimes it is necessary to interchange columns to get the desired form. For example, the first two operations of (7) carried out on

$$A = \begin{bmatrix} 1 & 4 & 6 & 0 \\ 4 & -1 & 2 & 0 \\ -7 & 6 & 2 & 1 \end{bmatrix} \text{ give } A \cong \begin{bmatrix} 1 & 4 & 6 & 0 \\ 0 & -17 & -22 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

whereupon interchanging columns 3 and 4 achieves the form that permits calculating rank not just as the number of non-null rows but, more easily, as the number of nonzero elements in the diagonal. The reader is urged to practice these steps on matrices given in the exercises.

#### 4. REDUCTION TO EQUIVALENT CANONICAL FORM

##### a. Row operations

Derivation of  $B$  in each of the preceding examples has come from row operations on  $A$ ; each such operation can be represented by an  $E$  matrix. Then with  $P$  being the product of these  $E$ 's, for example,  $P = E_3 E_2 E_1$ ,

$$B = PA = (PI)A = (E_3 E_2 E_1 I)A.$$

Hence  $P = E_3 E_2 E_1 I$  can be derived by carrying out on  $I$  the same row operations as have been made on  $A$  to derive  $B$ .

**Example 1 (continued).** Perform the operations of (5) on

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \begin{array}{l} \text{row 2} - 3(\text{row 1}) \\ \text{row 3} - 5(\text{row 1}) \end{array} \text{ gives } \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -5 & 0 & 1 \end{bmatrix}.$$

Then operation (6) carried out on this,

$$\text{row 3} - 2(\text{row 2}), \text{ gives } \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} = P. \quad (10)$$

With this we find that

$$PA = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ = B \text{ of (6)}. \quad (11)$$

This is the general procedure for obtaining  $P$ , the product of elementary operator matrices corresponding to a sequence of row operations on a matrix  $A$ :  $P$  is derived by carrying out the same sequence of row operations on an identity matrix of order equal to the number of rows in  $A$ .

**Example 2 (continued).** Operations (7), (8) and (9) carried out on  $I_4$  give

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -4 & 1 & 0 & 0 \\ -2 & 0 & 0 & 1 \\ -1 & 2 & 1 & 0 \end{bmatrix} \text{ and } PA = B \text{ of (9)}. \quad (12)$$

##### b. Column operations

Having obtained  $PA$  we can then reduce to zero the elements above the "diagonal," using column operations. The procedure is exactly as used to reduce to zero the elements below the "diagonal" except the operations are on columns rather than rows. An example makes this clear,

**Example 1 (continued).** On  $PA$  of (11) carry out the operations

$$\begin{array}{l} \text{col 2} - 2(\text{col 1}) \\ \text{col 3} - 4(\text{col 1}) \\ \text{col 4} - 3(\text{col 1}) \end{array} \text{ to get } PA \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -7 & -10 & -11 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (13)$$

Then from this matrix

$$\begin{array}{l} \text{col 3} - (10/7)(\text{col 2}) \\ \text{col 4} - (11/7)(\text{col 2}) \end{array} \text{ gives } PA \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -7 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (14)$$

and finally

$$\text{col 2} \times (1/-7) \text{ gives } PA \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = C. \quad (15)$$

$P$  of (10) was obtained by carrying out row operations on an identity matrix. So also can  $C$  of (15) be expressed as  $PAQ$  and  $Q$  obtained by carrying out on an identity matrix the column operations contained in (13), (14) and (15). The identity matrix in this case has order equal to the number of columns in  $A$ , in this example, 4. Thus, starting with  $I_4$ , the operations in

(13) and (14) yield first

$$\begin{bmatrix} 1 & -2 & -4 & -3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and then } \begin{bmatrix} 1 & -2 & -8/7 & 1/7 \\ 0 & 1 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (16)$$

respectively. Finally, the operation in (15) produces

$$\begin{bmatrix} 1 & 2/7 & -8/7 & 1/7 \\ 0 & -1/7 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{Q}. \quad (17)$$

Then with  $\mathbf{PA}$  of (11) we have, for  $\mathbf{C}$  of (15),

$$\mathbf{PAQ} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix} \\ \times \begin{bmatrix} 1 & 2/7 & -8/7 & 1/7 \\ 0 & -1/7 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{C}. \quad (18)$$

The reader should check that this is so.

### c. The equivalent canonical form

Result (18) is an example of the following theorem.

**Theorem.** Any non-null matrix  $\mathbf{A}$  of rank  $r$  is equivalent to

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{C} \quad (19)$$

where  $\mathbf{I}_r$  is the identity matrix of order  $r$ , and the null submatrices are of appropriate order to make  $\mathbf{C}$  the same order as  $\mathbf{A}$ . For  $\mathbf{A}$  of order  $m \times n$ ,  $\mathbf{P}$  and  $\mathbf{Q}$  are nonsingular matrices of order  $m$  and  $n$ , respectively, being products of elementary operators.

**Proof.** Since  $\mathbf{A}$  is non-null there is at least one element that is nonzero. Suppose it is  $a_{pq}$ . Then by successively multiplying  $\mathbf{A}$  by  $\mathbf{E}$ -type operators (premultiplying and

postmultiplying), the element  $a_{pq}$  can be brought to the leading position. Call the resulting matrix  $\mathbf{B}_1$ . Then  $\mathbf{B}_1 \cong \mathbf{A}$  and  $b_{11} = a_{pq} \neq 0$ . Now premultiply  $\mathbf{B}_1$  by  $\mathbf{P}$ -type operators,  $\mathbf{P}_{i1}(-b_{i1}/b_{11})$  for  $i = 2, 3, \dots, m$ , to reduce to zero all elements (except the first) of the first column  $\mathbf{B}_1$ ; likewise postmultiply  $\mathbf{B}_1$  by  $\mathbf{P}_{1j}(-b_{1j}/b_{11})$  for  $j = 2, 3, \dots, n$ , to reduce the first row (except the first element) to zeros. Call the resulting matrix  $\mathbf{B}_2$ . Then

$$A_{m \times n} \cong \mathbf{B}_1 \cong \mathbf{B}_2 = \begin{bmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & c_{m2} & \cdots & c_{mn} \end{bmatrix}$$

This process can now be repeated operating on the  $c_{ij}$ 's, provided that at least one of them is nonzero; and it can be repeated until, in a form similar to this, there are only zero terms remaining apart from the nonzero terms that have been brought to the diagonal at each step in the process. Furthermore, at each step the resulting matrix is equivalent to  $\mathbf{A}$  and has rank  $r$ , the same as  $\mathbf{A}$ ; and by the nature of the operations (of bringing nonzero elements into the diagonal and making off-diagonal elements zero) we know the final form is diagonal, as far as nonzero elements are concerned; and because the rank is  $r$  there must be exactly  $r$  of these elements. Multiplying each non-null column by the reciprocal of the nonzero term therein yields the final form as

$$A_{m \times n} \cong \mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where  $\mathbf{P}$  and  $\mathbf{Q}$  are the products of the elementary operators. QED

It is clear that the proof of the theorem is largely a formal statement of the procedures used in the example.

The matrix in (19) is called the *equivalent canonical form* or the *canonical form under equivalence*. Its importance is that it *always exists*. This means that for any non-null matrix  $\mathbf{A}$  (of real numbers) the form  $\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$  can always be obtained by carrying out row and column operations on  $\mathbf{A}$ . Furthermore, the row operations can always be represented by a matrix we call  $\mathbf{P}$ , which is a product of elementary operator matrices, and the column operations can always be represented by a matrix  $\mathbf{Q}$ , which is also a product of elementary operator matrices. Also, because  $\mathbf{P}$  and  $\mathbf{Q}$  are products of elementary operator matrices, which themselves are nonsingular and have inverses,  $\mathbf{P}$  and  $\mathbf{Q}$  are also nonsingular and have inverses:  $\mathbf{P}^{-1}$  and  $\mathbf{Q}^{-1}$  exist.

#### d. Non-uniqueness of $\mathbf{P}$ and $\mathbf{Q}$

An important property of  $\mathbf{P}$  and  $\mathbf{Q}$  should be noted: for any  $\mathbf{A}$ , the matrices  $\mathbf{P}$  and  $\mathbf{Q}$  are not unique. For example, column operation (15) could be eliminated and replaced by the row operation (row 2)  $\times$  (1/-7). This would give, from (10) and (16),

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 3/7 & -1/7 & 0 \\ 1 & -2 & 1 \end{bmatrix} \text{ and } \mathbf{Q} = \begin{bmatrix} 1 & -2 & -8/7 & 1/7 \\ 0 & 1 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with

$$\begin{aligned} \mathbf{PAQ} &= \begin{bmatrix} 1 & 0 & 0 \\ 3/7 & -1/7 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix} \\ &\times \begin{bmatrix} 1 & -2 & -8/7 & 1/7 \\ 0 & 1 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

just as in (18), even though the  $\mathbf{P}$  and  $\mathbf{Q}$  here are not the same as  $\mathbf{P}$  and  $\mathbf{Q}$  there.

#### e. Existence is assured

In Section 6.4c the existence of relationships among linearly dependent vectors was emphasized rather than the actual values of those relationships. The same is true here. It is the existence of  $\mathbf{P}$  and  $\mathbf{Q}$  that is much more important than their specific values. It is knowing that they always exist, without needing to know their specific values, that often permits us to establish useful results. Illustration of this is provided by the following theorem.

#### f. Full-rank factorization

The full-rank factoring of a matrix developed in (40) and (43) of Section 6.10 is now stated and proved formally as a theorem.

**Theorem.**  $\mathbf{A}_{p \times q}$  of rank  $r$  can always be factored as

$$\mathbf{A} = \mathbf{K}_{p \times r} \mathbf{L}_{r \times q} \quad (20)$$

where  $\mathbf{K}$  and  $\mathbf{L}$  have full column and full row rank, respectively.

**Proof.** There exist nonsingular matrices  $\mathbf{P}$  and  $\mathbf{Q}$  such that

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ and, equivalently, } \mathbf{A} = \mathbf{P}^{-1} \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1}.$$

Partition  $\mathbf{P}^{-1}$  and  $\mathbf{Q}^{-1}$  as

$$\mathbf{P}^{-1} = [\mathbf{K}_{p \times r} \quad \mathbf{W}_{p \times (p-r)}] \text{ and } \mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{L}_{r \times q} \\ \mathbf{Z}_{(q-r) \times q} \end{bmatrix}.$$

Then

$$\mathbf{A} = [\mathbf{K} \quad \mathbf{W}] \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{L} \\ \mathbf{Z} \end{bmatrix} = [\mathbf{K} \quad \mathbf{0}] \begin{bmatrix} \mathbf{L} \\ \mathbf{Z} \end{bmatrix} = \mathbf{KL} = \mathbf{K}_{p \times r} \mathbf{L}_{r \times q}. \quad (21)$$

Since  $\mathbf{P}$  is nonsingular, so is  $\mathbf{P}^{-1}$ , and so the columns of  $\mathbf{P}^{-1}$  are LIN. In particular, the  $r$  columns of  $\mathbf{K}$  are LIN, and so  $r_{\mathbf{K}} = r$ ; i.e.,  $\mathbf{K}$  has full column rank. (Of course, with  $\mathbf{A}$  being  $p \times q$  of rank  $r$ , it is clear that  $r \leq p$ , the number of rows in  $\mathbf{K}$ .)

Similar reasoning for rows of  $\mathbf{L}$  shows that  $\mathbf{L}$  has full row rank. QED

**Example.** Define

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \\ 4 & 9 & 3 & 24 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -2 & -1 & 1 \end{bmatrix} \quad (22)$$

$$\mathbf{Q} = \begin{bmatrix} 1 & -2 & -3 & 3 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It will be found that the equivalent canonical form of  $\mathbf{A}$  is

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

Therefore

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

is

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and so by (21), with the rank of  $\mathbf{A}$  being 2,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 1 & -1 & 4 \end{bmatrix}. \quad (23)$$

Carrying out the multiplication in (23) confirms its validity.

Note that  $\mathbf{A}$  of (22) is the same as  $\mathbf{A}$  in (41) of Section 6.9; but the factoring in (23) is different from either of the factorings shown in (42) of Section 6.9. There is no mystery in this. Whenever  $\mathbf{A} = \mathbf{KL}$  as in (20), it is also true that  $\mathbf{A} = (\mathbf{KS})(\mathbf{S}^{-1}\mathbf{L})$  for any nonsingular matrix  $\mathbf{S}$  of order  $r_A$ . Indeed, by applying this to (23) it will be found that using

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \quad \text{and then} \quad \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

for  $\mathbf{S}$  yields the two factorings shown in (42) of Section 6.9. (See Exercise 4.)

**Corollary.** When  $\mathbf{A}$  is nonsingular

$$\mathbf{PAQ} = \mathbf{I} \quad \text{and} \quad \mathbf{A} = \mathbf{P}^{-1}\mathbf{Q}^{-1};$$

this shows that every nonsingular matrix is a product of elementary operators.

## 5. RANK OF A PRODUCT MATRIX

One of the most useful results concerning rank involves inequality statements about the rank of a product of two matrices. It is simply stated, as follows.

**Theorem.**  $r_{\mathbf{AB}} \leq$  the lesser of  $r_A$  and  $r_B$ .

**Proof.** For  $r_A = r$  we know that nonsingular  $\mathbf{P}$  and  $\mathbf{Q}$  exist such that

$$\mathbf{PAQ} = \mathbf{C} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and equivalently} \quad \mathbf{PA} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1}.$$

[7.5]

Postmultiply both sides of the second equation by  $\mathbf{B}$ , and for  $\mathbf{B}$  having  $n$  columns partition  $\mathbf{Q}^{-1}\mathbf{B}$  as  $\mathbf{Q}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{S} \end{bmatrix}$  so that

$$\mathbf{PAB} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{0} \end{bmatrix}.$$

Then, because  $\mathbf{P}$  is a product of elementary operators,

$$r_{\mathbf{AB}} = r_{\mathbf{PAB}} = r \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{0} \end{bmatrix} \leq r \quad \text{because } \mathbf{T} \text{ has } r \text{ rows};$$

i.e.,  $r_{\mathbf{AB}} \leq r_A$ , since  $r = r_A$ . Similar reasoning, based on the canonical form of  $\mathbf{B}$ , gives  $r_{\mathbf{AB}} \leq r_B$ . QED

Even though this theorem does not involve an equality statement, it can often be applied twice to yield inequalities of the form  $\text{rank} \leq m$  and  $m \leq \text{rank}$  and hence  $\text{rank} = m$ . The following lemmas illustrate this use. The lemmas themselves find widespread use throughout matrix algebra, and the proofs of many of them illustrate how useful the theorem is.

**Lemma 1.** Multiplication by a nonsingular matrix does not alter rank.

**Proof.** Let  $\mathbf{Q} = \mathbf{TA}$  for  $\mathbf{T}$  nonsingular. Then, by the theorem  $r_{\mathbf{Q}} \leq r_A$ . But  $\mathbf{A} = \mathbf{T}^{-1}\mathbf{Q}$  so that the theorem also gives  $r_A \leq r_{\mathbf{Q}}$ . Hence  $r_{\mathbf{Q}} = r_A$ . QED  
This lemma is used repeatedly where problems of rank are concerned.

**Lemma 2.** For arbitrary  $\mathbf{z}$  the maximum number of LIN vectors  $\mathbf{Bz}$  is  $r_B$ .

**Proof.** Let  $\mathbf{Z}$  be a nonsingular matrix of the same order as  $\mathbf{z}$ . Then, by Lemma 1,  $r_{\mathbf{BZ}} = r_B$  and so the maximum number of LIN vectors  $\mathbf{Bz}$ , i.e., the maximum number of LIN columns in  $\mathbf{BZ}$ , is  $r_B$ . QED

This lemma plays an important role in solving linear equations (see Section 10.4).

**Lemma 3.** If  $\mathbf{AGA} = \mathbf{A}$ , then  $r_{\mathbf{GA}} = r_A$ .

**Proof.** Applying the theorem directly to  $\mathbf{GA}$  gives  $r_{\mathbf{GA}} \leq r_A$ , and applied to  $\mathbf{A} = \mathbf{AGA}$  it gives  $r_A \leq r_{\mathbf{GA}}$ . Therefore  $r_{\mathbf{GA}} = r_A$ . QED

Matrices  $\mathbf{G}$  of this form are called generalized inverses and are the subject of Chapter 8. They are used repeatedly in solving equations, as discussed in Chapter 9.

**Lemma 4.**  $r(\mathbf{A} \mid \mathbf{B}) \leq r(\mathbf{A}) + r(\mathbf{B})$ .

**Proof.**  $r(\mathbf{A} \mid \mathbf{B}) =$  number of LIN columns in  $(\mathbf{A} \mid \mathbf{B})$   
 $\leq$  (number of LIN columns in  $\mathbf{A}$ )  
 $+ (\text{number of LIN columns in } \mathbf{B})$   
 $\leq r(\mathbf{A}) + r(\mathbf{B})$ . QED

This lemma does not involve the theorem, but it is part of the next one, which does.

**Lemma 5.**  $r(\mathbf{A} + \mathbf{B}) \leq r(\mathbf{A} \mid \mathbf{B}) \leq r(\mathbf{A}) + r(\mathbf{B})$ .

**Proof.** Applying the theorem to the identity  $\mathbf{A} + \mathbf{B} \equiv [\mathbf{A} \mid \mathbf{B}] \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix}$  gives  $r(\mathbf{A} + \mathbf{B}) \leq r(\mathbf{A} \mid \mathbf{B})$  which, combined with Lemma 4, establishes Lemma 5. QED

This lemma is used to prove Lemma 6, which in turn is used for Lemma 7, which has important applications in solving linear equations, as seen in Section 9.4b.

**Lemma 6.** For  $\mathbf{A}$  square, of order  $n$ ,  $r(\mathbf{A}\mathbf{B}) \geq r(\mathbf{A}) + r(\mathbf{B}) - n$ .

**Proof.** For nonsingular  $\mathbf{P}$  and  $\mathbf{Q}$  such that

$$\mathbf{P}\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{I}_{r_A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \text{ define } \mathbf{X} = \mathbf{P}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r_A} \end{bmatrix} \mathbf{Q}^{-1}.$$

Then  $\mathbf{A} + \mathbf{X} = \mathbf{P}^{-1}\mathbf{Q}^{-1}$  and so, starting with Lemma 1,

$$\begin{aligned} r(\mathbf{B}) &= r(\mathbf{P}^{-1}\mathbf{Q}^{-1}\mathbf{B}) = r(\mathbf{A}\mathbf{B} + \mathbf{X}\mathbf{B}) \\ &\leq r(\mathbf{A}\mathbf{B}) + r(\mathbf{X}\mathbf{B}) \text{ by Lemma 5} \\ &\leq r(\mathbf{A}\mathbf{B}) + r(\mathbf{X}) \text{ because } r(\mathbf{X}\mathbf{B}) \leq r(\mathbf{X}) \\ &\leq r(\mathbf{A}\mathbf{B}) + n - r_A \text{ by the nature of } \mathbf{X}. \end{aligned} \quad \text{QED}$$

This lemma is known as Sylvester's law of nullity.

**Lemma 7.** For  $\mathbf{M}$  idempotent, of order  $n$ ,  $r(\mathbf{I} - \mathbf{M}) = n - r(\mathbf{M})$ .

**Proof.** Because  $\mathbf{I} = \mathbf{I} - \mathbf{M} + \mathbf{M}$ , Lemma 5 gives  $r(\mathbf{I}) \leq r(\mathbf{I} - \mathbf{M}) + r(\mathbf{M})$ , so that  $n - r(\mathbf{M}) \leq r(\mathbf{I} - \mathbf{M})$ . But because  $\mathbf{0} = (\mathbf{I} - \mathbf{M})\mathbf{M}$ , Lemma 6 gives  $r(\mathbf{0}) \geq r(\mathbf{I} - \mathbf{M}) + r(\mathbf{M}) - n$ , which is  $n - r(\mathbf{M}) \geq r(\mathbf{I} - \mathbf{M})$ . Therefore  $r(\mathbf{I} - \mathbf{M}) = n - r(\mathbf{M})$ . QED

This result is used in Section 9.4b. A simpler proof is available using eigenvalues, a topic discussed in Chapter 11 (see Exercise 11.18b).

**Lemma 8.** A matrix has left inverses only if it has full column rank.

**Proof.** Let  $\mathbf{L}$  be a left inverse  $\mathbf{A}_{r \times c}$ . Then  $\mathbf{L}\mathbf{A} = \mathbf{I}$  or, equivalently  $\mathbf{I}_{c \times c} = \mathbf{L}_{c \times r} \mathbf{A}_{r \times c}$ , so that the theorem gives  $r_1 = c \leq r_A$ . But, by definition of rank,  $r_A = r(\mathbf{A}_{r \times c}) \leq c$  and therefore  $r_A = c$ , i.e.,  $\mathbf{A}$  has full column rank. QED

**Converse.** A matrix has right inverses only if it has full row rank.

**Corollary.** A matrix has both a left inverse and a right inverse only if it is nonsingular; and each is the regular inverse.

These results on left and right inverses are consonant with the theorem in Section 5.10 on the existence of such inverses. One form of the inverses is given in Section 7e.

## 6. SYMMETRIC MATRICES

The symmetry inherent in symmetric matrices provides them with numerous special properties, some of which arise in canonical forms.

### a. Row and column operations

When  $\mathbf{E}$  is an elementary operator,  $\mathbf{E}\mathbf{A}$  represents row operations on  $\mathbf{A}$  and  $\mathbf{A}\mathbf{E}$  represents column operations on  $\mathbf{A}$ . But, as indicated in Section 1c, the column operations represented by  $\mathbf{A}\mathbf{E}$  are not always the same as the row operations represented by  $\mathbf{E}\mathbf{A}$ . In fact, from (1), since  $\mathbf{A}\mathbf{E}' = (\mathbf{E}\mathbf{A})'$ , the column operations represented by  $\mathbf{A}\mathbf{E}'$  are the same as the row operations on  $\mathbf{A}'$  represented by  $\mathbf{E}\mathbf{A}'$ . Thus column operations on  $\mathbf{A}$  represented by  $\mathbf{A}\mathbf{E}'$  are the same kind of operations as are row operations on  $\mathbf{A}$  represented by  $\mathbf{E}\mathbf{A}$ . For example, with

$$\mathbf{A} = \begin{bmatrix} r'_1 \\ r'_2 \end{bmatrix} = [\mathbf{c}_1 \quad \mathbf{c}_2] \text{ and } \mathbf{E} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix},$$

we have  $\mathbf{A}\mathbf{E}' = [\mathbf{c}_1 \quad \mathbf{c}_2] \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} = [\mathbf{c}_1 \quad 3\mathbf{c}_1 + \mathbf{c}_2]$  and

$$\mathbf{E}\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} r'_1 \\ r'_2 \end{bmatrix} = \begin{bmatrix} r'_1 \\ 3r'_1 + r'_2 \end{bmatrix}.$$

It is evident that  $\mathbf{A}\mathbf{E}'$  does to the columns of  $\mathbf{A}$  what  $\mathbf{E}\mathbf{A}$  does to the rows of  $\mathbf{A}$  (but  $\mathbf{A}\mathbf{E}$  does not). This is true generally, for any square matrix, and is



particularly pertinent for symmetric matrices, as will be seen. The same situation holds for  $\mathbf{P}$ , a product of elementary operators, used in place of  $\mathbf{E}$ .

### b. The diagonal form

The preceding result is of paramount importance in deriving the diagonal form of a symmetric matrix, the first step toward its equivalent canonical form. The derivation involves carrying out on  $\mathbf{A} = \mathbf{A}'$  the same operations on columns as are performed on rows. The operations can be represented as  $\mathbf{PAP}'$  and the result is a diagonal matrix.

**Example.** Consider

$$\mathbf{A} = \begin{bmatrix} 4 & 12 & 20 \\ 12 & 45 & 78 \\ 20 & 78 & 136 \end{bmatrix} \quad (24)$$

and carry out on  $\mathbf{A}$ , and on  $\mathbf{I}$ , the row operations that reduce to zero the elements below the diagonal of  $\mathbf{A}$ . This gives us  $\mathbf{PA}$  and  $\mathbf{P}$  as follows:

$$\begin{array}{l} \text{row 2} - 3(\text{row 1}) \\ \text{row 3} - 5(\text{row 1}) \end{array} : \mathbf{A} \cong \begin{bmatrix} 4 & 12 & 20 \\ 0 & 9 & 18 \\ 0 & 18 & 36 \end{bmatrix} \quad \text{and} \quad \mathbf{I} \cong \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -5 & 0 & 1 \end{bmatrix}$$

$$\text{row 3} - 2(\text{row 2}) : \mathbf{A} \cong \begin{bmatrix} 4 & 12 & 20 \\ 0 & 9 & 18 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{PA} \quad \text{and} \quad \mathbf{I} \cong \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} = \mathbf{P}$$

Because  $\mathbf{A}$  is symmetric the same operations made on columns of  $\mathbf{A}$  as have been made on rows of  $\mathbf{A}$  in getting  $\mathbf{P}$  will reduce to zero all the elements above the diagonal. Also, by the property of elementary operators illustrated and stated in the preceding subsection, these column operations can be achieved by post-multiplication by  $\mathbf{P}'$ . Hence in place of  $\mathbf{PAQ}$  for any matrix we have  $\mathbf{PAP}'$  for symmetric matrices. For the example

$$\mathbf{PAP}' = \begin{bmatrix} 4 & 12 & 20 \\ 0 & 9 & 18 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -3 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (25)$$

Result (25) is an example of the general result: for any symmetric matrix  $\mathbf{A}$  there always exists a nonsingular  $\mathbf{P}$ , a product of elementary operators, such that

$$\mathbf{PAP}' = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{for } r = r_{\mathbf{A}} \quad (26)$$

with  $\mathbf{D}_r$  being a diagonal matrix of  $r$  nonzero diagonal elements. This form, (26), is called the *diagonal form* of a symmetric matrix. In contrast to the equivalent canonical form for any matrix, given in (19),  $\mathbf{D}_r$  of (26) is not necessarily an identity matrix.

### c. The canonical form under congruence

We now show how (26) can be further reduced to the canonical form  $\mathbf{C}$  of (19). As distinct from the equivalent canonical form for any matrix, it is called the *congruent canonical form* (or the *canonical form under congruence*) of a symmetric matrix.

**Example** (continued). Consider pre-multiplying and post-multiplying (25) by a diagonal matrix whose diagonal elements are reciprocals of the square roots of the nonzero elements in  $\mathbf{PAP}'$  and unity otherwise:

$$\begin{bmatrix} 1/\sqrt{4} & 0 & 0 \\ 0 & 1/\sqrt{9} & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{PAP}' \begin{bmatrix} 1/\sqrt{4} & 0 & 0 \\ 0 & 1/\sqrt{9} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

On writing

$$\mathbf{R} = \begin{bmatrix} 1/\sqrt{4} & 0 & 0 \\ 0 & 1/\sqrt{9} & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{P} \quad (27)$$

we thus have

$$\mathbf{RAR}' = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (28)$$

This is the canonical form under congruence.

Results (27) and (28) typify the general case, wherein, from the general diagonal form of (26), the development of  $\mathbf{R}$  similar to that in (27) is

$$\mathbf{R} = \begin{bmatrix} \sqrt{\mathbf{D}_r^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{P} \quad (29)$$

where, because  $\mathbf{D}_r$  is diagonal with no element  $d_{ii}$  being zero,  $\sqrt{\mathbf{D}_r^{-1}}$  is diagonal with elements  $1/\sqrt{d_{ii}}$ . This is evident in (27). Then, similar to (28),

$$\mathbf{R}\mathbf{A}\mathbf{R}' = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{for } r = r_A \quad (30)$$

is the canonical form under congruence.

#### d. Two special provisions

The procedures just described differ from those of deriving the equivalent canonical form for any matrix, because of the desire to retain the symmetry inherent in a symmetric matrix. Achieving this requires two special provisions.

(i) *Null rows.* The procedure for interchanging rows to put non-null rows before null rows, as illustrated between (8) and (9) for nonsymmetric matrices, is the same for symmetric matrices except that symmetry must be retained.

**Example.** On

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 3 & 10 \\ 4 & 8 & 6 & 20 \\ 3 & 6 & 22\frac{1}{2} & 21 \\ 10 & 20 & 21 & 53 \end{bmatrix} \quad \text{and} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

the operations row 2 - 2(row 1), row 3 - 1½(row 1) and row 4 - 5(row 1) give

$$\mathbf{A} \cong \begin{bmatrix} 2 & 4 & 3 & 10 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 18 & 6 \\ 0 & 0 & 6 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{I} \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -1\frac{1}{2} & 0 & 1 & 0 \\ -5 & 0 & 0 & 1 \end{bmatrix}$$

Now, in bringing the null row to be the fourth row, the symmetry of the  $\begin{bmatrix} 18 & 6 \\ 6 & 3 \end{bmatrix}$  submatrix must be retained. This is achieved by interchanging rows 2 and 3, and then rows 3 and 4; i.e., "filtering" the null row to the bottom. The result is

$$\mathbf{A} \cong \begin{bmatrix} 2 & 4 & 3 & 10 \\ 0 & 0 & 18 & 6 \\ 0 & 0 & 6 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{I} \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1\frac{1}{2} & 0 & 1 & 0 \\ -5 & 0 & 0 & 1 \\ -2 & 1 & 0 & 0 \end{bmatrix}$$

Finally, row 3 - ½(row 2) gives

$$\mathbf{A} \cong \begin{bmatrix} 2 & 4 & 3 & 10 \\ 0 & 0 & 18 & 6 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{P}\mathbf{A} \quad \text{and} \quad \mathbf{I} \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1\frac{1}{2} & 0 & 1 & 0 \\ -4\frac{1}{2} & 0 & -\frac{1}{3} & 1 \\ -2 & 1 & 0 & 0 \end{bmatrix} = \mathbf{P}$$

and then the diagonal form is

$$\mathbf{P}\mathbf{A}\mathbf{P}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1\frac{1}{2} & 0 & 1 & 0 \\ -4\frac{1}{2} & 0 & -\frac{1}{3} & 1 \\ -2 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 3 & 10 \\ 4 & 8 & 6 & 20 \\ 3 & 6 & 22\frac{1}{2} & 21 \\ 10 & 20 & 21 & 53 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & -1\frac{1}{2} & -4\frac{1}{2} & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & -\frac{1}{3} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 18 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The canonical form for

$$\mathbf{R} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 1/\sqrt{18} & 0 & 0 \\ 0 & 0 & 1/\sqrt{1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{P} \quad \text{is then} \quad \mathbf{R}\mathbf{A}\mathbf{R}' = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

(ii) *Negative elements in the diagonal form.* For some symmetric matrices,  $\mathbf{D}_r$  of the diagonal form (26) will contain negative elements. In that case some elements  $1/\sqrt{d_{ii}}$  of  $\sqrt{\mathbf{D}_r^{-1}}$  will have the imaginary value  $\sqrt{-1}$  as a factor, as will the corresponding rows of  $\mathbf{R}$  in (29); i.e.,  $\mathbf{R}$  will not be a real matrix. Nevertheless, the canonical form under congruence, (30), still exists: it always exists but in some cases  $\mathbf{R}$  is not real.

An important class of matrices for which  $\mathbf{R}$  is always real is the class of non-negative definite (n.n.d.) matrices. For them,  $\mathbf{D}_r$  of (26) has only positive diagonal elements (see Section 7d) and so  $\mathbf{R}$  is real.

The occurrence of  $\sqrt{-1}$  in  $\mathbf{R}$  can be avoided by the following procedure, which yields a slight variant of the congruent canonical form.

Example. For

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & -2 & -1 & -4 \\ 2 & -1 & 3 & 0 \\ 2 & -4 & 0 & -6 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

$$\mathbf{PAP}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Premultiply  $\mathbf{PAP}'$  by a permutation matrix,  $\mathbf{S}$  say (and postmultiply by  $\mathbf{S}'$ ), to bring all the positive elements ahead of the negative ones in the diagonal. Then

$$\mathbf{SPAP}'\mathbf{S}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (31)$$

Now define  $\sqrt{\text{mod}(\mathbf{D}_r^{-1})}$  as the diagonal matrix of elements  $1/\sqrt{\text{mod}(d_{ii})}$  where  $\text{mod}(d_{ii})$  is  $d_{ii}$  ignoring sign. Then, similar to (29) and (30), define

$$\mathbf{R} = \begin{bmatrix} \sqrt{\text{mod}(\mathbf{D}_r^{-1})} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{SP} \quad \text{and get} \quad \mathbf{RAR}' = \begin{bmatrix} \mathbf{I}_{r-q} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where  $q$  is the number of negative elements in  $\mathbf{D}_r$ . The number of positive elements is  $r - q$ , and the difference between this and  $q$ , namely  $r - 2q$ , is called the *signature* of  $\mathbf{A}$ . The triplet of values  $[r - q, q, n - r]$  is sometimes (e.g., Oullette, 1981) called the *inertia*.

In the example,

$$\mathbf{RAR}' = \begin{bmatrix} 1/\sqrt{1} & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} 1/\sqrt{1} & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

the signature is  $3 - 2 = 1$ , and the inertia is  $(2, 1, 1)$ .

### e. Full-rank factorization

The factorization  $\mathbf{A} = \mathbf{KL}$  for  $\mathbf{K}$  and  $\mathbf{L}$  of full column and full row rank, respectively, is derived in Section 4f from the equivalent canonical form

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \text{i.e.,} \quad \mathbf{A} = \mathbf{P}^{-1} \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1},$$

of Section 4c. This is for any matrix  $\mathbf{A}$ . For symmetric  $\mathbf{A}$ , the analogous development comes from (30):

$$\mathbf{RAR}' = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \text{i.e.,} \quad \mathbf{A} = \mathbf{R}^{-1} \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{R}^{-1'} \quad (32)$$

Partitioning  $\mathbf{R}^{-1}$  as  $[\mathbf{K} \quad \mathbf{W}]$ , for  $\mathbf{K}$  having  $r_A$  LIN columns gives

$$\mathbf{A} = \mathbf{A}' = \mathbf{KK}', \quad \text{for } \mathbf{K} \text{ of full column rank.} \quad (33)$$

As with (30), so with (33): it always exists, but in some cases  $\mathbf{K}$  is not always real. It is real for n.n.d. matrices [see (35)].

## 7. NON-NEGATIVE DEFINITE MATRICES

Non-negative definite (n.n.d.) matrices are defined in Section 3.6. Essentially they are symmetric matrices  $\mathbf{A}$  such that  $\mathbf{x}'\mathbf{Ax} \geq 0$  for all  $\mathbf{x}$ . If  $\mathbf{x}'\mathbf{Ax} = 0$  only for  $\mathbf{x} = \mathbf{0}$ , then  $\mathbf{A}$  is positive definite (p.d.), but if  $\mathbf{x}'\mathbf{Ax} = 0$  for  $\mathbf{x} = \mathbf{0}$  and for some  $\mathbf{x} \neq \mathbf{0}$ , then  $\mathbf{A}$  is positive semidefinite (p.s.d.). We continue to limit discussion to real matrices and vectors, for which  $\mathbf{x}'\mathbf{x} > 0$  for  $\mathbf{x} \neq \mathbf{0}$ . Under these conditions n.n.d. matrices have a variety of special properties, some of which are especially helpful in applications to statistics.

### a. Diagonal elements and principal minors

An important property of a p.(s).d. matrix is that all its leading principal minors are positive (positive or zero). Proofs of this result, which for p.d. matrices is both a necessary and sufficient condition, are in many texts, but since most are lengthy none is given here. An elegant inductive proof in Seelye (1958) is commended to the interested reader (see Exercise 10.19).

Non-negative definiteness of  $\mathbf{A}$  is a property that is not affected by changing the sequence of rows and columns in  $\mathbf{A}$ , because that merely amounts to relabeling the elements of  $\mathbf{x}$  in  $\mathbf{x}'\mathbf{Ax}$ . Therefore the necessity part of this theorem applies to every principal minor of  $\mathbf{A}$ , and particularly to all diagonal elements of  $\mathbf{A}$  (which are principal minors of order 1) and to  $|\mathbf{A}|$

(which is the principal minor of largest order). Hence

$$\begin{aligned} \text{if } \mathbf{A} \text{ is p.d., } & a_{ii} > 0 \text{ and } |\mathbf{A}| > 0; \\ \text{if } \mathbf{A} \text{ is p.s.d., } & a_{ii} \geq 0 \text{ and } |\mathbf{A}| = 0. \end{aligned} \quad (34)$$

### b. Congruent canonical form

The diagonal form of an n.n.d. matrix never has negative elements, and so reduction to congruent canonical form involves only real numbers. We establish this result as a corollary to the second of two lemmas.

**Lemma.** When  $\mathbf{A}$  is p.(s.)d. so is  $\mathbf{PAP}'$  for nonsingular  $\mathbf{P}$ .

**Proof.**  $\mathbf{x}'\mathbf{Ax} = \mathbf{y}'\mathbf{PAP}'\mathbf{y}$  for  $\mathbf{y} = \mathbf{P}'^{-1}\mathbf{x}$ . Therefore, for any  $\mathbf{x} \neq \mathbf{0}$  for which  $\mathbf{x}'\mathbf{Ax} = 0$ , so will  $\mathbf{y}'\mathbf{PAP}'\mathbf{y}$  be zero for  $\mathbf{y} \neq \mathbf{0}$ . QED

**Lemma.** Diagonal elements in the diagonal matrix  $\mathbf{D}_r$  of the form  $\mathbf{PAP}'$  in (26) are non-negative for an n.n.d. matrix  $\mathbf{A}$ .

**Proof.**  $\mathbf{PAP}'$  is n.n.d. and so by (34) its diagonal elements are non-negative. QED

**Corollary.** In deriving the congruent canonical form (30) of an n.n.d. matrix,  $\mathbf{R}$  is real.

### c. Full-rank factorization

Because n.n.d. matrices are symmetric, they have the full-rank factorization  $\mathbf{A} = \mathbf{KK}'$  of (33). Since, by the preceding corollary,  $\mathbf{R}$  of the congruent form  $\mathbf{RAR}'$  is always real for n.n.d. matrices,  $\mathbf{K}$  is always real for them too. Hence when

$$\mathbf{A} \text{ is n.n.d., } \mathbf{A} = \mathbf{KK}' \text{ for } \mathbf{K} \text{ real and of full column rank.} \quad (35)$$

Furthermore, when  $\mathbf{A}$  is p.d.,  $\mathbf{K}$  is nonsingular.

These results have important applications in statistics where sums of squares are written as quadratic forms in n.n.d. and p.d. matrices.

### d. Quadratic forms as sums of squares

Quadratic forms occurring in statistics are almost always n.n.d. In that case (35) applies, and so

$$\begin{aligned} \mathbf{x}'\mathbf{Ax} &= \mathbf{x}'\mathbf{KK}'\mathbf{x} \\ &= \sum_{t=1}^r z_t^2 \text{ with } \mathbf{z} = \mathbf{K}'\mathbf{x} = \{z_t\}, \text{ for } t = 1, \dots, r \\ &= \sum_{i=1}^r y_i^2 \text{ with } \mathbf{y} = \mathbf{R}^{-1}\mathbf{x} = \{y_i\}, \text{ for } i = 1, \dots, n, \text{ the order of } \mathbf{A}, \end{aligned} \quad (36)$$

because of the partitioning  $\mathbf{R}^{-1} = [\mathbf{K} \quad \mathbf{W}]$  shown below (32).

**Examples.** It is easily seen that the quadratic form

$$\begin{aligned} q &= 4x_1^2 + 45x_2^2 + 136x_3^2 + 24x_1x_2 + 40x_1x_3 + 156x_2x_3 \\ &= \mathbf{x}'\mathbf{Ax} \text{ for } \mathbf{A} = \begin{bmatrix} 4 & 12 & 20 \\ 12 & 45 & 78 \\ 20 & 78 & 136 \end{bmatrix}, \end{aligned}$$

the same matrix  $\mathbf{A}$  as in (24). The congruent canonical form of  $\mathbf{A}$  is given in (28) with  $\mathbf{R}$  being, from (27),

$$\mathbf{R} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -1 & \frac{1}{3} & 0 \\ 1 & -2 & 1 \end{bmatrix}.$$

Hence from (28) and (36)

$$q = \mathbf{y}' \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{y} = y_1^2 + y_2^2$$

with

$$\begin{aligned} \mathbf{y} = \mathbf{R}'^{-1}\mathbf{x} &= \begin{bmatrix} \frac{1}{2} & -1 & 1 \\ 0 & \frac{1}{3} & -2 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \mathbf{x} = \begin{bmatrix} 2 & 6 & 10 \\ 0 & 3 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 + 6x_2 + 10x_3 \\ 3x_2 + 6x_3 \\ x_3 \end{bmatrix}. \end{aligned}$$

Thus, as can readily be verified by expansion,

$$q = y_1^2 + y_2^2 = (2x_1 + 6x_2 + 10x_3)^2 + (3x_2 + 6x_3)^2.$$

Note in passing that  $q$  is p.s.d. because it can be zero for values of the  $x_i$ 's other than all zero; for example,  $q = 0$  with  $x_1 = 1$ ,  $x_2 = -2$  and  $x_3 = 1$ . In contrast, for

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}, & \mathbf{PAP}' &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1\frac{1}{2} \end{bmatrix}, \\ \mathbf{P} &= \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} & \text{and } \mathbf{P}'^{-1} &= \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned}$$

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{P}^{-1}\mathbf{P}\mathbf{A}\mathbf{P}'\mathbf{P}^{-1}\mathbf{x} = \mathbf{x}'\mathbf{P}^{-1}(\mathbf{P}\mathbf{A}\mathbf{P}')\mathbf{P}^{-1}\mathbf{x}$$

$$= \mathbf{y}' \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1\frac{1}{2} \end{bmatrix} \mathbf{y} \quad \text{for } \mathbf{y} = \mathbf{P}^{-1}\mathbf{x} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 2(x_1 + x_2 + \frac{1}{2}x_3)^2 + 3x_2^2 + 1\frac{1}{2}x_3^2,$$

which is clearly positive for all  $\mathbf{x} \neq \mathbf{0}$ ; i.e.,  $\mathbf{A}$  is p.d.

When  $\mathbf{A}$  is not n.n.d., its diagonal form may contain negative elements, in which case the form (32) is used, and the quadratic  $\mathbf{x}'\mathbf{A}\mathbf{x}$  becomes the difference between two sums of squares:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{r-q} y_i^2 - \sum_{i=r-q+1}^r y_i^2.$$

#### e. Full row (column) rank matrices

In the factoring  $\mathbf{A} = \mathbf{K}\mathbf{K}'$  of (35), the matrix  $\mathbf{K}$  has full column rank. As a result,  $\mathbf{K}'\mathbf{K}$  is p.d. A more general result is given by the following theorem.

**Theorem.** For real  $\mathbf{X}$ , the product  $\mathbf{X}'\mathbf{X}$  is n.n.d.; it is p.d. when  $\mathbf{X}$  has full column rank and p.s.d. otherwise.

**Proof.**  $\mathbf{t}'\mathbf{X}'\mathbf{X}\mathbf{t} = (\mathbf{X}\mathbf{t})'\mathbf{X}\mathbf{t} \geq 0$  because it is the sum of squares of elements of  $\mathbf{X}\mathbf{t}$ . Hence  $\mathbf{X}'\mathbf{X}$  is n.n.d. When  $\mathbf{X}$  has full column rank its columns are LIN and so  $\mathbf{X}\mathbf{t} = \mathbf{0}$  only for  $\mathbf{t} = \mathbf{0}$  and so  $\mathbf{t}'\mathbf{X}'\mathbf{X}\mathbf{t} > 0$  for all  $\mathbf{t} \neq \mathbf{0}$ ; i.e.,  $\mathbf{X}'\mathbf{X}$  is p.d. And when  $\mathbf{X}$  has less than full column rank its columns are linearly dependent and so  $\mathbf{X}\mathbf{t} = \mathbf{0}$  for some  $\mathbf{t} \neq \mathbf{0}$ , for which  $\mathbf{t}'\mathbf{X}'\mathbf{X}\mathbf{t} = 0$ ; and so  $\mathbf{X}'\mathbf{X}$  is p.s.d. QED

**Corollary 1.** For real  $\mathbf{Y}$ , the product  $\mathbf{Y}\mathbf{Y}'$  is n.n.d.; it is p.d. when  $\mathbf{Y}$  has full row rank and p.s.d. otherwise.

**Corollary 2.** For real matrices, a left inverse of  $\mathbf{A}$  when it has full column rank is  $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ , and a right inverse when it has full row rank is  $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ .

#### EXERCISES

1. For  $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ 9 & 8 & 6 \end{bmatrix}$ ,  $\mathbf{E} = \mathbf{E}_{12}$ ,  $\mathbf{R} = \mathbf{R}_{33}(2)$  and  $\mathbf{P} = \mathbf{P}_{12}(3)$

show that

(a)  $\mathbf{EA} = \begin{bmatrix} 4 & 5 & 7 \\ 1 & 2 & 3 \\ 9 & 8 & 6 \end{bmatrix}$ ,  $\mathbf{AE} = \begin{bmatrix} 2 & 1 & 3 \\ 5 & 4 & 7 \\ 8 & 9 & 6 \end{bmatrix}$ ,  $\mathbf{EAE} = \begin{bmatrix} 5 & 4 & 7 \\ 2 & 1 & 3 \\ 8 & 9 & 6 \end{bmatrix}$ ;

(b)  $\mathbf{RA} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ 18 & 16 & 12 \end{bmatrix}$ ,  $\mathbf{AR} = \begin{bmatrix} 1 & 2 & 6 \\ 4 & 5 & 14 \\ 9 & 8 & 12 \end{bmatrix}$ ; find  $\mathbf{RAR}'$ .

(c)  $\mathbf{PA} = \begin{bmatrix} 13 & 17 & 24 \\ 4 & 5 & 7 \\ 9 & 8 & 6 \end{bmatrix}$ ,  $\mathbf{AP} = \begin{bmatrix} 1 & 5 & 3 \\ 4 & 17 & 7 \\ 9 & 35 & 6 \end{bmatrix}$ ; find  $\mathbf{PAP}'$ .

(d)  $\mathbf{REA} = \begin{bmatrix} 4 & 5 & 7 \\ 1 & 2 & 3 \\ 18 & 16 & 12 \end{bmatrix}$ ,  $\mathbf{REAP} = \begin{bmatrix} 4 & 17 & 7 \\ 1 & 5 & 3 \\ 18 & 70 & 12 \end{bmatrix}$ ;

(e)  $-|\mathbf{E}| = 1 = |\mathbf{P}| = \frac{1}{2}|\mathbf{R}|$ ;

(f)  $\mathbf{E} = \mathbf{E}^{-1} = \mathbf{E}'$ ; and  $\mathbf{R} = \mathbf{R}'$ ;

(g)  $\mathbf{R}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$  and  $\mathbf{P}^{-1} = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ .

#### 2. Find the rank of

$$\mathbf{A} = \begin{bmatrix} 1 & 6 \\ 2 & 9 \\ 4 & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 6 & 4 & -1 & 2 & 5 \\ 3 & 0 & -1 & 2 & 7 \\ 18 & 3 & 4 & -2 & 0 \\ 6 & 8 & 0 & 0 & -4 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 3 & 7 & 6 \\ 2 & 1 & 7 \\ 4 & 6 & 3 \\ -1 & -1 & 0 \\ 6 & 8 & 3 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 1 & 0 & 2 \\ -1 & 3 & 6 & 0 \\ 1 & 5 & 6 & 4 \\ 6 & 4 & -3 & 11 \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} 1 & -1 & 0 \\ 3 & 2 & -4 \\ 5 & 0 & -4 \\ 1 & -6 & 4 \end{bmatrix}, \quad \text{and} \quad \mathbf{F} = \begin{bmatrix} 21 & 6 & 3 \\ 12 & 16 & 36 \\ -63 & 13 & 18 \\ 0 & 93 & 81 \end{bmatrix}.$$

#### 3. Illustrate the theorem of Section 4f for matrices $\mathbf{A}$ , $\mathbf{D}$ and $\mathbf{E}$ of Exercise 2.

#### 4. For $\mathbf{A} = \mathbf{K}\mathbf{L}$ of (23), with $\mathbf{S}$ taken first as $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ and then as $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ show that $\mathbf{K}\mathbf{S}(\mathbf{S}^{-1}\mathbf{L})$ yields the factorings given in (42) of Section 6.9.

5. From each factoring derived in Exercise 3, develop a different factoring, in the manner of Exercise 4.

6. For  $A = \begin{bmatrix} 1 & 2 & 4 & 0 \\ -2 & -3 & -1 & 1 \\ 0 & 1 & 7 & 1 \\ -2 & -2 & 6 & 2 \\ -3 & -6 & -12 & 0 \end{bmatrix}$ ,

show the following:

- All minors of  $A$  of order 3 or more are zero.
  - The rank of  $A$  is 2.
  - For any pair of linearly independent rows, each of the other three rows is a linear function thereof.
  - $A$  can be expressed as  $XY$ , where  $X$  is  $5 \times 2$  and  $Y$  is  $2 \times 4$ .
7. Reduce the following matrices to diagonal form and thence to equivalent canonical form.

$$A_1 = \begin{bmatrix} 7 & 13 \\ 2 & 9 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 3 & 6 & 48 \\ 1 & 9 & 2 \\ 4 & 1 & 3 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 1 & 0 & -1 \\ 3 & -4 & 2 \\ 5 & -4 & 0 \\ 1 & 4 & -6 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 3 & 6 & 2 & 4 \\ 9 & 1 & 3 & 2 \\ 6 & -5 & 1 & -2 \end{bmatrix}$$

8. Reduce the following symmetric matrices to diagonal form and thence to canonical form under congruence.

$$B_1 = \begin{bmatrix} 4 & 4 & 10 \\ 4 & 20 & 18 \\ 10 & 18 & 29 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 4 & 6 & 12 \\ 6 & 8 & 1 \\ 12 & 1 & 5 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix}, \quad B_4 = \begin{bmatrix} 1 & 8 & 6 & 7 \\ 8 & 65 & 99 & 40 \\ 6 & 99 & 81 & 78 \\ 7 & 40 & 78 & 21 \end{bmatrix}$$

9. (a) Express

$$q = 3x_1^2 + 16x_2^2 + 139x_3^2 + 12x_1x_2 + 30x_1x_3 + 92x_2x_3$$

as a quadratic form  $x'Ax$  with  $A = A'$ . Find the congruent reduction of  $A$  and from it express  $q$  as a sum of squares of linear combinations of  $x_1$ ,  $x_2$  and  $x_3$ .

- (b) Repeat (a) for

$$q = 4x_1^2 + 21x_2^2 + 116x_3^2 + 16x_1x_2 + 24x_1x_3 + 88x_2x_3.$$

- Prove the corollary at the end of Section 7b.
- Prove the corollaries at the end of Section 7e.
- Prove, for real  $K$ , that  $I + KK'$  is p.d.
- Prove, for real  $X$ , that  $r(X'X) = r(X)$ .  
[Hints: Use the theorem of Section 5, equation (20) for  $X$  and Section 7e.]
- Using (20), suppose  $M = L'(K'AL')^{-1}K'$ . Show that  $AMA = A$  and  $MAM = M$ , and that  $AM$  is symmetric and  $MA$  is symmetric.  $M$  is called the Moore–Penrose inverse of  $A$ . It is discussed in Chapter 8.
- For  $A = XY'$  with  $X$  and  $Y$  of full column rank  $r_A$ , prove that  $A$  and  $A^2$  have the same rank if and only if  $Y'X$  is nonsingular.
- For  $K$  of full column rank, prove that  $T$  and  $KT$  have the same rank.
- Suppose  $A$  is a basis for a vector space  $S$ . Explain why  $A$  has full column rank. Prove that all bases of  $S$  contain the same number of vectors.
- Prove that the difference between two nonsingular matrices has the same rank as the difference between their inverses.
- When  $AB$  is nonsingular, prove that  $A$  has full row rank, and  $B$  has full column rank.

## CHAPTER 8

# GENERALIZED INVERSES

Recent years have seen development for rectangular and singular matrices of an analogue of the inverse of nonsingular matrices. The resulting matrices, known generically as generalized inverses, play an important role in understanding the solutions to linear equations  $Ax = y$  when  $A$  has no inverse. With this in mind, because such equations arise in many important statistical applications (e.g., those of chapters 14 and 15), we give here a brief account of generalized inverses and of some of their properties, especially as they pertain to solving linear equations. The reader wishing to pursue the topic in more detail is referred to books by Albert (1972), Ben-Israel and Greville (1974), Nashed (1976), Pringle and Rayner (1971) and Rao and Mitra (1971).

### 1. THE MOORE-PENROSE INVERSE

Given any matrix  $A$ , there is a unique matrix  $M$  such that

$$\begin{aligned} \text{(i) } AMA &= A & \text{(iii) } AM &\text{ is symmetric} \\ \text{(ii) } MAM &= M & \text{(iv) } MA &\text{ is symmetric.} \end{aligned} \quad (1)$$

This result is developed in Penrose (1955), on foundations laid by Moore (1920). The Penrose paper established not only the existence of  $M$  but also its uniqueness for a given  $A$ . One way of writing  $M$  is based on the factoring of  $A_{p \times q}$  as  $A = KL$  as in Section 7.4f, where  $K$  and  $L$  have full column and row rank respectively, equal to  $r_A$ . Then  $M$  of (1) is

$$M = L'(K'AL')^{-1}K'. \quad (2)$$

Readers should verify for themselves that (2) does satisfy (1), using the nonsingularity of  $K'K$  and  $LL'$  (see theorem and corollary, Section 7.7e) to do so.

**Example.** The factoring given by the theorem of Section 7.4f, namely  $A = KL$ , is exemplified by

$$A = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} = KL. \quad (3)$$

From this,

$$K'AL' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} = 6 \begin{bmatrix} 1 & -2 \\ -2 & 10 \end{bmatrix}.$$

Hence from (2)

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \left( \frac{1}{36} \right) \begin{bmatrix} 10 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 4 \end{bmatrix} = \frac{1}{18} \begin{bmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{bmatrix}. \quad (4)$$

It is easily verified that  $A$  and  $M$  satisfy the four equations in (1).

$M$  is generally known as the Moore-Penrose inverse of  $A$ ; and the four conditions in (1) are usually called the Penrose conditions. For  $A$  of order  $p \times q$ , the order of  $M$  is  $q \times p$ . When  $A$  is nonsingular,  $K$  and  $L$  are likewise and the Moore-Penrose inverse is the regular inverse,  $M = A^{-1}$ .

### 2. GENERALIZED INVERSES

The matrix  $M$  defined by the four Penrose conditions in (1) is unique for a given  $A$ . But there are many matrices  $G$  which satisfy just the first Penrose condition:

$$AGA = A. \quad (5)$$

Nevertheless, they are of such importance in solving linear equations that we direct most attention to those matrices  $G$  rather than to the Moore-Penrose inverse  $M$ .

Any matrix  $G$  satisfying (5) is called a generalized inverse of  $A$ ; and, by (5), when  $A$  is  $p \times q$  then  $G$  is  $q \times p$ . Although the name generalized inverse has not been adopted universally, it is widely used; other names are discussed in Section 3. Notice that  $G$  is "a" generalized inverse of  $A$  and not "the" generalized inverse, because for any given  $A$  there are generally many matrices  $G$  satisfying (5). The exception is when  $A$  is nonsingular, in which case there is only one  $G$  satisfying (5) and it is the regular inverse  $G = A^{-1} = M$ .

Two derivations of  $G$  are now given. A third is discussed in Section 4.

#### a. Derivation from row operations

As part of the development of the equivalent canonical form  $PAQ = C$  of Section 7.4, we often find that row operations alone, represented by  $P$ , reduce  $A$  to the form

$$PA = \begin{bmatrix} T & R \\ 0 & 0 \end{bmatrix}$$

where  $T$  is upper triangular (Section 1.6) and nonsingular of order  $r_A$ . Then

$$G = \begin{bmatrix} T^{-1} & 0 \\ 0 & 0 \end{bmatrix} P \quad (6)$$

is a generalized inverse of  $A$  satisfying (5). On the occasion when some column interchanges (represented by  $Q$ ) are needed to achieve the appropriate form of  $PA$ , we then have

$$PAQ = \begin{bmatrix} T & R \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad G = Q \begin{bmatrix} T^{-1} & 0 \\ 0 & 0 \end{bmatrix} P.$$

**Example (continued).** Row operations for deriving the equivalent canonical reduction of  $A$  in (3) are

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

so that  $G$  of (6) is

$$G = \begin{bmatrix} \left( \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right)^{-1} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (7)$$

#### b. Derivation from the diagonal form

Although in Chapter 7 the diagonal form is emphasized more for symmetric matrices (Section 7.6) than it is for nonsymmetric matrices one can always, as part of the equivalent canonical form derivation, obtain  $P$  and  $Q$  as products of elementary operators such that for any matrix  $A$  the product  $PAQ$  is diagonal; i.e.,

$$PAQ = \Delta = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} \quad (8)$$

where  $D_r$  is diagonal of order  $r = r_A$  with all diagonal elements nonzero. Then

$$\Delta^{-} = \begin{bmatrix} D_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

is a generalized inverse of  $\Delta$  with  $\Delta \Delta^{-} \Delta = \Delta$ ; and

$$G = Q \begin{bmatrix} D_r^{-1} & X \\ Y & Z \end{bmatrix} P \quad (9)$$

for any matrices  $X$ ,  $Y$  and  $Z$  of appropriate orders is a generalized inverse of  $A$  satisfying  $AGA = A$ . The reader should verify this (using  $A = P^{-1} \Delta Q^{-1}$ ).

#### Example

$$PAQ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

illustrates (8), so that from (9)

$$G = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & x_1 \\ 0 & \frac{1}{2} & x_2 \\ y_{11} & y_{12} & z_1 \\ y_{21} & y_{22} & z_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \quad (10)$$

for any values of the  $x$ 's,  $y$ 's and  $z$ 's is a generalized inverse of  $A$ , meaning that  $AGA = A$ . The reader can easily verify this.



Note the occurrence of arbitrary  $X$ ,  $Y$  and  $Z$  in (9) illustrated with  $x$ 's,  $y$ 's and  $z$ 's as elements in (10). These can be any values at all and the defining relationship  $AGA = A$  still holds. The presence of these arbitrary values emphasizes the existence of many generalized inverses  $G$  for any particular  $A$ . The nature of this arbitrariness is characterized further in Section 5. Only when  $A$  is square of full rank is there but one  $G$ , and that is the regular inverse,  $A^{-1}$ . The easiest form of  $G$  in (9) is, of course, with  $X, Y, Z$  null, giving

$$G = Q \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P.$$

### 3. OTHER NAMES AND SYMBOLS

A useful alternative symbol for  $G$  satisfying  $AGA = A$  is  $A^-$ ;

$$G \equiv A^- \quad (11)$$

A reflexive generalized inverse, to be denoted by  $A_r^-$ , is a matrix satisfying the first two Penrose conditions:

$$AA_r^-A = A \quad \text{and} \quad A_r^-AA_r^- = A_r^- \quad (12)$$

Given a generalized inverse  $A^-$  it is easily shown that

$$A_r^- = A^-AA^- \quad (13)$$

is a reflexive generalized inverse satisfying (12). So also is (9) with  $Z = YD_rX$ :

$$A_r^- = Q \begin{bmatrix} D_r^{-1} & X \\ Y & YD_rX \end{bmatrix} P. \quad (14)$$

Widely used symbols for the Moore-Penrose inverse, rather than  $M$  of (1), are  $A^+$  and  $A^\dagger$ . Analogous to (13) we then have

$$A^+ = A'(AA')^-A(A'A)^-A' \quad (15)$$

as an alternative to (2) for the Moore-Penrose inverse of  $A$ .

The literature abounds with names for these inverses: pseudo-inverse,  $p$ -inverse,  $g$ -inverse, effective inverse, normalized generalized inverse and weak generalized inverse, and others (e.g., Searle, 1971, p. 19). Any reader

encountering these needs to be sure of understanding the meanings given to them, because they are not universal. Some, for example, refer to a reflexive generalized inverse that also satisfies one but not both of the Penrose symmetry conditions, namely (iii) or (iv) in (1).

Notation also comes in many forms, one popular variant being to use subscripts for denoting which of the Penrose conditions are satisfied. For example, a  $g_1$ -inverse would be  $A^-$  or  $G$ , satisfying just condition (i),  $AGA = A$ ; similarly, a  $g_{12}$ -inverse would be a reflexive generalized inverse and the  $g_{1234}$ -inverse would be the Moore-Penrose inverse  $A^+$ . This notation is useful when dealing with matrices satisfying different combinations of the Penrose conditions. In our case, when focusing prime attention on those satisfying just the first condition, it is easier to use a single symbol  $G$  defined by  $AGA = A$ , although the alternative  $A^-$  is frequently used.

### 4. AN ALGORITHM

#### a. An easy form

Suppose that in  $A_{p \times q}$  the leading principal submatrix  $A_{11}$  is nonsingular of rank  $r_A$ . Then a generalized inverse obtainable from partitioning  $A$  as

$$A_{p \times q} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \text{is} \quad G_{q \times p} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad (16)$$

where the null matrices in  $G$  have appropriate order to make  $G$  be of order  $q \times p$ . Then verification of  $AGA = A$  involves using  $A_{22} = A_{21}A_{11}^{-1}A_{12}$ , which comes from (38) of Section 6.9.

**Example.** A generalized inverse of

$$A = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix} \quad \text{is} \quad G = \frac{1}{7} \begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

#### b. A general form

There is no need for the nonsingular submatrix of order  $r$  to be in the leading position. It can be anywhere in  $A$ , whereupon the following algorithm can be developed.

Suppose the nonsingular submatrix is not in the leading position. Let  $R$  and  $S$  be permutation matrices (Section 6.9) such that  $RAS$  brings a nonsingular submatrix of

order  $r$  to the leading position; i.e.,

$$\mathbf{RAS} = \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \quad (17)$$

with  $\mathbf{B}_{11}$  being nonsingular of order  $r$ . Then a generalized inverse of  $\mathbf{B}$  is

$$\mathbf{F} = \begin{bmatrix} \mathbf{B}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \text{ and } \mathbf{G} = \mathbf{SFR} \quad (18)$$

is a generalized inverse of  $\mathbf{A}$ . Now, because of the orthogonality of permutation matrices, (17) implies

$$\mathbf{A} = \mathbf{R}'\mathbf{B}\mathbf{S}' = \mathbf{R}' \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \mathbf{S}' \quad (19)$$

Insofar as  $\mathbf{B}_{11}$  is concerned this product represents the operations of returning the elements of  $\mathbf{B}_{11}$  to their original positions in  $\mathbf{A}$ . Hence, because (18) is equivalent to

$$\mathbf{G} = \mathbf{SFR} = (\mathbf{R}'\mathbf{F}'\mathbf{S}')' = \left\{ \mathbf{R}' \begin{bmatrix} (\mathbf{B}_{11}^{-1})' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{S}' \right\}', \quad (20)$$

we see by comparing (20) with (19) that the product involving  $\mathbf{R}'$  and  $\mathbf{S}'$  in (20) represents putting the elements of  $(\mathbf{B}_{11}^{-1})'$  into the corresponding positions (of  $\mathbf{G}$ ) that the elements of  $\mathbf{B}_{11}$  occupied in  $\mathbf{A}$ . Hence (20) gives the following algorithm for finding  $\mathbf{G}$ .

- (i) In  $\mathbf{A}$  find any nonsingular submatrix of order  $r_A$ . Denote it by  $\mathbf{W}$  [in place of  $\mathbf{B}_{11}$  in (17)]. It is not necessary for  $\mathbf{W}$  to come from  $r_A$  adjacent rows and columns of  $\mathbf{A}$ .
- (ii) Invert and transpose  $\mathbf{W}$ :  $(\mathbf{W}^{-1})'$ .
- (iii) In  $\mathbf{A}$  replace each element of  $\mathbf{W}$  by the corresponding element of  $(\mathbf{W}^{-1})'$ ; i.e., when  $a_{ij} = w_{st}$ , the  $(s, t)$ th element of  $\mathbf{W}$ , then replace  $a_{ij}$  by  $w^{t,s}$ , the  $(t, s)$ th element of  $\mathbf{W}^{-1}$ , equivalent to the  $(s, t)$ th element of the transpose of  $\mathbf{W}^{-1}$ .
- (iv) Replace all other elements of  $\mathbf{A}$  by zero.
- (v) Transpose the resulting matrix.
- (vi) The result is  $\mathbf{G}$ , a generalized inverse of  $\mathbf{A}$ .

Note that this procedure is *not* equivalent, in (iii), to replacing elements of  $\mathbf{W}$  in  $\mathbf{A}$  by the elements of  $\mathbf{W}^{-1}$  (and others by zero) and then in (v) transposing. It is if  $\mathbf{W}$  is symmetric. Nor is it equivalent to replacing, in (iii), elements of  $\mathbf{W}$  in  $\mathbf{A}$  by elements of  $\mathbf{W}^{-1}$  (and others by zero) and then in (v) not transposing (see Exercise 8). In general, the algorithm must be carried out exactly as described.

[8.5]

Example. The matrix

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

has the following matrices, among others, as generalized inverses:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1.5 & 2.5 \\ 0 & .5 & -.5 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & -.5 & 1.5 \\ 0 & 0 & 0 \\ 0 & .1 & -.1 \end{bmatrix} \text{ and } \begin{bmatrix} .25 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1.5 & 0 & .2 \end{bmatrix}$$

They are derived from inverting the  $2 \times 2$  submatrices

$$\begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 15 \\ 1 & 5 \end{bmatrix} \text{ and } \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}, \text{ respectively.}$$

The algorithm can be simplified slightly for symmetric matrices as discussed in Section 6.

## 5. ARBITRARINESS IN A GENERALIZED INVERSE

The non-uniqueness of matrices  $\mathbf{G}$  satisfying  $\mathbf{AGA} = \mathbf{A}$  is evident in (9), which holds for any  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  of appropriate order. More useful characterizations of arbitrariness are based on Urquhart (1969), as also described in Searle (1971). First, a generalization of (16) is that

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} - \mathbf{U}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{V} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{W}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix} \quad (21)$$

is a generalized inverse of  $\mathbf{A}$  for arbitrary  $\mathbf{U}$ ,  $\mathbf{V}$  and  $\mathbf{W}$ . An important consequence of (21) is the evidence it provides for the existence of nonsingular generalized inverses of singular matrices. For example, suppose  $\mathbf{U} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{W}$ ,  $\mathbf{V} = \mathbf{0}$  and  $\mathbf{W}$  is nonsingular. Then (21) gives

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{W} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}, \quad (22)$$

which is nonsingular. More generally, when  $\mathbf{A}$  is rectangular and  $\mathbf{W}$  has full row rank then  $\mathbf{G}$  of (22) has full row rank also, i.e.,  $r_G > r_A$ .

A further characterization of the arbitrariness is that when  $G$  is a generalized inverse of  $A$  so also is

$$G^* = GAG + (I - GA)T + S(I - AG) \quad (23)$$

for any matrices  $T$  and  $S$  of appropriate order. A consequence of (23) is the very useful result that given one particular generalized inverse  $G$ , all others can be generated from it by using (23) for all possible arbitrary matrices  $T$  and  $S$ . To see that this is so, suppose  $\hat{G}$  is some generalized inverse different from  $G$ . Then putting  $T = \hat{G}$  and  $S = G\hat{G}$  into (23) yields  $G^*$  as  $\hat{G}$ . The existence of a  $T$  and an  $S$  that achieve this shows that no matter what generalized inverse we have, it can always be written as (23) with an appropriate  $T$  and  $S$ ; i.e., (23) generates all generalized inverses of  $A$ .

## 6. SYMMETRIC MATRICES

Generalized inverses of symmetric matrices have some interesting properties. First, from the diagonal form of  $A = A'$ , namely

$$PAP' = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

developed in (26) of Section 7.6, generalized inverses corresponding to (9) are

$$G = P \begin{bmatrix} D_r^{-1} & X \\ Y & Z \end{bmatrix} P' \quad (24)$$

This shows at once that a generalized inverse of a symmetric matrix is not necessarily symmetric. However, transposing the equation  $AGA = A$  gives

$$AG'A = A, \quad (25)$$

so establishing that when  $G$  is a generalized inverse of the symmetric matrix  $A$ , then so is the transpose  $G'$ , whether  $G$  is symmetric or not.

### a. Non-negative definite matrices

Equation (35) of Section 7.7c indicates that when  $A$  is n.n.d. and symmetric,  $A = KK'$  for  $K$  of full column rank. Then, by Section 7.7e,  $K'K$  is nonsingular and so by using  $K'$  in (2) in place of  $L$ , the Moore-Penrose of  $A$  is

$$A^+ = K(K'AK)^{-1}K' = K(K'K)^{-2}K'. \quad (26)$$

$(K'K)^{-2}$  is, of course, the square of  $(K'K)^{-1}$ . And, as would be expected, (15) and (26) are equivalent (see Exercise 22).

Confining attention to n.n.d. matrices ensures that  $(K'K)^{-1}$  exists. As indicated in Section 7.6d(i), the canonical form

$$RAR' = \begin{bmatrix} I_{r_A} & 0 \\ 0 & 0 \end{bmatrix}$$

exists for all symmetric matrices; but for those which are not n.n.d.,  $R$  may have rows that have  $i = \sqrt{-1}$  as a factor. In that case,  $K$  of  $A = KK'$  in Section 7.7c will still have full column rank, but the imaginary factor in  $R$  may also be in  $K$ , and  $K'K$  may be singular (which it cannot be, for real  $K$ ). For example, the columns of

$$K = \begin{bmatrix} i & 0 \\ 0 & i \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

are LIN, but  $K'K = 0$  is singular.

### b. A general algorithm

The general form of the algorithm in Section 4b can be simplified when  $A$  is symmetric. Principal submatrices of  $A$  are then symmetric and so by using them, the transposing in (iii) and (v) can be ignored. The algorithm can then become as follows.

- (i) In  $A$  find a nonsingular principal submatrix,  $W$ , of order  $r_A$ .
- (ii) Invert  $W$ .
- (iii) In  $A$  replace each element  $w_{ij}$  of  $W$  by the element  $w^{ij}$  of  $W^{-1}$ .
- (iv) Replace all other elements of  $A$  by zero.
- (v) The result is  $G$ , a generalized inverse of  $A$ .

Of course, when  $A$  is symmetric and a nonsymmetric nonprincipal submatrix is used for  $W$ , then the general algorithm of Section 4b must be used.

### c. The matrix $X'X$

The matrix  $X'X$  has an important role in statistics where it arises in least squares equations  $X'Xb = X'y$ . Properties of generalized inverses of  $X'X$  are therefore of interest.

**Theorem.** When  $G$  is a generalized inverse of  $X'X$ :

- (i)  $G'$  is also a generalized inverse of  $X'X$ .
- (ii)  $XGX' = X$ ; i.e.,  $GX'$  is a generalized inverse of  $X$ .
- (iii)  $XGX'$  is invariant to  $G$ .
- (iv)  $XGX'$  is symmetric, whether  $G$  is or not.

**Proof.** (i) is a consequence of (25). Applying (7) of Section 3.1b yields (ii). To prove (iii) suppose  $F$  is a generalized inverse of  $X'X$  possibly different from  $G$ . Then (ii) gives  $XGX'X = XFX'X$  and (7) of Section 3.1b applied to this yields (iii). Finally, suppose  $T$  is a symmetric generalized inverse of  $X'X$ . ( $T = GX'XG'$  is always one possibility.) Then  $XTX'$  is symmetric. But  $XGX' = XTX'$  by (iii) and so  $XGX'$  is symmetric too. QED

**Corollary.**  $XGX'X = X$  and  $X'XGX' = X'XG'X' = X'$ ; and  $XGX' = XG'X'$ . In the latter, note that both expressions are symmetric, even without  $G$  equalling  $G'$ .

A result relating  $G$  to the Moore–Penrose inverse  $X^+$  of  $X$  is

$$XX^+ = XGX'. \quad (27)$$

This is established by applying (15) to  $X$ . It is important to note from (27) that  $X^+$  is not necessarily  $GX'$ , even though  $GX'$  is a generalized inverse of  $X$ . (See Exercise 7b.)

## EXERCISES

- Show that (2) and (4) satisfy (1).
- Show that (6) and (9), and the examples thereof, (7) and (10), all satisfy (5).
- Show that (13) and (14) satisfy (12).
- Show that (15) satisfies (1), and calculate (4) using (15).
- Show that (16), (21) and (23) satisfy (5).
- Show that (21) and (23) are consistent.  
[Hint: Use (16) for  $G$  and partition  $T$  and  $S$ .]
- For  $G$  satisfying  $X'XGX'X = X'X$  show that
  - $GX'$  is a generalized inverse of  $X$  satisfying three of the Penrose conditions.
  - $XX^+ = XGX'$ .
- Section 4b describes an algorithm for deriving a generalized inverse. Prove in general that neither one nor both of the matrix transpositions in steps (iii) and (v) of the algorithm can be omitted, but that when  $W = W'$  the transposition in (iii) can be. Illustrate these

results for

$$A_1 = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 2 & 2 & 6 \\ 2 & 3 & 8 \\ 6 & 8 & 22 \end{bmatrix}.$$

Also find a generalized inverse of  $A_2$  using a nonsymmetric non-principal submatrix for  $W$ .

9. Calculate generalized inverses of

$$A = \begin{bmatrix} 2 & 3 & 1 & -1 \\ 5 & 8 & 0 & 1 \\ 1 & 2 & -2 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & 2 \\ 7 & 8 & 10 & 7 \\ 2 & 1 & 1 & 6 \end{bmatrix}$$

using (6), (9), (16), (21) and (23). Check your result in each case.

10. Find the Moore–Penrose inverse of  $C$  using (2) and (15).

$$C = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 5 \\ 0 & 1 & -1 \\ 1 & 3 & -1 \end{bmatrix}.$$

- Find a generalized inverse of each of the following matrices:
  - $PAQ$  when  $P$  and  $Q$  are nonsingular.
  - $GA$  when  $G$  is a generalized inverse of  $A$ .
  - $kA$  when  $k$  is a scalar.
  - $ABA'$  when  $A$  is orthogonal and  $B$  is idempotent.
  - $J$  when  $J$  is square with every element unity.
- What kinds of matrices
  - are their own generalized inverses?
  - have their transposes as a generalized inverse?
  - have an identity matrix for a generalized inverse?
  - have every matrix of order  $p \times q$  for a generalized inverse?
  - have only nonsingular generalized inverses?
  - are their own Moore–Penrose inverses?
- Calculate the Moore–Penrose and two other generalized inverses of  $x = [7 \ 2 \ 3]'$ .

14. Using (9), or any method of deriving a generalized inverse, discuss the possibility of finding one that satisfies Penrose condition (i) and either or both of (iii) and (iv), but not (ii). Illustrate your results with an example.
15. Show that all generalized inverses of a full column rank matrix are left inverses satisfying three of the Penrose conditions.
16. Construct a generalized inverse in terms of  $X^-$  of each of  $LX$ ,  $XM$  and  $LXM$  where  $L$  and  $M$  are nonsingular.
17. Prove that  $B^-A^-$  is a generalized inverse of  $AB$  if and only if  $A^-ABB^-$  is idempotent.
18. Prove that  $r_G = r_A$  for  $AGA = A$  if and only if  $G$  is a reflexive generalized inverse.
19. When  $K$  is idempotent and  $Z^-$  is a generalized inverse of  $Z = KAK$ , prove that  $KZ^-K$  is also a generalized inverse of  $Z$ .
20. When  $G$  is a generalized inverse of  $A$ , show that it is also a generalized inverse of  $AG$  if and only if  $G^2$  is also a generalized inverse of  $A$ .
21. When  $G$  is a generalized inverse of  $A$ , with  $A$  being symmetric, show that  $G^2$  is a generalized inverse of  $A^2$  if  $GA$  is symmetric.
22. For  $M$  being the Moore–Penrose inverse of a symmetric matrix  $A$ , show that  $M^2$  is the Moore–Penrose inverse of  $A^2$ ; then show that (15) reduces to (25).
23. (a) Show that any generalized inverse of a non-null vector satisfies three of the Penrose conditions.  
(b) Find a generalized inverse of a vector  $t$  that involves only one nonzero element of  $t$ .  
(c) Find the Moore–Penrose inverse of  $t$ .
24. If  $P_{m \times q}$  and  $D_{m \times m}$  have rank  $m$ , show that  $D^{-1} = P(P'DP)^-P'$ .
25. When  $P = I - X(X'A^{-1}X)^{-1}X'A^{-1}$  with  $A$  symmetric and nonsingular, prove that  $P'(PAP')^-P = A^{-1}P$ .
26. When  $G$  is a generalized inverse of  $X'X$  define:

$$b = GX'y, \quad s = (y - Xb)'(y - Xb)$$

$$\text{and } b_0 = b - GQ(Q'GQ)^{-1}(Q'b - m) \quad \text{with } Q' = T'X.$$

Prove that

$$(a) \quad s = y'y - b'X'y;$$

$$(b) \quad Q'b_0 = m;$$

$$(c) \quad (y - Xb_0)'(y - Xb_0) = s + (Q'b - m)'(Q'GQ)^{-1}(Q'b - m).$$

27. When  $M$  is idempotent and  $V$  is nonsingular with  $MV = VM'$ , show that  $M'(MVM')^-$ ,  $M'(MVM')^-M$  and  $V^{-1}M$  are all generalized inverses of  $MV$ .
28. For  $A$  and  $X$  symmetric with  $AX = 0$ , and for  $X$  idempotent and  $A + X$  nonsingular, prove that  $(A + X)^{-1}$  is a generalized inverse of both  $A$  and  $X$ .
29. For  $A'$  and  $F$  both of full column rank with  $FF' = I - A^-A$ , prove that  $P = [A^- \quad F]$  is the inverse of  $L = \begin{bmatrix} A \\ F' \end{bmatrix}$ .
30. For  $X$  partitioned as  $X = [X_1 \quad X_2]$  with  $X_1$  of full column rank  $r_X$ , prove that  $X(X'X)^-X = X_1'(X_1'X_1)^{-1}X_1'$ .
31. For  $X$  of the preceding exercise, develop a generalized inverse of  $X'V^{-1}X$  that involves  $(X_1'V^{-1}X_1)^{-1}$ .  
[Hint: Use  $X_2 = X_1P$  for some  $P$ .]
32. When  $CJ = 0_{n \times n} = JC$  for  $J$  having every element unity, show that provided  $C + \lambda J$  is nonsingular  
(a)  $(C + \lambda J)^{-1}$  is a generalized inverse of  $C$ .  
(b)  $(C + \lambda J)^{-1} - (1/\lambda n^2)J$  is the Moore–Penrose inverse of  $C$ .
33. When  $M$  has full column rank, show that the Moore–Penrose inverse of  $MM'$  is  $M(M'M)^{-2}M'$ . Hence prove that if  $(MM')^-$  is any generalized inverse of  $MM'$  then  $M'(MM')^-M = I$ .
34. (a) Prove that all left and right inverses are reflexive generalized inverses that also satisfy one of the two Moore–Penrose symmetry conditions.  
(b) Prove that all generalized inverses of a full row (column) rank matrix are right (left) inverses.
35. For symmetric  $V$ , for  $P = X'V^{-1}X$  and for  $d = 1/(1 + \beta'P\beta)$ , prove that  
(a)  $X(X'V^{-1}X)^-X'V^{-1}X = X$ .  
(b)  $V^{-1}XP^-X'V^{-1} - dV^{-1}X\beta\beta'X'V^{-1}$  is a generalized inverse of  $XP^-X' - dP\beta\beta'P$ .

36. For  $M$  of full column rank and  $X = MT$ , prove that  $X$  and  $X'(MM')^{-1}X$  have the same rank.
37. The matrix

$$P = I - X(X'X)^{-1}X'$$

has a number of features that are useful in statistical applications. For  $X$  having  $n$  rows, verify that  $P$  has the following properties:

- (i)  $P$  is symmetric, idempotent and invariant to  $(X'X)^{-1}$ .
- (ii)  $PX = 0$  and  $XP = 0$ .
- (iii)  $r_P = n - r_X$ .
- (iv) Columns of  $P$  are orthogonal to and LIN of columns of  $X$ .
- (v) All  $n$ -order vectors are linear combinations of the columns of  $X$  and  $P$ . [Columns of  $X$  and  $P$  span Euclidean  $n$ -space (see Section 6.11c) and the column (row) space of  $P$  is the orthogonal complement of the column space of  $X$  (see Section 6.11f).]
- (vi) Any  $B$  for which  $BX = 0$  has rows that are linear combinations of columns (rows) of  $P$ ; hence  $B = KP$  for some  $K$ . (The row space of  $B$  is in the column space of  $P$ , the orthogonal complement of the column space of  $X$ .)
- (vii)  $P$  has zero row sums when row sums of  $X$  are all the same and nonzero.
- (viii)  $P$  has zero row sums when a column of  $X$  is  $\mathbf{1}$ .
- (ix)  $I - (1/n)J_n$  is a special case of  $P$ , with  $X = \mathbf{1}_n$ .
- (x) If  $A$  is symmetric and satisfies  $AX = 0$ , then  $A = PLP$  for some  $L$ .

See also Section 12.2 and Exercise 13.5.

38. Prove that a generalized inverse of  $X'V^{-1}X$  for symmetric  $R$  is  $(X'R^{-1}X)^{-1}X'R^{-1}X(X'V^{-1}X)^{-1}$ .
39. Prove that
- (a)  $r_{AB} = r_A$  implies  $B(AB)^{-1}$  is a generalized inverse of  $A$ .
  - (b)  $r_{AB} = r_B$  implies  $(AB)^{-1}A$  is a generalized inverse of  $B$ .
  - (c)  $r_{ABC} = r_B$  implies  $C(ABC)^{-1}A$  is a generalized inverse of  $B$ .
40. Prove
- (a)  $(I - KK^{-1})L^{-1}LK = -(I - KK^{-1})(I - L^{-1}L)K$ .
  - (b)  $LKK^{-1}(I - L^{-1}L) = -L(I - KK^{-1})(I - L^{-1}L)$ .
  - (c)  $(LK)^{-1} = K^{-1}L^{-1} - K^{-1}(I - L^{-1}L)[(I - KK^{-1})(I - L^{-1}L)]^{-1}(I - KK^{-1})L^{-1}$ .

## CHAPTER 9

### SOLVING LINEAR EQUATIONS

Linear equations in several unknowns can be represented in matrix form as  $Ax = y$  where  $x$  is the vector of unknowns,  $y$  is a vector of known values and  $A$  is the matrix of coefficients. Equations of this nature have been mentioned on a number of occasions in preceding chapters. So long as  $A$  has an inverse the solution is  $x = A^{-1}y$ , a situation demanding that  $A$  be square and that its rank equal its order. The more general case is now considered, that of having  $p$  equations in  $q$  unknowns,

$$A_{p \times q} x_{q \times 1} = y_{p \times 1},$$

with the rank of  $A$  being  $r_A = r$ . The instances when the solution is  $x = A^{-1}y$  are then particular cases of this, namely when  $p = q = r$ .

Throughout this chapter we deal with linear equations only, of the kind  $3x_1 + 7x_2 + 2x_3 = 13$ , i.e., linear in scalar unknowns,  $x_1$ ,  $x_2$  and  $x_3$ . At no time are equations involving powers of the unknowns considered. Thus by "equations" we will always mean "linear equations." The manner of presentation established in Chapter 8 is continued, with important results being stated formally as theorems.

#### 1. EQUATIONS HAVING MANY SOLUTIONS

The equation  $3x_1 = 12$  has but a single solution,  $x_1 = 4$ . Likewise the two simultaneous equations  $2x_1 + x_2 = 7$  and  $x_1 + 3x_2 = 16$  have the single solution  $x_1 = 1$  and  $x_2 = 5$ . But consider the equations

$$\begin{aligned} 2x_1 + 3x_2 + x_3 &= 14 \\ x_1 + x_2 + x_3 &= 6 \\ 3x_1 + 5x_2 + x_3 &= 22 \end{aligned} \tag{1}$$

By direct substitution it will be found that the values  $x_1 = 4$ ,  $x_2 = 2$  and  $x_3 = 0$  are a solution, and so is the set of values  $x_1 = 6$ ,  $x_2 = 1$  and  $x_3 = -1$ . Indeed, it is easily verified that each of the following sets of values is a solution.

TABLE 9.1. FIVE SOLUTIONS TO EQUATIONS (1)

Unknowns	Solutions <sup>a</sup>				
	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$	$\bar{x}_5$
$x_1$	4	6	0	-6	3070
$x_2$	2	1	4	7	-1531
$x_3$	0	-1	2	5	-1533

<sup>a</sup>The notation  $\bar{x}$  is used for a solution, to distinguish it from  $x$ , the vector of unknowns.

Readers accustomed to thinking of a set of linear equations as having only a single solution may find this idea of many solutions somewhat novel. Nevertheless, it is a situation of wide occurrence, especially in the analysis of linear statistical models as described in Chapter 15.

## 2. CONSISTENT EQUATIONS

### a. Definition

Consider the two equations:

$$\begin{aligned}x_1 + x_2 &= 5 \\ 2x_1 + 2x_2 &= 11.\end{aligned}$$

Obviously if one is true the other cannot be, for the second is incompatible, or inconsistent, with the first. Considered as a set of linear equations they are said to be *inconsistent*. Similarly, consider

$$\begin{aligned}2x_1 + 3x_2 + x_3 &= 14 \\ x_1 + x_2 + x_3 &= 6 \\ 3x_1 + 5x_2 + x_3 &= 19;\end{aligned}\tag{1a}$$

subtracting twice the first from the sum of the second and third leads to the meaningless result  $0 = -3$ . This is because the left-hand side of the third equation is twice that of the first minus that of the second, but without the same relationship being true of the right-hand sides. This incompatibility

[9.2]

leads to the equations being described as inconsistent. Now consider equations (1) again. They differ from (1a) only in the right-hand side of the last equation, which is 22 instead of 19. A consequence of the 22, however, is that in (1) the right-hand side of the last equation is twice that of the first minus that of the second, the same relationship that holds for the left-hand sides. Now consider these equations in matrix form  $Ax = y$ :

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix}.\tag{2}$$

The relationship which exists among the individual equations of (1) applies to both the rows of  $A$  and the corresponding elements of  $y$ : the last row of  $A$  (element of  $y$ ) is twice the first minus the second. Equations of this nature, in which linear relationships existing among the rows of  $A$  also hold among the corresponding elements of  $y$ , are said to be *consistent*. A formal definition is as follows.

**Definition.** The linear equations  $Ax = y$  are defined as consistent if linear relationships existing among the rows of  $A$  also exist among the corresponding elements of  $y$ .

The definition does not require that linear relationships exist among the rows of  $A$ ; but if they do, the same relationships must also exist among the corresponding elements of  $y$  for the equations to be consistent. When  $A$  has full row rank there are no linear relationships among the rows of  $A$  and therefore none that the elements of  $y$  must satisfy. Therefore equations  $Ax = y$  with  $A$  of full row rank (including  $A$  nonsingular) are always consistent.

### b. Existence of solutions

The importance of the concept of consistency is that consistent equations can be solved and inconsistent equations cannot. This fact represents the starting point for a series of theorems on the solution of linear equations.

**Theorem 1.** A set of linear equations can be solved if, and only if, they are consistent.

**Proof.** Suppose  $A_{p \times q}$  has rank  $r$ . Interchange rows of  $A$  and elements of  $y$ , and also columns of  $A$  and elements of  $x$ , so that the leading  $r \times r$  submatrix of  $A$ , to be called  $A_{11}$ , is nonsingular. Then by (37) of Section 6.9,  $A$  can be factored as

$$A = \begin{bmatrix} A_{11} & A_{11}L \\ KA_{11} & KA_{11}L \end{bmatrix} = \begin{bmatrix} I \\ K \end{bmatrix} A_{11} [I \ L]\tag{3}$$

for some matrices  $\mathbf{K}$  and  $\mathbf{L}$ . Partition  $\mathbf{x}$  and  $\mathbf{y}$  conformably with (3) so that the equations  $\mathbf{Ax} = \mathbf{y}$  are

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{K} \end{bmatrix} \mathbf{A}_{11} [\mathbf{I} \quad \mathbf{L}] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}. \quad (4)$$

Relationships between the two groups of rows of  $\mathbf{A}$  in (3) are represented by  $\mathbf{K}$ . Therefore if the equations  $\mathbf{Ax} = \mathbf{y}$  are consistent, the same relationships exist among the elements of  $\mathbf{y}$  and so  $\mathbf{y}_2 = \mathbf{Ky}_1$ ; whereupon (4) becomes

$$\mathbf{A}_{11}(\mathbf{x}_1 + \mathbf{Lx}_2) = \mathbf{y}_1$$

and

$$\mathbf{KA}_{11}(\mathbf{x}_1 + \mathbf{Lx}_2) = \mathbf{Ky}_1.$$

Clearly, every  $\mathbf{x}_1$  and  $\mathbf{x}_2$  that satisfy the first of these also satisfy the second; and from the first

$$\mathbf{x}_1 = \mathbf{A}_{11}^{-1} \mathbf{y}_1 - \mathbf{Lx}_2 = \mathbf{A}_{11}^{-1}(\mathbf{y}_1 - \mathbf{A}_{11} \mathbf{Lx}_2), \quad (5)$$

where  $\mathbf{A}_{11} \mathbf{L}$  is a submatrix of  $\mathbf{A}$  in (3). Therefore any vector  $\mathbf{x}_2$ , together with the corresponding  $\mathbf{x}_1$  given by using that  $\mathbf{x}_2$  in (5), constitute a solution  $\mathbf{x}$  that satisfies  $\mathbf{Ax} = \mathbf{y}$ ; i.e., if the equations are consistent they have a solution.

Conversely, suppose  $\tilde{\mathbf{x}}$  is a solution. Then from (4)

$$\begin{bmatrix} \mathbf{A}_{11}(\tilde{\mathbf{x}}_1 + \mathbf{L}\tilde{\mathbf{x}}_2) \\ \mathbf{KA}_{11}(\tilde{\mathbf{x}}_1 + \mathbf{L}\tilde{\mathbf{x}}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

and so  $\mathbf{y}_2 = \mathbf{Ky}_1$ ; i.e., if the equations have a solution they are consistent. QED

**Example.** Equations (1) are consistent.  $\mathbf{A}$  has rank 2, and its leading submatrix of order  $2 \times 2$  is nonsingular. Hence by (5) the elements of a solution vector are given by

$$\mathbf{x}_1 = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}^{-1} \left[ \begin{pmatrix} 14 \\ 6 \end{pmatrix} - \begin{pmatrix} x_3 \\ x_3 \end{pmatrix} \right] = \begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 14 - x_3 \\ 6 - x_3 \end{bmatrix} = \begin{bmatrix} 4 - 2x_3 \\ 2 + x_3 \end{bmatrix}.$$

Thus a solution vector is

$$\tilde{\mathbf{x}} = \begin{bmatrix} 4 - 2x_3 \\ 2 + x_3 \\ x_3 \end{bmatrix} \quad (6)$$

for any value  $x_3$ . In particular,

$$\text{if } x_3 = 0, \tilde{\mathbf{x}} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} \text{ is a solution, as is } \tilde{\mathbf{x}} = \begin{bmatrix} 6 \\ 1 \\ -1 \end{bmatrix} \text{ for } x_3 = -1.$$

These are two of the solutions shown in Table 9.1. The reader should verify that the other solutions given there can also, for appropriate values of  $x_3$ , be derived from (6).

That equations (1) have many solutions is evident from the form of the general solution given in (6) for, by direct substitution, we find that

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{bmatrix} \begin{bmatrix} 4 - 2x_3 \\ 2 + x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix}$$

no matter what the value of  $x_3$  is. This means, of course, that there is an infinite number of solutions.

The arbitrariness of the solutions in the example just discussed is true also of (5), the general form of solution. There, for  $\tilde{\mathbf{x}}_1$  calculated from (5) using any  $\tilde{\mathbf{x}}_2$ , the vector  $\tilde{\mathbf{x}} = [\mathbf{x}'_1 \quad \mathbf{x}'_2]'$  will be a solution for any  $\tilde{\mathbf{x}}_2$ . Thus (5) not only represents a method of obtaining solutions to equations  $\mathbf{Ax} = \mathbf{y}$ , but it also demonstrates the availability of many solutions. Depending on the nonexistence or existence of  $\mathbf{x}_2$  there will be either a single, unique solution or an infinite number of solutions. When  $r = q$ , i.e., when rank equals the number of unknowns, there is no vector  $\mathbf{x}_2$  and solution (5) reduces to  $\mathbf{x} = \mathbf{A}_{11}^{-1} \mathbf{y}_1$ , which is the sole solution. If, in addition,  $p = r = q$ , then  $\mathbf{A}$  and  $\mathbf{A}_{11}$  are the same and the sole solution is the familiar  $\mathbf{x} = \mathbf{A}^{-1} \mathbf{y}$ . When  $r < q$ , then  $\mathbf{x}_2$  exists and there are many solutions, indeed an infinite number of them.

**Example.** The only solution to

$$\begin{bmatrix} 2 & 3 \\ 1 & 1 \\ 3 & 5 \\ 4 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 22 \\ 18 \\ 0 \end{bmatrix} \text{ is } \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 14 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$



## c. Tests for consistency

It is clear from Theorem 1 that consistency is a property required of linear equations before one attempts to solve them, for otherwise they have no solution and trying to look for one is pointless. Before discussing procedures for solving consistent equations we therefore consider methods of testing whether equations are consistent or not. If they are consistent they can be solved, otherwise they cannot.

The most common test for consistency is one involving the partitioned matrix  $[A \ y]$  formed by adding  $y$  as an extra column to  $A$ . Such a matrix is often referred to as an *augmented matrix*,  $A$  augmented by  $y$ . It is used in the following theorem, which provides the test.

**Theorem 2.** The equations  $Ax = y$  are consistent if, and only if, the rank of the augmented matrix  $[A \ y]$  equals the rank of  $A$ .

**Proof.** Write  $A_0$  for  $A_{11}[I \ L]$  in (3). Then if equations  $Ax = y$  are consistent,  $y_2 = Ky_1$  as noted in the proof of Theorem 1, and so the augmented matrix is, using (3),

$$[A \ y] = \begin{bmatrix} A_0 & y_1 \\ KA_0 & Ky_1 \end{bmatrix} = \begin{bmatrix} I \\ K \end{bmatrix} [A_0 \ y_1].$$

This obviously has the same number of LIN rows as  $A$ , and hence the same rank as  $A$ . Conversely, if  $[A \ y]$  and  $A$  have the same rank then so do

$$\begin{bmatrix} A_0 & y_1 \\ KA_0 & y_2 \end{bmatrix} \text{ and } \begin{bmatrix} A_0 \\ KA_0 \end{bmatrix};$$

and this can be true only if  $y_2 = Ky_1$ , i.e., if  $Ax = y$  is consistent. QED

The test for consistency based on Theorem 2 requires ascertaining the rank of  $[A \ y]$  and the rank  $A$ . If the ranks are equal the equations are consistent, otherwise they are inconsistent. Since some methods for solving equations utilize the diagonal form of  $A$ , we give an alternative test for consistency that can be used in conjunction with them.

**Theorem 2a.** If  $A$  is a matrix of  $p$  rows and rank  $r$ , and if  $PAQ$  is a diagonal form of  $A$ , then the equations  $Ax = y$  are consistent if, and only if, the last  $p - r$  elements of  $Py$  are zero.

**Proof.** With  $PAQ$  being a diagonal form of  $A$ ,  $PA$  can be written as  $PA = \begin{bmatrix} A_r \\ 0 \end{bmatrix}$  where  $A_r$  is a matrix of  $r$  LIN rows,  $P$  being a product of elementary operators. Therefore, if the equations  $Ax = y$  are consistent, so are  $P Ax = Py$ ; i.e., so are  $\begin{bmatrix} A_r x \\ 0 \end{bmatrix} = Py$ . Hence the last  $p - r$  elements of  $Py$  must be zero. Conversely, if the last

$p - r$  elements of  $Py$  are zero, the equations  $\begin{bmatrix} A_r \\ 0 \end{bmatrix} x = Py$  are consistent, and since  $P^{-1}$  exists the equations  $P^{-1} \begin{bmatrix} A_r \\ 0 \end{bmatrix} x = y$  are also; i.e.,  $Ax = y$  is a set of consistent equations. QED

As part of the calculations for obtaining  $PAQ$  the test for consistency based on this theorem is patently simple:  $r$ , the rank of  $A$ , is obtained as the number of nonzero elements in  $PAQ$ , and if the last  $p - r$  elements of  $Py$  are zero the equations are consistent, otherwise they are inconsistent.

**Example.** For the matrix  $A$  of equations (2) the diagonal form  $PAQ$  is

$$\begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3/2 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore the rank of  $A$  is 2, and for the equations to be consistent the last  $3 - 2 = 1$  rows of  $Py$  must be zero, which they are,

$$Py = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix} = \begin{bmatrix} 14 \\ -1 \\ 0 \end{bmatrix}.$$

We have just seen how consistent equations  $Ax = y$  can be solved using expression (5). This requires (i) ascertaining the rank  $r$  of  $A$ , (ii) locating  $r$  linearly independent rows of  $A$  and  $r$  linearly independent columns among those rows, and (iii) interchanging, if need be, rows and columns of  $A$  and rows of  $x$  and  $y$  to recreate  $A$  in a form suited to partitioning. There is usually no difficulty in carrying out these steps when only a few equations are involved, but when the number of equations is at all large any one of the steps may be quite tedious. Fortunately, however, there are more uniformly applicable methods that can be used not only when the number of unknowns equals the number of equations (square  $A$ ), as in the preceding example, but also when the number of unknowns differs from the number of equations (rectangular  $A$ ). First we distinguish equations that have just one solution from those that have many solutions.

## 3. EQUATIONS HAVING ONE SOLUTION

Consistent equations  $Ax = y$  have just one solution when  $A$  has full column rank (including the case of  $A$  square and nonsingular, whereupon

the equations are always consistent, with solution  $A^{-1}y$ ). Otherwise, for  $A$  of less than full column rank, there are many solutions. We begin with  $A$  of full column rank.

An easily obtained solution of  $Ax = y$  when  $A$  has full column rank comes from observing that  $L$  in (3) does not then exist. The solution (5) therefore becomes

$$x = A_{11}^{-1}y_1. \quad (7)$$

Another derivation uses any left inverse of the full column rank matrix  $A$  (see Lemma 8, Section 7.5). For  $L$  being a left inverse of  $A$ , with  $LA = I$ ,

$$x = Ly \quad (8)$$

is the solution. A third expression uses  $(A'A)^{-1}A'$  of Section 7.7e for a left inverse of  $A$  so that

$$x = (A'A)^{-1}A'y \quad (9)$$

is the solution. All three expressions (7), (8) and (9) give the same solution, a proof of which is left to the reader, as Exercise 8.

**Example.** For equations

$$\begin{bmatrix} 3 & 5 \\ 2 & 7 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 13 \\ 16 \\ 19 \end{bmatrix}$$

(7) gives the solution as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 2 & 7 \end{bmatrix}^{-1} \begin{bmatrix} 13 \\ 16 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 7 & -5 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 13 \\ 16 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

An example of (8) is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 29 & -49 & 22 \\ -13 & 25 & -11 \end{bmatrix} \begin{bmatrix} 13 \\ 16 \\ 19 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and (9) is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 14 & 38 \\ 38 & 155 \end{bmatrix}^{-1} \begin{bmatrix} 90 \\ 348 \end{bmatrix} = \frac{1}{726} \begin{bmatrix} 155 & -38 \\ -38 & 14 \end{bmatrix} \begin{bmatrix} 90 \\ 348 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Notice that  $A$  being square and nonsingular is a particular case of  $A$  having full column rank, in which case all three of (7), (8) and (9) reduce to  $x = A^{-1}y$  as one would expect. With this in mind we can say that whenever  $A$  has full column rank, consistent equations  $Ax = y$  have just one solution; otherwise they have infinitely many solutions. If the equations are inconsistent they have no solution.

#### 4. DERIVING SOLUTIONS USING GENERALIZED INVERSES

When there are many solutions to  $Ax = y$  we make use of generalized inverse matrices. This occurs when  $A$  has less than full column rank, i.e.,  $A$  square and singular, or  $A$  rectangular and rank less than the number of columns. The case of  $A$  having full row rank does not need to be treated separately.

##### a. Obtaining a solution

The relationship between a generalized inverse of  $A$  and consistent equations  $Ax = y$  is set out in the following theorem adapted from Rao (1962).

**Theorem 3.** The consistent equations  $Ax = y$  for  $y \neq 0$  have a solution  $x = Gy$  if, and only if,  $AGA = A$ .

**Proof.** Write  $a_j$  as the  $j$ th column of  $A$ . Consider the equations  $Ax = a_j$ . They have a solution: the elementary vector (see Section 3.1e)  $e_j$ . Therefore the equations  $Ax = a_j$  are consistent (by Theorem 1). Consequently, if the equations  $Ax = y$  are consistent with a solution  $x = Gy$ , the consistent equations  $Ax = a_j$  have a solution  $x = Ga_j$ . Substitution into  $Ax = a_j$  gives  $AGA_j = a_j$ . This is true for all  $j$ , i.e., for all columns of  $A$ . Hence  $AGA = A$ .

Conversely, if  $AGA = A$  then  $AGAx = Ax$  and with  $Ax = y$  this means  $AGy = y$ , i.e.,  $A(Gy) = y$ . Hence  $x = Gy$  is a solution of  $Ax = y$ . QED

**Example.** For equations

$$Ax = y \equiv \begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \end{bmatrix} x = \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix} \quad (10)$$

a generalized inverse of  $A$  is, from the algorithm of Section 8.4a,

$$G = \begin{bmatrix} -1 & 3 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad (11)$$

hence

$$\bar{x} = Gy = \begin{bmatrix} -1 & 3 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} \quad (12)$$

is a solution, as is easily verified by direct substitution.

The important part of Theorem 3 insofar as applications are concerned is the sufficiency part of it, that if  $AGA = A$ , then  $x = Gy$  is a solution. In this respect Theorem 3 indicates how a solution of consistent equations  $Ax = y$  can be obtained: find a generalized inverse  $G$  of  $A$  and  $x = Gy$  is a solution. (The symbol  $\bar{x}$  distinguishes a solution from the vector of unknowns,  $x$ .)

An important subtlety exists in the proof of the necessity part of Theorem 3, that if  $x = Gy$  is a solution then  $AGA = A$ . This is to be understood as meaning that if  $x = Gy$  is a solution for *all* vectors  $y$  for which  $Ax = y$  are consistent, then  $AGA = A$ . Insistence on "all vectors  $y$ " is needed because in particular cases there will be matrices  $K$  such that a solution can be written as  $x = Ky$  but without  $K$  being a generalized inverse of  $A$ . For example,

$$I_3 x = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} \text{ has solution } x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

but

$$K = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is not a generalized inverse of  $I_3$ .

Note, in passing, the useful result

$$AGy = y \quad (13)$$

which occurs in the proof of the theorem, and which can also be derived directly (by premultiplying  $Ax = y$  by  $AG$ ). A similar kind of result for  $\bar{x}$  being a solution to  $Ax = y$  is

$$GA\bar{x} = Gy. \quad (14)$$

This comes directly from premultiplying  $A\bar{x} = y$  by  $G$ .

### b. Obtaining many solutions

When  $Ax = y$  has many solutions the following theorem provides a method for deriving them from the one given by Theorem 3.

**Theorem 4.** When  $A$  has  $q$  columns and  $G$  is a generalized inverse of  $A$ , the consistent equations  $Ax = y$  have solutions

$$\bar{x} = Gy + (GA - I)z \quad (15)$$

for any arbitrary vector  $z$  of order  $q$ .

**Proof.** For  $\bar{x} = Gy + (GA - I)z$ ,

$$A\bar{x} = AGy + A(GA - I)z = y + (AGA - A)z \quad \text{because of (13)}$$

$$= y \quad \text{because } AGA = A.$$

QED

**Example (continued).** From (10) and (11)

$$H = GA = \begin{bmatrix} -1 & 3 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (16)$$

and so (15) is

$$\begin{aligned} \bar{x} &= \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} + \left\{ \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} \\ &= \begin{bmatrix} 4 + 2z_3 + 3z_4 \\ 2 - z_3 - z_4 \\ -z_3 \\ -z_4 \end{bmatrix} \end{aligned} \quad (17)$$

for any values of  $z_3$  and  $z_4$ . The reader can verify by substitution that (17) does satisfy (10).

It is sometimes convenient to use the notation

$$H \equiv GA \quad \text{and} \quad \bar{x} = Gy + (H - I)z. \quad (18)$$

Three useful properties of  $H$  are: (i)  $H$  is idempotent, (ii)  $r_H = r_A$  and (iii) for  $A$  having  $q$  columns,  $r(H - I) = r(I - H) = q - r_A$ . Property (i) is obvious from consideration of  $H^2$ , and (ii) and (iii) are proven in Lemmas 3 and 7 of Section 7.5.

### c. All possible solutions

Having shown that (15), or equivalently (18), generates many solutions for a specific  $G$ , the question immediately arises as to whether all possible

solutions can be so generated—for that specific generalized inverse  $\mathbf{G}$ . They can. The following theorem states this formally.

**Theorem 5.** All possible solutions of consistent equations  $\mathbf{Ax} = \mathbf{y}$  can be generated from  $\bar{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z}$  for any specific  $\mathbf{G}$  by using all possible values of the arbitrary vector  $\mathbf{z}$ .

**Proof.** Suppose  $\bar{\mathbf{x}}$  is a solution for  $\mathbf{x}$  to  $\mathbf{Ax} = \mathbf{y}$ . Choose  $\mathbf{z} = -\bar{\mathbf{x}}$ . Then  $\bar{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z} = \mathbf{Gy} - (\mathbf{GA} - \mathbf{I})\bar{\mathbf{x}}$  reduces to  $\bar{\mathbf{x}}$ . Hence by appropriate choice of  $\mathbf{z}$ , any solution  $\bar{\mathbf{x}}$  can be put in the form of  $\bar{\mathbf{x}}$ ; i.e., all solutions can be generated by  $\bar{\mathbf{x}}$ . QED

The importance of this theorem is that one need derive only one generalized inverse  $\mathbf{G}$  of  $\mathbf{A}$  in order to generate all solutions to  $\mathbf{Ax} = \mathbf{y}$ . There are no solutions other than those that can be generated from  $\bar{\mathbf{x}}$  of (15) or equivalently (18).

The generation of all possible solutions by means of  $\bar{\mathbf{x}} = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}$  is tied to the arbitrariness of  $\mathbf{z}$ . The arbitrariness in  $\mathbf{G}$ , as demonstrated in Section 8.5, can also lead to generation of all solutions, as indicated by the next theorem.

**Theorem 6.** All solutions of  $\mathbf{Ax} = \mathbf{y}$  for  $\mathbf{y} \neq \mathbf{0}$  can be generated from  $\bar{\mathbf{x}} = \mathbf{Gy}$  by using all possible generalized inverses  $\mathbf{G}$  of  $\mathbf{A}$ .

Proof of this theorem is based on the following lemma.

**Lemma.** For  $\mathbf{z}_{q \times 1}$  arbitrary and  $\mathbf{y}_{p \times 1}$  known and non-null there exists an arbitrary matrix  $\mathbf{X}$  such that  $\mathbf{z} = \mathbf{Xy}$ .

**Proof of lemma.** Since  $\mathbf{y} \neq \mathbf{0}$  one element,  $y_k$  say, is nonzero. Write  $\mathbf{z} = \{z_i\}$  and  $\mathbf{X} = \{x_{ij}\}$  with  $i = 1, \dots, q$  and  $j = 1, \dots, p$ ; and let  $x_{ij} = z_i/y_k$  for  $j = k$ , and  $x_{ij} = 0$  otherwise. Then  $\mathbf{Xy} = \mathbf{z}$  and  $\mathbf{X}$  is arbitrary. QED

**Example of lemma.** With  $q = 2$  and  $p = 3$

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2}z_1 & 0 \\ 0 & \frac{1}{2}z_2 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \\ 5 \end{bmatrix}.$$

**Proof of Theorem 6.** For a generalized inverse  $\mathbf{G}$  of  $\mathbf{A}$ , a solution to  $\mathbf{Ax} = \mathbf{y}$  is  $\mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z}$  for  $\mathbf{z}$  arbitrary. Letting  $\mathbf{z} = -\mathbf{Xy}$  for  $\mathbf{X}$  arbitrary, as in the lemma, this solution becomes

$$\mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z} = [\mathbf{GAG} + (\mathbf{I} - \mathbf{GA})\mathbf{X} + (-\mathbf{G})(\mathbf{AG} - \mathbf{I})]\mathbf{y} = \mathbf{G}^*\mathbf{y}$$

for  $\mathbf{G}^*$  of Section 8.5; i.e., any solution can be put in the form  $\mathbf{G}^*\mathbf{y}$ . QED

Relationships between the different solutions are considered in Exercises 15 and 16.

#### d. Combinations of solutions

Theorems 3–6 indicate how to obtain solutions of  $\mathbf{Ax} = \mathbf{y}$  using  $\mathbf{y}$  and generalized inverses of  $\mathbf{A}$ . Further solutions can also be obtained in the form of linear combinations of those already derived, as given by the next theorem.

**Theorem 7.** When  $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_t$  are any  $t$  solutions to consistent equations  $\mathbf{Ax} = \mathbf{y}$  for  $\mathbf{y} \neq \mathbf{0}$ , then  $\sum_{i=1}^t \lambda_i \bar{\mathbf{x}}_i$  is a solution if and only if  $\sum_{i=1}^t \lambda_i = 1$ .

**Proof.** Define  $\bar{\mathbf{x}} = \sum \lambda_i \bar{\mathbf{x}}_i$  where, throughout this proof, summation is over  $i = 1, 2, \dots, t$ ; and observe that

$$\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}\sum \lambda_i \bar{\mathbf{x}}_i = \sum \lambda_i \mathbf{A}\bar{\mathbf{x}}_i = \sum \lambda_i \mathbf{y} = \mathbf{y}\sum \lambda_i.$$

Then if  $\bar{\mathbf{x}}$  is a solution

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} = \mathbf{y}\sum \lambda_i, \quad \text{so implying } \sum \lambda_i = 1;$$

and if  $\sum \lambda_i = 1$  then

$$\mathbf{A}\bar{\mathbf{x}} = \mathbf{y}, \quad \text{implying that } \bar{\mathbf{x}} \text{ is a solution.} \quad \text{QED}$$

**Example (continued).** Three different solutions to (10) are, from (17),

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 0 \\ 4 \\ 2 \\ 0 \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} -3 \\ 4 \\ -1 \\ 3 \end{bmatrix}. \quad (19)$$

From these it will be found that both

$$\bar{\mathbf{x}}_4 = \frac{1}{2}\bar{\mathbf{x}}_1 + \frac{1}{4}\bar{\mathbf{x}}_2 + \frac{1}{4}\bar{\mathbf{x}}_3 = \begin{bmatrix} -\frac{1}{4} \\ 3\frac{1}{2} \\ \frac{1}{4} \\ 1\frac{1}{4} \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{x}}_5 = 4\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - 2\bar{\mathbf{x}}_3 = \begin{bmatrix} 10 \\ 0 \\ 0 \\ -2 \end{bmatrix}$$

are also solutions.

Simply stated, the theorem says that a linear combination of solutions is itself a solution if (and only if) the coefficients in the linear combination sum to unity.

## 5. LINEARLY INDEPENDENT SOLUTIONS

The solutions to  $\mathbf{Ax} = \mathbf{y}$  for  $\mathbf{A}$  of order  $p \times q$  are vectors of order  $q$ . Therefore no more than  $q$  of them can be LIN (see Section 6.5c). However,

in view of the relationships among solutions that have just been discussed and the related methods of deriving solutions there are, in fact, usually fewer than  $q$  LIN solutions as the following theorem shows.

**Theorem 8.** For  $\mathbf{A}$  having  $q$  columns, consistent equations  $\mathbf{Ax} = \mathbf{y}$  for  $\mathbf{y} \neq \mathbf{0}$  have  $q - r_{\mathbf{A}} + 1$  LIN solutions.

**Proof.** From (18), solutions are  $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}$  for  $\mathbf{H} = \mathbf{GA}$  and arbitrary  $\mathbf{z}$ . From (iii) following (18),  $r(\mathbf{H} - \mathbf{I}) = q - r_{\mathbf{A}}$  and so from Lemma 2 of Section 7.5, the maximum number of LIN vectors  $(\mathbf{H} - \mathbf{I})\mathbf{z}$  for arbitrary  $\mathbf{z}$  is  $q - r_{\mathbf{A}}$ . To represent such a set of vectors, suppose that

$$(\mathbf{H} - \mathbf{I})\mathbf{z}_i \text{ for } i = 1, 2, \dots, q - r_{\mathbf{A}} \text{ are LIN.} \quad (20)$$

We will show, using (20), that

- (i)  $\tilde{\mathbf{x}}_i = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i$  are  $q - r_{\mathbf{A}}$  LIN solutions,
- (ii)  $\mathbf{Gy}$  is LIN of the  $\tilde{\mathbf{x}}_i$ ,
- (iii) any other solution is a linear combination of  $\mathbf{Gy}$  and the  $\tilde{\mathbf{x}}_i$ .

Assume that  $\tilde{\mathbf{x}}_i$  of (i) are not LIN, so that

$$\sum \lambda_i [\mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i] = \mathbf{0}$$

is true for some  $\lambda$ 's not all zero, and where, throughout this proof, summation is over  $i = 1, 2, \dots, q - r_{\mathbf{A}}$ . Then

$$\mathbf{Gy} = -\sum \lambda_i (\mathbf{H} - \mathbf{I})\mathbf{z}_i / \sum \lambda_i.$$

Premultiplying this by  $\mathbf{A}$  and using (13) gives  $\mathbf{y} = \mathbf{0}$ , which is explicitly excluded by the theorem. Hence the assumption is false and so (i) is proved.

To prove (ii) assume  $\mathbf{Gy}$  is not LIN of the  $\tilde{\mathbf{x}}_i$ , so that

$$\mathbf{Gy} = \sum k_i [\mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i] = \sum k_i \mathbf{Gy} + \sum k_i (\mathbf{H} - \mathbf{I})\mathbf{z}_i \quad (21)$$

for some  $k$ 's not all zero. Since (21) is to be true identically, for all possible sets of  $\mathbf{z}_i$ 's, and the left-hand side  $\mathbf{Gy}$  contains no  $\mathbf{z}_i$ 's, the right-hand side must be likewise. Hence

$$\sum k_i (\mathbf{H} - \mathbf{I})\mathbf{z}_i = \mathbf{0}. \quad (22)$$

But the  $(\mathbf{H} - \mathbf{I})\mathbf{z}_i$  are LIN and so the only way for (22) to be true is for all the  $k$ 's to be zero—which contradicts the assumption. Thus (ii) is proven.

Let  $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}$  be a solution different from the  $\tilde{\mathbf{x}}_i$  with  $\mathbf{Gy}$  of (i) and (ii). By the definition of the  $(\mathbf{H} - \mathbf{I})\mathbf{z}_i$  in (20),  $(\mathbf{H} - \mathbf{I})\mathbf{z}$  is a linear combination of those vectors in (20); i.e.,

$$(\mathbf{H} - \mathbf{I})\mathbf{z} = \sum k_i (\mathbf{H} - \mathbf{I})\mathbf{z}_i \quad (23)$$

for some  $k$ 's not all zero. Then

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z} \\ &= \mathbf{Gy} + \sum k_i (\mathbf{H} - \mathbf{I})\mathbf{z}_i \\ &= \sum k_i [\mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i] + (1 - \sum k_i) \mathbf{Gy}, \end{aligned} \quad (24)$$

which is a linear combination of the  $\tilde{\mathbf{x}}_i$  and  $\mathbf{Gy}$  of (i) and (ii). So (iii) is proved.

The consequence of (i), (ii) and (iii) is that for  $(\mathbf{H} - \mathbf{I})\mathbf{z}_i$ , with  $i = 1, 2, \dots, q - r_{\mathbf{A}}$  being LIN,  $\mathbf{Gy}$  and  $\tilde{\mathbf{x}}_i = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i$  form a set of  $q - r_{\mathbf{A}} + 1$  LIN solutions such that every other solution is a linear combination of them. QED

The proof of the theorem shows not only that there are just  $q - r_{\mathbf{A}} + 1$  LIN solutions but also that one possible set of such solutions is

$$\mathbf{Gy} \text{ and } \tilde{\mathbf{x}}_i = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i \quad (25)$$

for  $\mathbf{z}_i$  arbitrary but chosen so that

$$(\mathbf{H} - \mathbf{I})\mathbf{z}_i \text{ for } i = 1, 2, \dots, q - r_{\mathbf{A}} \text{ are LIN.} \quad (26)$$

**Example (continued).** Taking solutions from (12) and (19),

$$\mathbf{Gy} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix} \text{ and } \tilde{\mathbf{x}}_2 = \begin{bmatrix} 0 \\ 4 \\ 2 \\ 0 \end{bmatrix} \quad (27)$$

are a set of  $q - r_{\mathbf{A}} + 1 = 4 - 2 + 1 = 3$  LIN solutions illustrating (25). The other solution in (19) can be expressed as a linear combination of these three:

$$\begin{bmatrix} -3 \\ 4 \\ -1 \\ 3 \end{bmatrix} = -1\frac{1}{2} \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 0 \\ 4 \\ 2 \\ 0 \end{bmatrix}.$$

Notice that the sum of the coefficients in this linear combination is unity,  $-1\frac{1}{2} + 3 - \frac{1}{2} = 1$ , in keeping with Theorem 8. And, of course, every such combination of the vectors in (27) is a solution.

Although (25) is an easy set of  $q - r_{\mathbf{A}} + 1$  LIN solutions to derive, (24) of the proof shows the possibility of obtaining such solutions without one of them being  $\mathbf{Gy}$ . This is done by using for the vector  $(\mathbf{H} - \mathbf{I})\mathbf{z}$  in  $\mathbf{x} = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}$  not only a

set of  $q - r_A$  LIN vectors  $(\mathbf{H} - \mathbf{I})\mathbf{z}_i$ , but also a linear combination of them,  $(\mathbf{H} - \mathbf{I})\mathbf{z}$  as in (23).

**Example.** An example of this is  $\bar{\mathbf{x}}_3$  of (19). The LIN vectors  $(\mathbf{H} - \mathbf{I})\mathbf{z}_i$  corresponding to  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$  of (19) are

$$(\mathbf{H} - \mathbf{I})\mathbf{z}_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad (\mathbf{H} - \mathbf{I})\mathbf{z}_2 = \begin{bmatrix} -4 \\ 2 \\ 2 \\ 0 \end{bmatrix}.$$

Then for (23) with  $k_1 = 3$  and  $k_2 = -\frac{1}{2}$

$$(\mathbf{H} - \mathbf{I})\mathbf{z} = 3 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -4 \\ 2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -7 \\ 2 \\ -1 \\ 3 \end{bmatrix}$$

giving

$$\bar{\mathbf{x}} = \mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -7 \\ 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 4 \\ -1 \\ 3 \end{bmatrix}$$

which is  $\bar{\mathbf{x}}_3$  of (19). Thus  $\bar{\mathbf{x}}_1$ ,  $\bar{\mathbf{x}}_2$  and  $\bar{\mathbf{x}}_3$  are 3 ( $= q - r_A + 1$ ) LIN solutions without one of them being  $\mathbf{G}\mathbf{y}$  as in (25). Nevertheless, (25) is usually the easiest to use.

## 6. AN INVARIANCE PROPERTY

Our final theorem on solving linear equations concerns the invariance of certain linear combinations of elements of a solution. The result has important applications in estimation procedures in analysis of variance situations in statistics. We begin by noting some features of the example.

**Example (continued).** For every solution  $\bar{\mathbf{x}}$  to (10) the linear combination of elements of the solution represented by the inner product  $[3 \ 2 \ 4 \ 7]\bar{\mathbf{x}}$  has the value 16. Thus on defining  $\mathbf{k}'_1 = [3 \ 2 \ 4 \ 7]$  and using the solution  $\mathbf{G}\mathbf{y}$  of (12)

$$\mathbf{k}'_1\mathbf{G}\mathbf{y} = [3 \ 2 \ 4 \ 7] \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} = 16. \quad (28)$$

Similarly for  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$  of (19)

$$\mathbf{k}'_1\bar{\mathbf{x}}_1 = [3 \ 2 \ 4 \ 7] \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix} = 16 \quad \text{and} \quad \mathbf{k}'_1\bar{\mathbf{x}}_2 = [3 \ 2 \ 4 \ 7] \begin{bmatrix} 0 \\ 4 \\ 2 \\ 1 \end{bmatrix} = 16. \quad (29)$$

The same is true of  $\mathbf{k}'_1\bar{\mathbf{x}}_3$  for  $\bar{\mathbf{x}}_3$  of (19), and of  $\mathbf{k}'_1\bar{\mathbf{x}}_4$  and  $\mathbf{k}'_1\bar{\mathbf{x}}_5$  for solutions  $\bar{\mathbf{x}}_4$  and  $\bar{\mathbf{x}}_5$  given below (19): their values are 16. And for every solution  $\bar{\mathbf{x}}$  of (10) the value of  $\mathbf{k}'_1\bar{\mathbf{x}}$  is 16.

Similar results hold for  $\mathbf{k}'_2 = [10 \ 1 \ 19 \ 29]$ : the value of  $\mathbf{k}'_2\bar{\mathbf{x}}$  is 42 for every solution  $\bar{\mathbf{x}}$  to (10). For example,

$$\mathbf{k}'_2\bar{\mathbf{x}}_4 = [10 \ 1 \ 19 \ 29] \begin{bmatrix} -\frac{1}{4} \\ 3\frac{1}{2} \\ \frac{1}{4} \\ 1\frac{1}{3} \end{bmatrix} = 42. \quad (30)$$

The invariance property suggested by these examples is that for certain  $\mathbf{k}'$ 's the value of  $\mathbf{k}'\bar{\mathbf{x}}$  is invariant to whatever solution  $\bar{\mathbf{x}}$  is used. This is so. First note a feature of this invariance: for  $\mathbf{k}'_2$  different from  $\mathbf{k}'_1$  the value of  $\mathbf{k}'_2\bar{\mathbf{x}}$  is different from  $\mathbf{k}'_1\bar{\mathbf{x}}$ . The value of  $\mathbf{k}'_1\bar{\mathbf{x}}$  is invariant to whatever solution is used for  $\bar{\mathbf{x}}$ , and the value of  $\mathbf{k}'_2\bar{\mathbf{x}}$ , although different from  $\mathbf{k}'_1\bar{\mathbf{x}}$ , is also invariant to  $\bar{\mathbf{x}}$ . It is only for certain vectors  $\mathbf{k}'$  that this invariance property exists. They are defined by Theorem 9.

**Theorem 9.** The value of  $\mathbf{k}'\bar{\mathbf{x}}$  is the same for all solutions  $\bar{\mathbf{x}}$  to  $\mathbf{A}\mathbf{x} = \mathbf{y}$  if and only if  $\mathbf{k}' = \mathbf{k}'\mathbf{H}$ .

**Proof.** From Theorem 4, any solution is  $\bar{\mathbf{x}} = \mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}$  for some  $\mathbf{G}$  and arbitrary  $\mathbf{z}$ . Therefore if  $\mathbf{k}' = \mathbf{k}'\mathbf{H}$  we have  $\mathbf{k}'\bar{\mathbf{x}} = \mathbf{k}'\mathbf{H}\mathbf{G}\mathbf{y} + \mathbf{k}'\mathbf{H}(\mathbf{H} - \mathbf{I})\mathbf{z} = \mathbf{k}'\mathbf{G}\mathbf{A}\mathbf{G}\mathbf{y}$ , because  $\mathbf{H}$  is idempotent (Section 4b), and on using (13) we get  $\mathbf{k}'\bar{\mathbf{x}} = \mathbf{k}'\mathbf{G}\mathbf{y}$ . This result holds even for a solution  $\bar{\mathbf{x}}$  based on a generalized inverse different from  $\mathbf{G}$  because from Theorems 7 and 8 we know that  $\bar{\mathbf{x}}$  is a linear combination  $\sum \lambda_i \bar{\mathbf{x}}_i$  of  $q - r_A + 1$  LIN solutions  $\bar{\mathbf{x}}_i = \mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i$  with  $\sum \lambda_i = 1$ . In that case

$$\mathbf{k}'\bar{\mathbf{x}} = \mathbf{k}'\sum \lambda_i \bar{\mathbf{x}}_i = \sum \lambda_i \mathbf{k}'\bar{\mathbf{x}}_i = \sum \lambda_i \mathbf{k}'\mathbf{G}\mathbf{y} = \mathbf{k}'\mathbf{G}\mathbf{y}.$$

Conversely, if  $\mathbf{k}'\bar{\mathbf{x}}$  is to be the same for all solutions  $\bar{\mathbf{x}}$  it must equal  $\mathbf{k}'\mathbf{G}\mathbf{y}$  because  $\mathbf{G}\mathbf{y}$  is one solution. Then for any solution  $\bar{\mathbf{x}} = \mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}$  we have

$$\mathbf{k}'\bar{\mathbf{x}} = \mathbf{k}'[\mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}] = \mathbf{k}'\mathbf{G}\mathbf{y} + \mathbf{k}'(\mathbf{H} - \mathbf{I})\mathbf{z}$$

true identically for all  $\mathbf{z}$ 's. Therefore  $\mathbf{k}'(\mathbf{H} - \mathbf{I})\mathbf{z} = \mathbf{0}$  for all  $\mathbf{z}$ 's and so  $\mathbf{k}'(\mathbf{H} - \mathbf{I}) = \mathbf{0}$ ; i.e.,  $\mathbf{k}' = \mathbf{k}'\mathbf{H}$ . QED

The important part of this theorem is the sufficiency part, that for any  $k'$  satisfying  $k' = k'H$  the value of  $k'\bar{x}$  is  $k'Gy$  no matter what solution is used for  $\bar{x}$ .

**Example** (continued). The value  $k'\bar{x} = 16$  obtained in (28) and (29) is, as in (28),  $k'_1Gy$ . And in (30) the value 42 is  $k'_2Gy$ :

$$k'_2Gy = [10 \quad 1 \quad 19 \quad 29] \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} = 42.$$

There are two useful corollaries to Theorem 9.

**Corollary 1.**  $k'\bar{x}$  is invariant to  $\bar{x}$  when  $k' = w'H$  for any  $w$ .

This is so because if  $k' = w'H$ , then  $k'H = w'H^2 = w'H = k'$  and the theorem is satisfied. Furthermore, for  $k' = w'H$  the value of  $k'\bar{x} = k'Gy$  is  $w'HGy = w'GAGy = w'Gy$ , after making use of (13).

**Corollary 2.** There are only  $r_A$  LIN vectors  $k'$  satisfying  $k' = k'H$ .

This is because  $r_H = r_A$  as at the end of Section 4b, and then by Lemma 2 of Section 7.5 there are only  $r_A$  LIN vectors  $k = Hw$  for arbitrary  $w$ .

The value of Corollary 1 is that it permits vectors  $k'$  for which  $k'\bar{x}$  is invariant to  $\bar{x}$  to be written in arbitrary form  $w'H$  with the corresponding value of  $k'\bar{x}$  involving the same arbitrary elements:  $k'\bar{x} = w'Gy$ . For example, with  $H$  of (16) the expression  $k' = w'H$  is

$$k' = [w_1 \quad w_2 \quad w_3 \quad w_4] \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ = [w_1 \quad w_2 \quad 2w_1 - w_2 \quad 3w_1 - w_2]. \quad (31)$$

And then, using  $Gy$  from (12)

$$k'\bar{x} = w'Gy = [w_1 \quad w_2 \quad w_3 \quad w_4] \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} = 4w_1 + 2w_2. \quad (32)$$

This means that for any values of  $w_1$  and  $w_2$  we have  $k'$  of (31) such that  $k'\bar{x}$  of (32) is invariant to  $\bar{x}$ , and  $k'\bar{x}$  is as given by (32). Illustration of this is available in (28) and (29) where for  $k'_1$  we have  $w_1 = 3$  and  $w_2 = 2$  so that (32) gives  $k'_1\bar{x} = 4(3) + 2(2) = 16$ ; and in (30) where for  $k'_2$  the  $w$ 's are  $w_1 = 10$  and  $w_2 = 1$  giving (32) as  $k'_2\bar{x} = 4(10) + 2(1) = 42$ .

7. EQUATIONS  $Ax = 0$ 

## a. General properties

Theorems 7 and 8, concerning linear combinations of solution vectors and the number of LIN solution vectors, exclude the case  $y = 0$ , i.e., they exclude the equation  $Ax = 0$ . It is therefore discussed separately.

First, by the definition of consistent equations given in Section 2, the equations  $Ax = 0$  are always consistent: no matter what linear relationships exist among the rows of  $A$  they also occur among the elements of  $0$ . Therefore from Theorem 1,  $Ax = 0$  can always be solved. One obvious solution is  $x = 0$ . And from Section 3 this is the only solution when  $A$  has full column rank (square and nonsingular, or rectangular).

The many non-null solutions when  $A$  has less than full column rank (square and singular, or rectangular) are, from (15),

$$\bar{x} = (GA - I)z \quad (33)$$

for arbitrary  $z$ . As seen at the start of the proof of Theorem 8 there are  $q - r_A$  such LIN vectors. Along with  $x = 0$  they constitute  $q - r_A + 1$  solutions but they cannot be described as  $q - r_A + 1$  LIN solutions because null vectors are excluded from the definition of LIN vectors [see Section 6.4b(iv)].

In Theorem 7 the condition  $\sum \lambda_i = 1$  on  $\sum \lambda_i \bar{x}_i$  being a solution is not needed for equations  $Ax = 0$ . This is because when  $A\bar{x}_i = 0$  then  $A(\sum \lambda_i \bar{x}_i) = \sum \lambda_i A\bar{x}_i = \sum \lambda_i 0 = 0$ . Hence any linear combination of any number of solutions of  $Ax = 0$  is also a solution.

Finally, since  $G0 = 0$ , application of Theorem 9 to  $Ax = 0$  shows that for  $k' = k'H$  the value of  $k'\bar{x}$  is 0 for all such  $k'$  and for all solutions  $\bar{x}$ . This is one case where the invariance property produces a unique value.

**Example.** With the same  $A$  as has been used in the equations in (10), consider

$$\begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \end{bmatrix} x = 0.$$

Non-null solutions are, from (16) and (33),

$$\bar{x} = (H - I)z = \begin{bmatrix} 0 & 0 & 2 & 3 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} z = \begin{bmatrix} 2z_3 + 3z_4 \\ -z_3 - z_4 \\ -z_3 \\ -z_4 \end{bmatrix}$$

so that two LIN solutions are

$$\bar{x}_1 = \begin{bmatrix} -5 \\ 2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \bar{x}_2 = \begin{bmatrix} -12 \\ 5 \\ 3 \\ 2 \end{bmatrix}.$$

Any linear combination of these is a solution, e.g.,

$$\bar{x}_3 = -3\bar{x}_1 + \bar{x}_2 = \begin{bmatrix} 3 \\ -1 \\ 0 \\ -1 \end{bmatrix};$$

and for  $k'_1 = [3 \ 2 \ 4 \ 7]$  and  $k'_2 = [10 \ 1 \ 19 \ 29]$  of (29) and (30) we find that  $k'_1\bar{x} = k'_2\bar{x} = 0$  for all of these solutions—and for all others.

#### b. Orthogonal solutions

For  $a'$  being any row of  $A$  and  $\bar{x}$  being a solution to  $Ax = 0$ , it is clear that  $a'\bar{x} = 0$ ; i.e.,  $\bar{x}$  is orthogonal to the rows of  $A$ . Therefore, by Exercise 6.10, the following lemma is true.

**Lemma 1.** Solutions to  $Ax = 0$  are orthogonal to and LIN of rows of  $A$ . This gives rise to a second lemma concerning the existence of solutions orthogonal to each other.

**Lemma 2.** Solutions to  $Ax = 0$  can always be found that are orthogonal to one another.

**Proof.** Suppose  $A$  has order  $p \times q$  and rank  $r$ . Then  $Ax = 0$  has  $q - r$  LIN solutions. Suppose  $x_1$  is a solution. Consider the equations in  $x$ :

$$\begin{bmatrix} A \\ x'_1 \end{bmatrix} x = 0. \quad (34)$$

Because  $x_1$  is a solution to  $Ax = 0$ , Lemma 1 indicates that  $x'_1$  and rows of  $A$  are LIN. Therefore (34) has  $q - r - 1$  LIN solutions. Suppose  $x_2$  is such a solution. Then clearly  $x_2$  satisfies  $Ax = 0$  and is also orthogonal to  $x_1$ . Similarly, if  $q - r > 2$  a third solution can be obtained by solving

$$\begin{bmatrix} A \\ x'_1 \\ x'_2 \end{bmatrix} x = 0$$

which has  $q - r - 2$  LIN solutions for  $x$ . This process can be continued until  $q - r$  solutions have been obtained, all orthogonal to each other.

**Example.** In the equations

$$\begin{bmatrix} 3 & 5 & 2 & -4 & -6 \\ 1 & 2 & -4 & -6 & 7 \\ 8 & 13 & 10 & -6 & -25 \end{bmatrix} x = 0, \quad (35)$$

the third row of the matrix is thrice the first minus the second, so that the matrix has rank 2; and there are  $5 - 2 = 3$  LIN solutions. One is

$$x'_1 = [1 \ 1 \ 1 \ 1 \ 1].$$

A second solution, orthogonal to  $x_1$ , is obtained by solving

$$\begin{bmatrix} 3 & 5 & 2 & -4 & -6 \\ 1 & 2 & -4 & -6 & 7 \\ 1 & 1 & 1 & 1 & 1 \\ 8 & 13 & 10 & -6 & -25 \end{bmatrix} x = 0.$$

This can be done using (33) in the form

$$x_2 = \left\{ \left[ \begin{pmatrix} 3 & 5 & 2 & -4 & -6 \\ 1 & 2 & -4 & -6 & 7 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right] \begin{bmatrix} 3 & 5 & 2 & -4 & -6 \\ 1 & 2 & -4 & -6 & 7 \\ 1 & 1 & 1 & 1 & 1 \\ 8 & 13 & 10 & -6 & -25 \end{bmatrix} - I_5 \right\} z. \quad (36)$$

For  $-\alpha$  and  $-\beta$  being the last two elements of  $z$ , this reduces after a little arithmetic to

$$x'_2 = [-(30\alpha + 81\beta)/9 \quad (28\alpha + 51\beta)/9 \quad (-7\alpha + 21\beta)/9 \quad \alpha \quad \beta].$$

Hence

$$x'_2 = [-19 \ 15 \ 0 \ 3 \ 1]$$

is one solution to (35) that is orthogonal to  $x_1$ . Finally, a third solution that



is orthogonal to both  $x_1$  and  $x_2$  is obtained by solving

$$\begin{bmatrix} 3 & 5 & 2 & -4 & -6 \\ 1 & 2 & -4 & -6 & 7 \\ 1 & 1 & 1 & 1 & 1 \\ -19 & 15 & 0 & 3 & 1 \\ 8 & 13 & 10 & -6 & -25 \end{bmatrix} \mathbf{x} = \mathbf{0}. \quad (37)$$

This can be done in the form

$$\mathbf{x}_3 = \left\{ \frac{1}{1017} \begin{bmatrix} -48 & 99 & 492 & -30 & 0 \\ -23 & 111 & 490 & 28 & 0 \\ 260 & -282 & -631 & -7 & 0 \\ -189 & 72 & 666 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 3 & 5 & 2 & -4 & -6 \\ 1 & 2 & -4 & -6 & 7 \\ 1 & 1 & 1 & 1 & 1 \\ -19 & 15 & 0 & 3 & 1 \\ 8 & 13 & 10 & -6 & -25 \end{bmatrix} - \mathbf{I}_5 \right\} \mathbf{z} \quad (38)$$

where the non-null  $4 \times 4$  matrix in the first part of this expression is the inverse of the principal leading submatrix of order 4 in (37). This solution reduces to

$$\mathbf{x}'_3 = \left[ \frac{1443}{1017} \quad \frac{1433}{1017} \quad \frac{-4172}{1017} \quad \frac{2313}{1017} \quad -1 \right] \gamma$$

for arbitrary  $\gamma$ , it being the last element of  $\mathbf{z}$  in (38). Hence one value for  $\mathbf{x}'_3$  is

$$\mathbf{x}'_3 = [1443 \quad 1433 \quad -4172 \quad 2313 \quad -1017].$$

It is easily checked that  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\mathbf{x}_3$  all satisfy (35) and are orthogonal to each other.

### c. Orthogonal vector spaces

Vector spaces are discussed in Section 6.11. The concept of orthogonality is now added to that of vector spaces. The row space of  $\mathbf{A}_{p \times q}$  of rank  $r$  has dimension  $r$ . A set of  $q - r$  orthogonal solutions to  $\mathbf{Ax} = \mathbf{0}$  spans a vector space of dimension  $q - r$  with every vector in it being orthogonal to every vector in the row space of  $\mathbf{A}$ .

The vector space spanned by the solutions (sometimes called the *solution space*) is therefore said to be the *orthogonal complement* of the row space of  $\mathbf{A}$ . This solution space is also called the null space of  $\mathbf{A}$ , as defined in Section 6.11.

## 8. A COMPLETE EXAMPLE

The example threaded through the preceding sections is somewhat spread out so we give here a complete example. Consider  $\mathbf{Ax} = \mathbf{y}$  as

$$\begin{bmatrix} 5 & 2 & -1 & 2 \\ 2 & 2 & 3 & 1 \\ 1 & 1 & 4 & -1 \\ 2 & -1 & -3 & -1 \\ 3 & 0 & 1 & -2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 7 \\ 9 \\ 5 \\ -6 \\ -1 \end{bmatrix}.$$

Row operations represented by

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -5 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ -1 & 2 & -1 & 1 & 0 \\ -1 & 2 & -2 & 0 & 1 \end{bmatrix} \quad \text{give} \quad \mathbf{PA} = \begin{bmatrix} 1 & 1 & 4 & -1 \\ 0 & -3 & -21 & 7 \\ 0 & 0 & -5 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

so that  $r_A = 3$ , and the last two elements of  $\mathbf{Py}$  are  $-7 + 18 - 5 - 6 = 0$  and  $-7 + 18 - 10 - 1 = 0$ . Hence by Theorem 2a the equations are consistent, and so from Theorem 1 they can be solved. A generalized inverse of  $\mathbf{A}$ , which can be obtained from either of Sections 8.2 or 8.4, is

$$\mathbf{G} = \frac{1}{15} \begin{bmatrix} 5 & -9 & 8 & 0 & 0 \\ -5 & 21 & -17 & 0 & 0 \\ 0 & -3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with

$$\mathbf{H} = \mathbf{GA} = \begin{bmatrix} 1 & 0 & 0 & -7/15 \\ 0 & 1 & 0 & 28/15 \\ 0 & 0 & 1 & -9/15 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (39)$$

Solutions given by (18) are therefore

$$\begin{aligned} \bar{x} &= \frac{1}{15} \begin{bmatrix} 5 & -9 & 8 & 0 & 0 \\ -5 & 21 & -17 & 0 & 0 \\ 0 & -3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 7 \\ 9 \\ 5 \\ -6 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & -7/15 \\ 0 & 0 & 0 & 28/15 \\ 0 & 0 & 0 & -9/15 \\ 0 & 0 & 0 & -1 \end{bmatrix} z \\ &= \frac{1}{15} \begin{bmatrix} -6 - 7z \\ 69 + 28z \\ 3 - 9z \\ -15z \end{bmatrix} \end{aligned} \quad (40)$$

where  $z$  is an arbitrary scalar.

Because  $q = 4$  and  $r_A = 3$ , there are  $q - r_A + 1 = 4 - 3 + 1 = 2$  LIN solutions: two possibilities are obtained by putting  $z = 0$  and  $z = -3$  in (40) giving

$$\bar{x}_1 = \begin{bmatrix} -0.4 \\ 4.6 \\ 0.2 \\ 0 \end{bmatrix} \quad \text{and} \quad \bar{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 3 \end{bmatrix},$$

respectively. Any other solution will be a linear combination of these two with coefficients summing to unity. For example,

$$\bar{x}_3 = 2\bar{x}_1 - \bar{x}_2 = \begin{bmatrix} -1.8 \\ 10.2 \\ -1.6 \\ -3.0 \end{bmatrix}$$

is a solution.

With an arbitrary vector  $w'$ , invariant values  $k'\bar{x}$  are available using (39) for  $k'$  of the form

$$k' = w'H = [w_1 \quad w_2 \quad w_3 \quad (-7w_1 + 28w_2 - 9w_3)/15]; \quad (41)$$

and from (40) the value of  $k'\bar{x}$  is then

$$k'\bar{x} = k'Gy = (-6w_1 + 69w_2 + 3w_3)/15. \quad (42)$$

There are  $r = 3$  LIN vectors of the form (41), three examples being  $k'_1 = [3 \ 0 \ 1 \ -2]$ ,  $k'_2 = [4 \ 1 \ 5 \ -3]$  and  $k'_3 = [1 \ 4 \ 0 \ 7]$ , the corresponding invariant values from (42) being  $k'_1\bar{x} = -1$ ,  $k'_2\bar{x} = 4$  and  $k'_3\bar{x} = 18$ .

Any other  $k'$  of the form (41) will be a linear combination of  $k'_1$ ,  $k'_2$  and  $k'_3$  and the corresponding value of  $k'\bar{x}$  will be the same linear combination of  $k'_1\bar{x}$ ,  $k'_2\bar{x}$  and  $k'_3\bar{x}$ . For example, for

$$k'_4 = [11 \ -1 \ 30 \ -25] = -5k'_1 + 7k'_2 - 2k'_3,$$

$$k'_4\bar{x} = -5(k'_1\bar{x}) - 7(k'_2\bar{x}) - 2(k'_3\bar{x}) = -5(-1) + 7(4) - 2(18) = -3.$$

The reader should verify for appropriate choice of  $w'$  that  $k'_4$  and  $k'_4\bar{x}$  satisfy (41) and (42), respectively.

The procedures described here may appear somewhat lengthy for the small examples chosen, but advantages become apparent when dealing with a large number of equations, say 20 or more. At all stages, the matrix operations are readily amenable to computer processing.

## 9. LEAST SQUARES EQUATIONS

Suppose we have a set of inconsistent equations  $Ax = y$ . Despite their inconsistency, we may seek an approximate solution and decide to use as that solution the value of  $x$  which minimizes the distance between  $y$  and  $Ax$ . Distance in this context is defined as the norm of the vector of differences,  $y - Ax$ . Hence we minimize

$$(y - Ax)'(y - Ax) = x'A'Ax - 2x'A'y + y'y,$$

which yields (see Section 12.8) equations  $A'A\bar{x} = A'y$ . The solution to these is then taken as the approximate solution to  $Ax = y$ , and it is considered best in the sense of yielding the minimum distance between  $y$  and  $Ax$ .

The method of least squares in statistics, applied to predicating the expected value of a vector of observations  $y$  being  $E(y) = Xb$  (where  $X$  is a known rectangular matrix), leads to exactly the same sort of equations, namely to

$$X'X\hat{b} = X'y. \quad (43)$$

These equations arise in many settings in statistics, and the theorems of this chapter are important in both solving the equations and using the solutions. Thus, although equations (43) occur in regression analysis where  $X'X$  is usually of full rank and so  $\hat{b} = (X'X)^{-1}X'y$ , they are also applicable more generally; for example, in experimental design, in analysis of variance and with survey data, where  $X'X$  is often not of full rank. Using a generalized

inverse of  $X'X$  then gives

$$\mathbf{b}^0 = (X'X)^{-1} X'y \quad (44)$$

as a solution. The notation  $\mathbf{b}^0$ , in contrast to the more familiar  $\hat{\mathbf{b}}$  of (43), is used to emphasize that  $\mathbf{b}^0$  for some particular  $(X'X)^{-1}$  is just one of many possible solutions and as such cannot be considered a unique estimator of  $\mathbf{b}$ . Nevertheless, many useful and practical features of the statistical analysis are invariant to the choice of  $(X'X)^{-1}$ , i.e., choice of solution  $\mathbf{b}^0$ , so that availability of many different solutions is not really bothersome.

First, the vector of predicted values is  $\hat{\mathbf{y}} = X\mathbf{b}^0 = X(X'X)^{-1} X'y$ , and by Theorem 9 this is invariant to  $(X'X)^{-1}$ , i.e., to  $\mathbf{b}^0$ . So also is the error sum of squares

$$\text{SSE} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = [\mathbf{y} - X(X'X)^{-1} X'y]'[\mathbf{y} - X(X'X)^{-1} X'y], \quad (45)$$

for the same reason. Further use of the theorem simplifies (45) to

$$\text{SSE} = \mathbf{y}'[\mathbf{I} - X(X'X)^{-1} X']\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{b}^0 X'y$$

where the final expression here is a very useful form:  $\mathbf{y}'\mathbf{y}$  is the total sum of squares of the observations in  $\mathbf{y}$ , and  $\mathbf{b}^0 X'y$  is the inner product of the vector of solutions with the vector of right-hand sides of (42). Finally, although  $\mathbf{b}^0$  is a biased estimator of  $\mathbf{b}$ , we know from Theorem 9 that  $\mathbf{k}'\mathbf{b}^0$  is invariant to  $\mathbf{b}^0$  for  $\mathbf{k}' = \mathbf{k}'(X'X)^{-1} X'X$ ; and it turns out that not only is  $\mathbf{k}'\mathbf{b}^0$  an estimator of  $\mathbf{k}'\mathbf{b}$  having desirable features (unbiasedness and minimum variance) but also that under these conditions  $\mathbf{k}'\mathbf{b}$  is often a linear combination of the parameters in  $\mathbf{b}$  that is of great practical use and interest. Hence an estimator with desirable features can be obtained for useful functions of the unknown parameters even when there are no unique estimators of the parameters themselves, i.e., no unique solution  $\mathbf{b}^0$ .

These and other similar results are explored in detail in Chapters 14 and 15, and they are dealt with extensively in Searle (1971). Since they are applicable to regression, to analysis of variance of designed experiments and of survey data and to covariance analysis, and also for hypothesis testing in all these situations, there is widespread application in statistics of results whose development benefits from using the generalized inverse.

#### EXERCISES

1. Find a set of LIN solutions to each of the following sets of equations; in each case also find one other solution and show it

relationship to the LIN solutions

$$(a) \begin{bmatrix} 6 & 2 & 0 \\ -1 & 0 & 3 \\ 3 & 2 & 9 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 8 \\ 2 \\ 14 \end{bmatrix} \quad (b) \begin{bmatrix} 1 & 0 & 2 \\ 4 & 1 & 7 \\ 3 & 2 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 12 \\ 46 \\ 27 \end{bmatrix}$$

$$(c) \begin{bmatrix} 1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 8 \\ -1 \\ 6 \\ 1 \end{bmatrix}$$

$$(d) \begin{bmatrix} 4 & -9 & -1 & 2 \\ 3 & 1 & 0 & 1 \\ 10 & -7 & -1 & 4 \\ 25 & -2 & -1 & 9 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 7 \\ 5 \\ 17 \\ 42 \end{bmatrix}$$

$$(e) \begin{bmatrix} 6 & 1 & 4 & 2 & 1 \\ 3 & 0 & 1 & 4 & 2 \\ -3 & -2 & -5 & 8 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 11 \\ 6 \\ -4 \end{bmatrix}$$

$$(f) \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 10 \\ 26 \end{bmatrix}$$

2. The third column in

$$\mathbf{M} = \begin{bmatrix} 1 & 6 & 8 \\ 4 & 17 & 12 \\ 7 & 28 & 16 \\ 4 & 10 & -8 \end{bmatrix}$$

is a linear combination of the other two columns. How many LIN solutions are there to  $\mathbf{M}\mathbf{x} = [19 \ 42 \ 65 \ 8]'$ ? Find such a set of solutions and from them derive one other solution.

3. Solve the two sets of equations

$$\begin{bmatrix} 2 & 4 & 3 & 5 \\ 1 & 9 & 7 & 0 \\ 4 & 22 & 17 & 5 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 21 \\ 33 \\ 87 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 4 & 3 & 2 \\ 2 & 9 & 7 & 1 \\ 5 & -3 & -2 & -5 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 8 \\ 31 \\ 7 \end{bmatrix}$$

using the same generalized inverse in both cases, and find a linear transformation connecting the general solutions of the two sets of equations.

4. Find a set of LIN solutions to

$$\frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

5. With

$$\begin{bmatrix} 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ -1\frac{1}{2} & 2\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 \end{bmatrix} = \mathbf{I} \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ -1\frac{1}{2} & 2\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 9 \\ -1 \\ 0 \end{bmatrix}$$

what are the equations to which

$$\tilde{\mathbf{x}} = \begin{bmatrix} \alpha \\ 9 - 6\alpha + 10\beta \\ -1 + \alpha - 5\beta \\ \beta \end{bmatrix}$$

is a solution? Give reasons for your answer.

6. Three different machines, A, B, and C, are to be used 8 hours a day in manufacturing four different products  $t, u, v$  and  $w$ . The number of hours that each machine must be used in making one unit of each product is shown in the accompanying table;

Machine	Product			
	$t$	$u$	$v$	$w$
A	1	2	1	2
B	7	0	2	0
C	1	0	0	4

i.e., to make one  $v$  requires 1 hour of machine A and 2 hours of machine B.

- In terms of the numbers of each product made in an 8-hour day, what equations represent full utilization of all machines?
- Solve the equations derived in (a).
- Negative solutions to the equations are meaningless, but fractional solutions have meaning. What conditions therefore apply to the solution obtained in (a)?

- How many of each product are made if production of  $t$  is maximized whilst maintaining full production of all machines?
  - What production of the four products maximizes production of  $w$ , subject to maintaining full utilization of all machines?
7. (a) Repeat Exercise 6 for full utilization of all machines over a 40-hour week rather than an 8-hour day. Confine solutions to being integers.  
 (b) Repeat (a) for machine C unavailable 8 hours a week.
8. (a) Use (23) of Chapter 8 to derive the general form of any left inverse of  $A$  of full column rank.  
 (b) Prove that  $(A'A)^{-1}A'y$  is then the only solution of  $Ax = y$ .  
 (c) Illustrate these results with the example in Section 3.
9. Find a common solution to the following equations:

$$\begin{bmatrix} 4 & 5 & 1 \\ -5 & -6 & 2 \\ -1 & -1 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 19 \\ -20 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 5 & 7 \\ -5 & -6 & 3 \\ 9 & 11 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 25 \\ -19 \\ 44 \end{bmatrix}$$

10. Provide a second proof of Corollary 2 of Theorem 9.
11. Find orthogonal solutions for the example in Section 7a.
12. When  $G$  and  $F$  are both generalized inverses of  $A$  and  $z = (G - F)y + (I - FA)w$ , show that the solution  $\tilde{\mathbf{x}} = G'y + (GA - I)z$  to  $Ax = y$  is identical to  $Fy + (FA - I)w$ .
13. When  $G$  and  $F$  are both generalized inverses of  $A$ , what is the simplest form of solution for  $w$  to
- $$(I - GA)w = (F - G)y + (FA - I)z?$$
14. When  $k' = k'GA$  for  $G$  a generalized inverse of  $A$ , show that  $k' = k'G^*A$  for  $G^*$  being any other generalized inverse of  $A$ .
15. When  $G$  and  $\hat{G}$  are two different generalized inverses of  $A$ , with corresponding solutions to  $Ax = y$  being  $\tilde{\mathbf{x}}$  and  $\hat{\mathbf{x}}$  respectively, as given by Theorem 4, what is the simplest relationship of  $\hat{z}$  to  $z$  that makes  $\hat{\mathbf{x}}$  equal to  $\tilde{\mathbf{x}}$ ?

16. Repeat Exercise 15 when  $G$  and  $\hat{G}$  are such that

$$G = \hat{G}\hat{A}\hat{G} + (I - \hat{G}\hat{A})T + S(I - \hat{A}\hat{G}).$$

17. Put solution (5) in the form of (15).  
 18. With  $K$  of (4) being real, prove that (4) reduces to (5). [Hint: first prove that  $I + K'K$  is p.s.d.]  
 19. For  $b^0$  being a solution to  $X'Xb^0 = X'y$ , prove that  $k'b^0$  is invariant to  $b^0$  if and only if  $k' = t'X$ .

## CHAPTER 10

## PARTITIONED MATRICES

The procedure of partitioning a matrix into submatrices is introduced in Section 2.10. The general specifications of a partitioned matrix are also given there, together with the operations of transposing and of multiplying partitioned matrices. This chapter deals with other operations on partitioned matrices.

## 1. ORTHOGONAL MATRICES

It is sometimes useful to partition an orthogonal matrix  $P$  as  $P = [A \ B]$ . Since orthogonality implies  $PP' = I$  and  $P'P = I$ ,

$$[A \ B] \begin{bmatrix} A' \\ B' \end{bmatrix} = I, \quad \text{i.e., } AA' + BB' = I, \quad (1)$$

and

$$\begin{bmatrix} A' \\ B' \end{bmatrix} [A \ B] = I, \quad \text{i.e., } \begin{bmatrix} A'A & A'B \\ B'A & B'B \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \quad (2)$$

Equation (2) means that

$$A'A = I, \quad A'B = 0, \quad B'A = 0 \quad \text{and} \quad B'B = I. \quad (3)$$

Although orthogonality of  $P$  leads to  $A'A = I$  and  $B'B = I$  in (3), note that neither  $A$  nor  $B$  is orthogonal because, with  $P$  being square, the partitioning  $P = [A \ B]$  means that both  $A$  and  $B$  are rectangular. ( $A'A = I$  implies orthogonality of  $A$  only if  $A$  is square; see Section 5.7d.)

A partitioning of  $P$  into four submatrices is the subject of Exercise 2.

## 2. DETERMINANTS

Consider the partitioned matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (4)$$

assuming that  $\mathbf{A}$  and  $\mathbf{D}$  are square and  $\mathbf{A}$  is nonsingular. [The identity (4) is easily verified by multiplication.] In Section 4.3e, (11) gives  $|\mathbf{KL}| = |\mathbf{K}||\mathbf{L}|$  for square  $\mathbf{K}$  and  $\mathbf{L}$  and (8), using  $|\mathbf{M}'| = |\mathbf{M}|$ , gives

$$\begin{vmatrix} \mathbf{R}' & \mathbf{X}' \\ \mathbf{0} & \mathbf{T}' \end{vmatrix} = \begin{vmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{X} & \mathbf{T} \end{vmatrix} = |\mathbf{R}||\mathbf{T}|,$$

so that by taking the determinant of both sides of (4) we have

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}|. \quad (5)$$

And with  $\mathbf{D}^{-1}$  existing it can be shown in similar fashion that

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{D}||\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}|. \quad (6)$$

Therefore, providing both  $\mathbf{A}$  and  $\mathbf{D}$  are nonsingular,

$$|\mathbf{A}||\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| = |\mathbf{D}||\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}|. \quad (7)$$

The reader should create and work some simple numerical examples to illustrate the validity of (5), (6) and (7).

**Illustration.** Exercise 3 suggests using (7) to find determinant expressions for the quadratic form  $\mathbf{x}'\mathbf{P}^{-1}\mathbf{x}$ . An interesting use of this quadratic form is available in Searle (1963) concerning the efficiency of ancestor records in animal selection. It is shown there that the correlation between an animal's true additive genetic merit and an estimate thereof based on  $n$  generations of ancestors takes the form  $r_n^2 = \mathbf{t}'_n \mathbf{A}_n^{-1} \mathbf{t}_n$  with

$$\mathbf{t}'_n = [p \quad p^2 \quad p^3 \quad \dots \quad p^n], \quad (8)$$

and

$$\mathbf{A}_n = \begin{bmatrix} kp & p^2 & p^3 & \dots & p^n \\ p^2 & kp^2 & p^3 & \dots & p^n \\ p^3 & p^3 & kp^3 & \dots & p^n \\ \vdots & \vdots & \vdots & \dots & \vdots \\ p^n & p^n & p^n & \dots & kp^n \end{bmatrix}, \quad (9)$$

and  $p = \frac{1}{2}$  and  $k = 1/h$ , where  $h$  is the genetic parameter known as heritability. By utilizing (7) and properties of inverse matrices, it is shown that

$$r_n^2 = \frac{2h-1}{2h} + \frac{(1-h)^2}{2h(1-hr_{n-1}^2)}. \quad (10)$$

The correlation  $r_n$  represents the efficiency of  $n$  generations of ancestor records, and from this result the relative efficiency of successive generations has been derived (Searle, 1963).

Suppose  $\mathbf{A} = \mathbf{A}' = (a_{ij})$  is positive definite. Then the first step in expanding  $|\mathbf{A}|$  by cofactors is

$$\begin{aligned} |\mathbf{A}| &= a_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{23} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \dots & \vdots \\ a_{2n} & a_{3n} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{12} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{13} & a_{23} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \dots & a_{nn} \end{vmatrix} \\ &= a_{11} |\mathbf{A}_{11}| + \begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A}_{11} \end{vmatrix} \end{aligned}$$

where  $|\mathbf{A}_{11}|$  is the cofactor of  $a_{11}$  in  $|\mathbf{A}|$ , and  $\mathbf{a}'$  is the first row of  $\mathbf{A}$  but with  $a_{11}$  omitted. Hence by (6),

$$\begin{aligned} |\mathbf{A}| &= a_{11} |\mathbf{A}_{11}| + |\mathbf{A}_{11}| (0 - \mathbf{a}' \mathbf{A}_{11}^{-1} \mathbf{a}) \\ &= a_{11} |\mathbf{A}_{11}| - (\mathbf{a}' \mathbf{A}_{11}^{-1} \mathbf{a}) |\mathbf{A}_{11}|. \end{aligned}$$

Now, because  $\mathbf{A}$  is p.d.,  $|A_{11}|$  is positive (Section 7.7a) and  $\mathbf{a}'\mathbf{A}_{11}^{-1}\mathbf{a}$  is non-negative [Exercise 5.15(i)]. Therefore  $|\mathbf{A}| \leq a_{11}|A_{11}|$ . Extension of this argument leads to the result that for  $\mathbf{A}$  being p.d. of order  $n$ , we have  $|\mathbf{A}| \leq a_{11} a_{22} \cdots a_{nn}$ .

**Illustration.** An application of this result to statistics is that the determinant of a correlation matrix (Section 13.1) is less than unity.

Application of the result to  $\mathbf{A} = \mathbf{B}\mathbf{B}'$  for  $\mathbf{B} = \{b_{ij}\}$  being any nonsingular matrix gives what is known as the Hadamard inequality:

$$|\mathbf{B}|^2 \leq \prod_{i=1}^n \left( \sum_{j=1}^n b_{ij}^2 \right).$$

### 3. INVERSES

Observe that for nonsingular  $\mathbf{R}$  and  $\mathbf{S}$ ,

$$\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{X} & \mathbf{S} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0} \\ -\mathbf{S}^{-1}\mathbf{X}\mathbf{R}^{-1} & \mathbf{S}^{-1} \end{bmatrix}, \quad (11)$$

as may be verified by multiplication. Using this and its transpose on (4) therefore gives

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix},$$

which can be rewritten as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{I} \end{bmatrix} (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} [-\mathbf{C}\mathbf{A}^{-1} \quad \mathbf{I}]. \quad (12)$$

Starting from

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix} \quad (13)$$

instead of (4), one can also derive

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\mathbf{D}^{-1}\mathbf{C} \end{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} [\mathbf{I} \quad -\mathbf{B}\mathbf{D}^{-1}], \quad (14)$$

analogous to (12). It is left to the reader (Exercise 16) to show explicitly that each submatrix of (14) equals its counterpart in (12). Results (12) and (14) have many uses, especially in statistics. For special forms of them, e.g., for symmetric matrices, see Exercise 7.

### 4. SCHUR COMPLEMENTS

The matrix

$$\mathbf{S} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \quad (15)$$

occurring in (7) and (12) is called the Schur complement of  $\mathbf{A}$  in

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad (16)$$

when  $\mathbf{A}$  is nonsingular. And for singular  $\mathbf{A}$ , the matrix  $\mathbf{D} - \mathbf{C}\mathbf{A}^{-}\mathbf{B}$  is said to be the generalized Schur complement relative to  $\mathbf{A}^{-}$ . Extensive properties of these complements are to be found in the literature, but this book is not the place for their development. Typical of results on rank, for example, is  $r_{\mathbf{Q}} = r_{\mathbf{A}} + r_{\mathbf{S}}$  for nonsingular  $\mathbf{A}$ ; and numerous results concerning inverses are of the form

$$(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}. \quad (17)$$

Readers interested in these and other extensions are referred to Marsaglia and Styan (1974a, b), especially for results on rank, to Henderson and Searle (1981a) for results on inverses and to Ouellette (1981) for general discussion of uses in statistics. All four papers have extensive references.

### 5. GENERALIZED INVERSES

The regular inverse of a partitioned matrix has just been dealt with. We now consider analogous results for generalized inverses, making great use of Marsaglia and Styan (1974b) in so doing. Not only does it contain useful results concerning rank conditions but it also has many important references.

Let  $\mathbf{A}^{-}$  be any particular generalized inverse of  $\mathbf{A}$ —in contrast to representing all generalized inverses. Marsaglia and Styan show that with

$$\mathbf{S} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-}\mathbf{B} \quad (18)$$

similar to (15), the matrix

$$Q = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \text{ has } Q^* = \begin{bmatrix} A^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^-B \\ I \end{bmatrix} S^- [-CA^- \quad I] \quad (19)$$

as a generalized inverse if and only if

$$r_Q = r_A + r_S; \quad (20)$$

i.e., if and only if  $r_Q = r_A + r_S$ . Proof of (20) is somewhat lengthy and beyond the scope of this book, but certain features of the result merit attention. First,  $Q^*$  of (19) is exactly analogous to the regular inverse of a nonsingular matrix given in (12). Second, the need for specifying  $A^-$  as being a particular generalized inverse of  $A$  is seen in the condition (20). For some generalized inverses of  $A$  it will be found that  $r_Q$  exceeds  $r_A + r_S$  and so (20) is not satisfied, whereupon  $Q^*$  is not a generalized inverse of  $Q$ ; but if, for other generalized inverses of  $A$ , (20) is satisfied, then  $Q^*$  is a generalized inverse of  $Q$ . Third, in contrast to  $A^-$ , the generalized inverse  $S^-$  in (19) can be any generalized inverse of  $S$ .

Corollaries of (19) and (20) are that  $Q^*$  is a reflexive generalized inverse if and only if  $A^-$  and  $S^-$  are reflexive; and, under certain conditions (including the need for  $A^-$  and  $S^-$  to be Moore-Penrose inverses),  $Q^*$  can be the Moore-Penrose inverse of  $Q$ . Details are to be found at (14) of Marsaglia and Styan (1974b).

**Example.** The rank of

$$Q = \begin{bmatrix} 2 & 2 & 1 & 2 & 2 \\ 4 & 4 & 1 & 3 & 9 \\ 8 & 7 & 1 & 7 & 13 \\ 12 & 11 & 10 & 22 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (21)$$

is 3: the first three rows are LIN and the last row is the sum of rows 2 and 3. Hence  $r_Q = 3$ . From (21)

$$A = \begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix} \text{ with } r_A = 1, \text{ and } A^- = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} \quad (22)$$

as a possible generalized inverse. Then it is easy to calculate

$$S = D - CA^-B = \begin{bmatrix} -1 & 5 \\ -2 & 10 \end{bmatrix}$$

so that  $r_S = 1$ . Hence  $r_Q = 3 \neq r_A + r_S = 1 + 1 = 2$  and so (20) is not satisfied, and (19) will not yield a generalized inverse of  $Q$  for the  $A^-$  of (22). However, using

$$A^- = \begin{bmatrix} \frac{1}{2} & 1 \\ -2 & 0 \end{bmatrix} \text{ gives } S = D - CA^-B = \begin{bmatrix} 3 & -39 \\ 6 & -54 \end{bmatrix} \quad (23)$$

for which  $r_S = 2$ . Hence  $r_Q = 3 = r_A + r_S = 1 + 2$  and so (20) is satisfied when  $A^-$  of (23) is used, and then (19) yields a generalized inverse of  $Q$ . The reader should check that this is so.

An important consequence of the result  $X(X'X)^-X'X = X$  given in Section 8.6c is, on applying (19) to

$$Q = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}, \quad (24)$$

that

$$Q^* = \begin{bmatrix} (X_1'X_1)^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X_1'X_1)^-X_1'X_2 \\ I \end{bmatrix} S^- [-X_2'X_1(X_1'X_1)^- \quad I]$$

$$\text{for } S = X_2'M_1X_2 \text{ with } M_1 = I - X_1(X_1'X_1)^-X_1'$$

and

$$Q^{**} = \begin{bmatrix} 0 & 0 \\ 0 & (X_2'X_2)^- \end{bmatrix} + \begin{bmatrix} I \\ -(X_2'X_2)^-X_2'X_1 \end{bmatrix} T^- [I \quad -X_1'X_2(X_2'X_2)^-]$$

$$\text{for } T = X_1'M_2X_1 \text{ with } M_2 = I - X_2(X_2'X_2)^-X_2'$$

are always generalized inverses of  $Q$ . This is so because  $Q$  in (24) is such that, by the use of  $X(X'X)^-X'X = X$ , the rank condition (20) for the existence of such partitioned generalized inverses is always satisfied. Marsaglia and Styan (1974b) show details.

Notice that  $Q^*$  and  $Q^{**}$  are generalized inverse forms corresponding to  $Q^{-1}$  for the nonsingular form of  $Q$  of (24) given in Exercise 7 and do, of course, reduce to  $Q^{-1}$  when  $Q$  is nonsingular.

## 6. DIRECT SUMS

Direct sums and direct products are matrix operations defined in terms of partitioned matrices. They are discussed in this and the next section.



The direct sum of two matrices  $\mathbf{A}$  and  $\mathbf{B}$  is defined as

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \quad (25)$$

and extends very simply to more than two matrices:

$$\mathbf{A} \oplus \mathbf{B} \oplus \mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{bmatrix}$$

and

$$\bigoplus_{i=1}^k \mathbf{A}_i = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}_k \end{bmatrix} = \text{diag}\{\mathbf{A}_i\} \text{ for } i = 1, \dots, k.$$

The definition (25) and its extensions apply whether or not the submatrices are of the same order; and all null matrices are of appropriate order. For example,

$$[1 \ 2 \ 3] \oplus \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 8 & 9 \end{bmatrix}. \quad (26)$$

Transposing a direct sum gives the direct sum of the transposes. The rank of a direct sum is the sum of the ranks, as is evident from the definition of rank (Section 6.8). It is clear from (25) that  $\mathbf{A} \oplus (-\mathbf{A}) \neq \mathbf{0}$  unless  $\mathbf{A}$  is null. Also,

$$(\mathbf{A} \oplus \mathbf{B}) + (\mathbf{C} \oplus \mathbf{D}) = (\mathbf{A} + \mathbf{C}) \oplus (\mathbf{B} + \mathbf{D})$$

only if the necessary conditions of conformability for addition are met. Similarly,

$$(\mathbf{A} \oplus \mathbf{B})(\mathbf{C} \oplus \mathbf{D}) = \mathbf{AC} \oplus \mathbf{BD}$$

provided that conformability for multiplication is satisfied. Should  $\mathbf{A}$  and  $\mathbf{B}$  be nonsingular, the preceding result leads to

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

The direct sum  $\mathbf{A} \oplus \mathbf{B}$  is square only if  $\mathbf{A}$  is  $p \times q$  and  $\mathbf{B}$  is  $q \times p$ . The determinant of  $\mathbf{A} \oplus \mathbf{B}$  equals  $|\mathbf{A}||\mathbf{B}|$  if both  $\mathbf{A}$  and  $\mathbf{B}$  are square, but otherwise it is zero or nonexistent.

## 7. DIRECT PRODUCTS

The direct product of two matrices  $\mathbf{A}_{p \times q}$  and  $\mathbf{B}_{m \times n}$  is defined as

$$\mathbf{A}_{p \times q} \otimes \mathbf{B}_{m \times n} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix} \quad (27)$$

and is sometimes called the Kronecker product or Zehfuss product (see Henderson et al., 1981). Clearly, (27) is partitioned into as many submatrices as there are elements of  $\mathbf{A}$ , each submatrix being  $\mathbf{B}$  multiplied by an element of  $\mathbf{A}$ . Therefore the elements of the direct product consist of all possible products of an element of  $\mathbf{A}$  multiplied by an element of  $\mathbf{B}$ . It has order  $pm \times qn$ . For example,

$$[1 \ 2 \ 3] \otimes \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 6 & 7 & 12 & 14 & 18 & 21 \\ 8 & 9 & 16 & 18 & 24 & 27 \end{bmatrix}.$$

The transpose of a direct product is the direct product of the transposes—as is evident from transposing (27).

Direct products have many useful and interesting properties, some of which are as follows.

- (i) In contrast to  $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ , we have  $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$ .
- (ii) For  $\mathbf{x}$  and  $\mathbf{y}$  being vectors:  $\mathbf{x}' \otimes \mathbf{y} = \mathbf{y}\mathbf{x}' = \mathbf{y} \otimes \mathbf{x}'$ .
- (iii) For  $\lambda$  being a scalar:  $\lambda \otimes \mathbf{A} = \lambda\mathbf{A} = \mathbf{A} \otimes \lambda = \mathbf{A}\lambda$ .
- (iv) For partitioned matrices, although

$$[\mathbf{A}_1 \ \mathbf{A}_2] \otimes \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B} \ \mathbf{A}_2 \otimes \mathbf{B}],$$

$$\mathbf{A} \otimes [\mathbf{B}_1 \ \mathbf{B}_2] \neq [\mathbf{A} \otimes \mathbf{B}_1 \ \mathbf{A} \otimes \mathbf{B}_2].$$

- (v) Provided conformability requirements for regular matrix multiplication are satisfied,  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{X} \otimes \mathbf{Y}) = \mathbf{AX} \otimes \mathbf{BY}$ .
- (vi) For  $\mathbf{A}$  and  $\mathbf{B}$  square and nonsingular,  $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ .

(vii) For  $\mathbf{D}_k$  being a diagonal matrix of order  $k$  with elements  $d_i$ :

$$\mathbf{D}_k \otimes \mathbf{A} = d_1 \mathbf{A} \oplus d_2 \mathbf{A} \oplus \cdots \oplus d_k \mathbf{A},$$

$$\mathbf{A}_{p \times q} \otimes \mathbf{D}_k = (\mathbf{A} \otimes \mathbf{I}_k)(\mathbf{I}_q \otimes \mathbf{D}) = (\mathbf{A} \otimes \mathbf{I}) \left( \bigoplus_{i=1}^q \mathbf{D} \right).$$

(viii) Rank and trace obey product rules:  $r_{\mathbf{A} \otimes \mathbf{B}} = r_{\mathbf{A}} r_{\mathbf{B}}$  and  $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$ .

(ix) Provided  $\mathbf{A}$  and  $\mathbf{B}$  are square,  $|\mathbf{A}_{p \times p} \otimes \mathbf{B}_{m \times m}| = |\mathbf{A}|^m |\mathbf{B}|^p$ .

And, for the sake of completeness, although it comes from Chapter 11:

(x) Eigenvalues of  $\mathbf{A} \otimes \mathbf{B}$  are products of eigenvalues of  $\mathbf{A}$  with those of  $\mathbf{B}$ .

These results are readily illustrated with simple numerical examples, and most of them are not difficult to prove (see Exercise 17).

Sometimes  $\mathbf{A} \otimes \mathbf{B}$  is referred to as the *right direct product* to distinguish it from  $\mathbf{B} \otimes \mathbf{A}$  which is then called the *left direct product*; and on rare occasions the right-hand side of (27) will be found defined as  $\mathbf{B} \otimes \mathbf{A}$ . Whatever names are used, it is apparent from (27) that for  $\mathbf{A}_{p \times q} = \{a_{ij}\}$  and  $\mathbf{B}_{m \times n} = \{b_{rs}\}$  the elements of both  $\mathbf{A} \otimes \mathbf{B}$  and  $\mathbf{B} \otimes \mathbf{A}$  consist of all possible products  $a_{ij}b_{rs}$ . In fact,  $\mathbf{B} \otimes \mathbf{A}$  is simply  $\mathbf{A} \otimes \mathbf{B}$  with the rows and columns each in a different order. Thus

$$\mathbf{B} \otimes \mathbf{A} = \mathbf{P}(\mathbf{A} \otimes \mathbf{B})\mathbf{Q} \quad (28)$$

where  $\mathbf{P}$  and  $\mathbf{Q}$  are permutation matrices (Section 6.9), depending only on orders of  $\mathbf{A}$  and  $\mathbf{B}$  and not on their elements. Extensive discussion of these particular permutation matrices is to be found in Henderson and Searle (1981a).

**Illustration.** A problem in genetics that uses direct products is concerned with  $n$  generations of random mating starting with the progeny obtained from crossing two autotetraploid plants which both have genotype AAaa. Normally the original plants would produce gametes AA, Aa and aa in the proportion 1 : 4 : 1. But suppose the proportion is  $u : 1 - 2u : u$  where, for example,  $u$  might take the value  $(1 - \alpha)/6$ , for  $\alpha$  being a measure of "diploidization" of the plants:  $\alpha = 0$  is the case of autotetraploids with chromosome segregation and  $\alpha = 1$  is the diploid case with all gametes being Aa. The question now is, what are the genotypic frequencies in the population after  $n$  generations of random mating? Let  $\mathbf{u}_i$  be the vector of gametic frequencies and  $\mathbf{f}_i$  the vector of genotype frequencies in the  $i$ th

generation of random mating, where  $\mathbf{u}_0$  is the vector of gametic frequencies in the initial plants. Then

$$\mathbf{u}_0 = \begin{bmatrix} u \\ 1 - 2u \\ u \end{bmatrix}$$

and  $\mathbf{f}_{i+1} = \mathbf{u}_i \otimes \mathbf{u}_i$  for  $i = 0, 1, 2, \dots, n$ . Furthermore, the relationship between  $\mathbf{u}_i$  and  $\mathbf{f}_i$  at any generation is  $\mathbf{u}_i = \mathbf{B}\mathbf{f}_i$  where

$$\mathbf{B} = \begin{bmatrix} 1 & \frac{1}{2} & u & \frac{1}{2} & u & 0 & u & 0 & 0 \\ 0 & \frac{1}{2} & 1 - 2u & \frac{1}{2} & 1 - 2u & \frac{1}{2} & 1 - 2u & \frac{1}{2} & 0 \\ 0 & 0 & u & 0 & u & \frac{1}{2} & u & \frac{1}{2} & 1 \end{bmatrix}.$$

Thus

$$\begin{aligned} \mathbf{f}_i &= \mathbf{u}_{i-1} \otimes \mathbf{u}_{i-1} = \mathbf{B}\mathbf{f}_{i-1} \otimes \mathbf{B}\mathbf{f}_{i-1} = (\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-1} \otimes \mathbf{f}_{i-1}) \\ &= (\mathbf{B} \otimes \mathbf{B})[(\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2}) \otimes (\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2})] \\ &= (\mathbf{B} \otimes \mathbf{B})[(\mathbf{B} \otimes \mathbf{B}) \otimes (\mathbf{B} \otimes \mathbf{B})][(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2}) \otimes (\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2})]. \end{aligned}$$

It is easily seen (and can be verified by induction) that

$$\mathbf{f}_i = \otimes^2 \mathbf{B}(\otimes^4 \mathbf{B})(\otimes^8 \mathbf{B}) \cdots (\otimes^{2^{i-1}} \mathbf{B})(\otimes^{2^i} \mathbf{u}_0)$$

where  $\otimes^n \mathbf{B}$  means the direct product of  $n\mathbf{B}$ 's.

## EXERCISES

1. Prove that

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{R} & \mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{P} \\ \mathbf{R} \end{bmatrix} [\mathbf{X} \quad \mathbf{Y}].$$

2. Assume orthogonality for

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

- Write down the two sets of equations to be derived from  $\mathbf{P}\mathbf{P}' = \mathbf{I}$  and  $\mathbf{P}'\mathbf{P} = \mathbf{I}$ .
- Show that each set reduces to three equations.

- (c) One equation is  $\mathbf{AC}' + \mathbf{BD}' = \mathbf{0}$ . Use two of the other equations to show that this is consistent with one of the other five.
3. (a) Use the determinant of  $\begin{bmatrix} \mathbf{P} & \mathbf{x} \\ \mathbf{x}' & 1 \end{bmatrix}$  to prove that
- $$\mathbf{x}'\mathbf{P}^{-1}\mathbf{x} = 1 - |\mathbf{P} - \mathbf{xx}'|/|\mathbf{P}|.$$
- (b) Try using  $-1$  in place of  $1$  in the matrix.
4. Prove the following results:
- (a) The leading element of  $\mathbf{P}^{-1}$  is  $(1 - |\mathbf{Q}|/|\mathbf{P}|)$  where  $\mathbf{Q}$  is  $\mathbf{P}$  with its leading element reduced by unity.
- (b) The sum of the elements of  $\mathbf{P}^{-1}$  is  $(1 - |\mathbf{R}|/|\mathbf{P}|)$  where  $\mathbf{R}$  is  $\mathbf{P}$  with every element reduced by unity.
- (c)  $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = |\mathbf{I} + \mathbf{xy}'| - 1$ .
5. Prove: (a)  $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$   
 (b)  $|\mathbf{R} + \mathbf{ZDZ}'| = |\mathbf{R}| |\mathbf{I} + \mathbf{DZ}'\mathbf{R}^{-1}\mathbf{Z}|$ .
6. Verify (12) and (14).
7. Show that, for all necessary inverses existing,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{I} \end{bmatrix} (\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1} \begin{bmatrix} -\mathbf{B}'\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\mathbf{C}^{-1}\mathbf{B}' \end{bmatrix} (\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1} \begin{bmatrix} \mathbf{I} & -\mathbf{BC}^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{X}'_1\mathbf{X}_2 \\ \mathbf{X}'_2\mathbf{X}_1 & \mathbf{X}'_2\mathbf{X}_2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2 \\ \mathbf{I} \end{bmatrix} \mathbf{W} \begin{bmatrix} -\mathbf{X}'_2\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1} & \mathbf{I} \end{bmatrix}$$

$$\text{for } \mathbf{W}^{-1} = \mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2 \quad \text{with } \mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}'_2\mathbf{X}_2)^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1 \end{bmatrix} \mathbf{T} \begin{bmatrix} \mathbf{I} & -\mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1} \end{bmatrix}$$

$$\text{for } \mathbf{T}^{-1} = \mathbf{X}'_1\mathbf{M}_2\mathbf{X}_1 \quad \text{with } \mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2.$$

8. Verify that for  $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$  and  $\mathbf{M}_1$  and  $\mathbf{M}_2$  of Exercise 7

$$\begin{aligned} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' &= \mathbf{I} - \mathbf{M}_1 + \mathbf{M}_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{M}_1 \\ &= \mathbf{I} - \mathbf{M}_2 + \mathbf{M}_2\mathbf{X}_1(\mathbf{X}'_1\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{M}_2. \end{aligned}$$

9. Write down a numerical example of a skew symmetric matrix (see Section 3.1e) of order 3 and one of order 4, and denote them by  $\mathbf{M}_3$  and  $\mathbf{M}_4$ , respectively. Then (a) show that  $|\mathbf{M}_3| = 0$  but  $|\mathbf{M}_4| > 0$  and (b) invert  $\mathbf{M}_4$ .
10. Extend Exercise 9 to matrices of order  $k$ ,  $\mathbf{M}_k$  being any skew symmetric matrix of order  $k$ .
- (a) Show for  $k$  even that  $\mathbf{M}_k^{-1}$  is skew symmetric.
- (b) Show that  $|\mathbf{M}_k| = 0$  for  $k$  odd and  $|\mathbf{M}_k| > 0$  for  $k$  even. [Hint: Use partitioning and induction.]

11. By partitioning  $\begin{bmatrix} -5 & 3 & 5 & 1 & 0 \\ -4 & 8 & 10 & 0 & -1 \\ 2 & 13 & 11 & 2 & 1 \\ 0 & 1 & 1 & 3 & 2 \\ 3 & 1 & 0 & 7 & 5 \end{bmatrix}$ , find its inverse.

12. Prove:

$$(\mathbf{A} + \mathbf{BCB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}.$$

13. Given the existence of all the necessary inverses, show that

$$(a) \begin{bmatrix} \mathbf{I} & \mathbf{P} \\ \mathbf{Q} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{I} - \mathbf{PQ})^{-1} & -(\mathbf{I} - \mathbf{PQ})^{-1}\mathbf{P} \\ -\mathbf{Q}(\mathbf{I} - \mathbf{PQ})^{-1} & \mathbf{I} + \mathbf{Q}(\mathbf{I} - \mathbf{PQ})^{-1}\mathbf{P} \end{bmatrix},$$

$$(b) \begin{bmatrix} \mathbf{0} & \mathbf{P} \\ \mathbf{Q} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} -(\mathbf{PQ})^{-1} & (\mathbf{PQ})^{-1}\mathbf{P} \\ \mathbf{Q}(\mathbf{PQ})^{-1} & \mathbf{I} - \mathbf{Q}(\mathbf{PQ})^{-1}\mathbf{P} \end{bmatrix},$$

$$(c) \begin{bmatrix} \mathbf{0} & \mathbf{P} \\ \mathbf{Q} & \mathbf{R} \end{bmatrix}^{-1} = \begin{bmatrix} -(\mathbf{PR}^{-1}\mathbf{Q})^{-1} & (\mathbf{PR}^{-1}\mathbf{Q})^{-1}\mathbf{PR}^{-1} \\ \mathbf{R}^{-1}\mathbf{Q}(\mathbf{PR}^{-1}\mathbf{Q})^{-1} & \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Q}(\mathbf{PR}^{-1}\mathbf{Q})^{-1}\mathbf{PR}^{-1} \end{bmatrix},$$

$$(d) \begin{bmatrix} \mathbf{S} & \mathbf{P} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{S}^{-1} & -\mathbf{S}^{-1}\mathbf{PR}^{-1} \\ \mathbf{0} & \mathbf{R}^{-1} \end{bmatrix}.$$

14. Use (19) to calculate  $Q^*$  for  $Q$  of (21), first using (22) and then (23). Verify by calculating  $QQ^*Q$  that  $Q^*$  based on (22) is not a generalized inverse whereas based on (23) it is.
15. In three different ways calculate the determinant and inverse of

$$\begin{bmatrix} 2 & 1 & 4 & -2 & -1 & -4 \\ 3 & 0 & -1 & -3 & 0 & 1 \\ 2 & 3 & 4 & -2 & -3 & -4 \\ 4 & 2 & 8 & 6 & 3 & 12 \\ 6 & 0 & -2 & 9 & 0 & -3 \\ 4 & 6 & 8 & 6 & 9 & 12 \end{bmatrix}$$

16. Show that each submatrix of (12) equals its counterpart in (14).
17. Establish properties (i) to (ix) of Section 7. For (ix), make use of (v), (28) and the determinant of  $I \otimes A$ .
18. Illustrate properties (vi), (viii) and (ix) of Section 7 using

$$A = \begin{bmatrix} 4 & -2 \\ 5 & -3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 \\ 2 & 5 \end{bmatrix}.$$

19. Suppose  $A$  is p.d. Define the conformable partitionings

$$B = \begin{bmatrix} A & b \\ b' & \beta \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} x \\ \lambda \end{bmatrix}.$$

Prove that  $B$  is p.d. if and only if all its principal leading minors are positive. (This is the basis of the inductive proof of Seelye, mentioned in Section 7.7a.)

[Hint: Consider  $w'Bw$ ; and use (6).]

20. Show that

- (a)  $(x' \otimes B)(A \otimes y) = B y x' A$ .  
 (b)  $(\lambda \otimes y)(x' \otimes \mu) = \mu \lambda y x'$ .  
 (c)  $(B \otimes B)(x \otimes x) = (B \otimes B x)x \neq B(x \otimes B x)$ .  
 (d)  $(A \otimes B)^{-1}$  exists only if  $A^{-1}$  and  $B^{-1}$  exist.

- (e)  $A \otimes (B \oplus C) \neq (A \otimes B) \oplus (A \otimes C)$   
 (f)  $(A \oplus B) \otimes C = (A \otimes C) \oplus (B \otimes C)$ .
21. Suppose  $A$  is symmetric, of order  $n$  and rank  $n-1$ , with  $1'A = 0$ . Suppose also that  $r$  is a vector of order  $n$  with  $1'r = 0$ . Let  $s$  be any  $n-1$  elements of  $r$  and  $B$  the corresponding submatrix of  $A$ . Prove that  $s'B^{-1}s$  is invariant to  $s$ .
22. Prove the Hadamard inequality of Section 2.

# CHAPTER 11

## EIGENVALUES AND EIGENVECTORS

### 1. INTRODUCTION: AGE DISTRIBUTION VECTORS

Studies of animal populations often include investigating the distribution of individuals according to age. For an appropriate unit of time, depending on the species concerned, such a distribution can be summarized as  $\mathbf{n}_t$ :

$$\mathbf{n}_t = [n_{0t} \quad n_{1t} \quad n_{2t} \cdots n_{it} \cdots n_{kt}]'$$

where  $n_{it}$  is the number of individuals that are aged  $i$  at time  $t$ , with  $i$  and  $t$  being measured in the selected unit of time. Such individuals, although described as being of age  $i$ , are all those whose exact age at time  $t$  is at least  $i$  units of time but less than  $i + 1$ . In this way  $\mathbf{n}_t$  represents the distribution of the population according to age at time  $t$ , for all ages up to but not including  $k + 1$ . Hence  $\mathbf{n}_t$  is called the age distribution vector at time  $t$ .

If we knew  $\mathbf{n}_t$ , one item of interest would be to ascertain  $\mathbf{n}_{t+1}$ , the age distribution vector at time  $t + 1$ , and to investigate its relationship to  $\mathbf{n}_t$ . This is achieved by means of a matrix containing age-specific fertility and survival rates as elements. For the sake of simplicity the development is confined to females and uses

$p_i$  = probability a female aged  $i$  at time  $t$  survives to time  $t + 1$ ,

and

$f_i$  = number of daughters alive at time  $t + 1$  that were born during time  $t$  to  $t + 1$  to a female aged  $i$  at time  $t$ .

Consider  $\mathbf{n}_{t+1}$  with elements  $n_{i,t+1}$  for  $i = 0, 1, \dots, k$ . By definition,

$$\begin{aligned} n_{0,t+1} &= \text{number of females aged 0 to 1 at time } t + 1 \\ &= \text{number of females alive at time } t + 1 \\ &\quad \text{that were born during time } t \text{ to } t + 1. \end{aligned}$$

Hence from the definition of  $f_0, f_1, \dots, f_k$ ,

$$n_{0,t+1} = f_0 n_{0t} + f_1 n_{1t} + \cdots + f_k n_{kt}.$$

Similarly,

$$\begin{aligned} n_{1,t+1} &= \text{number of females aged 1 to 2 at time } t + 1 \\ &= \text{survivors of those females aged 0 to 1 at time } t \end{aligned}$$

and by the definition of  $p_0$  this means  $n_{1,t+1} = p_0 n_{0,t}$ . Likewise  $n_{2,t+1} = p_1 n_{1,t}$ ; and in general,  $n_{i,t+1} = p_{i-1} n_{i-1,t}$  for  $i = 1, 2, \dots, k$ . These results are brought together in the matrix equation

$$\mathbf{n}_{t+1} = \begin{bmatrix} n_{0,t+1} \\ n_{1,t+1} \\ n_{2,t+1} \\ \vdots \\ n_{k,t+1} \end{bmatrix} = \begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_k \\ p_0 & 0 & 0 & \cdots & 0 \\ 0 & p_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & p_{k-1} & 0 \end{bmatrix} \begin{bmatrix} n_{0,t} \\ n_{1,t} \\ n_{2,t} \\ \vdots \\ n_{k,t} \end{bmatrix} = \mathbf{M}\mathbf{n}_t, \quad (1)$$

where  $\mathbf{M}$  is the  $(k + 1) \times (k + 1)$  matrix whose only nonzero elements are  $f$ 's in the first row and  $p$ 's on the subdiagonal. Thus the relationship between  $\mathbf{n}_{t+1}$  and  $\mathbf{n}_t$  is  $\mathbf{n}_{t+1} = \mathbf{M}\mathbf{n}_t$ .

Bernadelli (1941) and Leslie (1945, 1948) are early references that discuss many implications of this relationship, especially Leslie, who treats the topic at great length. For the moment one question only is pertinent: does a time ever come when the age distribution becomes stable? Stability is here taken to mean that corresponding elements of age distribution vectors at one time and the next would be the same, apart from a constant factor. Should this state of affairs occur at time  $t = T$  say,  $n_{i,T+1}$  would equal  $\lambda n_{i,T}$ , i.e.,  $n_{i,T+1} = \lambda n_{i,T}$  for some constant  $\lambda$  for  $i = 0, 1, \dots, k + 1$ . Hence the relationship between  $\mathbf{n}_T$  and  $\mathbf{n}_{T+1}$  would be  $\mathbf{n}_{T+1} = \lambda \mathbf{n}_T$ . But using  $\mathbf{n}_{T+1} = \mathbf{M}\mathbf{n}_T$  (as already established) gives  $\mathbf{M}\mathbf{n}_T = \lambda \mathbf{n}_T$ . The question of the existence of a stable age distribution therefore reduces to asking: does there exist a

vector  $\mathbf{n}_T$  and a scalar  $\lambda$  such that  $\mathbf{M}\mathbf{n}_T = \lambda\mathbf{n}_T$ ? The answer is "yes." This chapter deals with such vectors and scalars and develops consequences of their existence.

## 2. DERIVATION OF EIGENVALUES

Generalization of the question just posed starts with a square matrix  $\mathbf{A}$ . Then consider equations

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \quad (2)$$

for  $\mathbf{u}$  a vector and  $\lambda$  a scalar and ask: under what conditions (other than  $\mathbf{u} = \mathbf{0}$ ) can  $\mathbf{u}$  and  $\lambda$  exist so that (2) is true? The answer lies in writing (2) as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}, \quad (3)$$

from which we know that if  $\mathbf{A} - \lambda\mathbf{I}$  is nonsingular the only solution is  $\mathbf{u} = \mathbf{0}$ . But if  $\mathbf{A} - \lambda\mathbf{I}$  is singular, a non-null solution for  $\mathbf{u}$  can be obtained from (33) of Section 9.7 as

$$\mathbf{u} = [(\mathbf{A} - \lambda\mathbf{I})^{-1}(\mathbf{A} - \lambda\mathbf{I}) - \mathbf{I}]\mathbf{z}, \quad (4)$$

using a generalized inverse of  $\mathbf{A} - \lambda\mathbf{I}$  and where  $\mathbf{z}$  is arbitrary. We return to this element of arbitrariness in Section 4. In the meantime observe that (4) exists so long as  $\mathbf{A} - \lambda\mathbf{I}$  is singular, i.e., so long as

$$|\mathbf{A} - \lambda\mathbf{I}| = 0. \quad (5)$$

Therefore (5) is the condition for  $\mathbf{u}$  and  $\lambda$  to exist such that (2) is true, i.e., pick  $\lambda$  so that the determinant of  $\mathbf{A} - \lambda\mathbf{I}$  is zero.

Equation (5) is called the *characteristic equation* of  $\mathbf{A}$ . For  $\mathbf{A}$  of order  $n$ , the characteristic equation is a polynomial equation in  $\lambda$  of order  $n$ , with  $n$  roots to be denoted by  $\lambda_1, \lambda_2, \dots, \lambda_n$ , some of which may be zero. These roots are called *latent roots*, *characteristic roots*, *eigenvalues* or  *$\lambda$ -roots*. Corresponding to each root  $\lambda_i$  is a vector  $\mathbf{u}_i$  satisfying (2):

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad \text{for } i = 1, \dots, n, \quad (6)$$

and these vectors  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  are correspondingly called *latent vectors*, *characteristic vectors* or *eigenvectors*. "Characteristic" or "eigen" are the usual names; "eigen" is used in this book.

**Example 1.** The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

has characteristic equation

$$\left| \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0; \quad \text{i.e., } \begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0. \quad (7)$$

Expanding the determinant in (7) gives

$$(1-\lambda)^2 - 36 = 0; \quad \text{i.e., } \lambda = -5 \text{ or } 7.$$

Note that characteristic equations are, from (5), always of the form shown in (7): equated to zero is the determinant of  $\mathbf{A}$  amended by subtracting  $\lambda$  from each diagonal element.

Methods for obtaining eigenvectors corresponding to solutions of a characteristic equation are discussed in Section 4, but meanwhile it can be seen here that

$$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5 \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7 \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad (8)$$

these being examples of (2). Thus  $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$  is an eigenvector corresponding to the eigenvalue  $-5$ , and  $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$  is a vector for the root  $7$ .

**Illustration.** Bernadelli (1941) hypothesizes a species of beetle "which lives three years only, and which propagates in its third year of life. Let the survival rate of the first age group be  $\frac{1}{2}$ , of the second  $\frac{1}{3}$ , and assume that each female in the age group 2 to 3 produces, on the average, 6 new living females." Then the matrix of fertility and survival rates,  $\mathbf{M}$  of the preceding section, is

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

Since  $|\mathbf{M}| = 1$ , and  $\mathbf{M} - \lambda\mathbf{I}$  is  $\mathbf{M}$  with  $-\lambda$  in place of its diagonal zeros,  $|\mathbf{M} - \lambda\mathbf{I}| = 0$  for  $\lambda = 1$ . Hence  $\mathbf{M}$  has an eigenvalue of unity. Corresponding to this,  $\begin{bmatrix} 6 & 3 & 1 \end{bmatrix}$  is an eigenvector; i.e.,

$$\begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \\ 1 \end{bmatrix}.$$

This is an example of (1). From it we can conclude that when any population of these beetles starts out with an age distribution in the proportion 6 : 3 : 1, it will always stay in that proportion.

**Example 2.** The characteristic equation for

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \text{ is } \begin{vmatrix} 2-\lambda & 2 & 0 \\ 2 & 1-\lambda & 1 \\ -7 & 2 & -3-\lambda \end{vmatrix} = 0.$$

Diagonal expansion of the determinant (Section 4.6) gives

$$-\lambda^3 + \lambda^2(2 + 1 - 3) - \lambda \left\{ \begin{vmatrix} 2 & 2 \\ 2 & 1 \end{vmatrix} + \begin{vmatrix} 2 & 0 \\ -7 & -3 \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ 2 & -3 \end{vmatrix} \right\} + |\mathbf{A}| = 0,$$

which, after simple arithmetic, reduces to

$$\lambda^3 - 13\lambda + 12 = 0, \text{ equivalent to } (\lambda - 1)(\lambda^2 + \lambda - 12) = 0.$$

Solutions are  $\lambda = 1, 3$  and  $-4$ , and these are the eigenvalues of  $\mathbf{A}$ .

### 3. ELEMENTARY PROPERTIES OF EIGENVALUES

Several properties of eigenvalues that stem directly from their definition (2) and the resulting characteristic equation (5) command attention before considering eigenvectors.

#### a. Eigenvalues of powers of a matrix

The defining equation is  $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$  of (2). Premultiplying this by  $\mathbf{A}$  and using (2) again gives

$$\mathbf{A}^2\mathbf{u} = \mathbf{A}\lambda\mathbf{u} = \lambda\mathbf{A}\mathbf{u} = \lambda(\lambda\mathbf{u}) = \lambda^2\mathbf{u}.$$

Comparing this equation with (2), we see that it defines  $\lambda^2$  as being an eigenvalue of  $\mathbf{A}^2$ . Similarly,  $\lambda^3$  is an eigenvalue of  $\mathbf{A}^3$ :  $\mathbf{A}^3\mathbf{u} = \mathbf{A}\lambda^2\mathbf{u} = \lambda^2\mathbf{A}\mathbf{u} = \lambda^3\mathbf{u}$ , and in general  $\lambda^k$  is an eigenvalue of  $\mathbf{A}^k$ :

$$\mathbf{A}^k\mathbf{u} = \lambda^k\mathbf{u}. \quad (9)$$

Furthermore, when  $\mathbf{A}$  is nonsingular (2) gives  $\mathbf{u} = \mathbf{A}^{-1}\lambda\mathbf{u} = \lambda\mathbf{A}^{-1}\mathbf{u}$  and so

$$\mathbf{A}^{-1}\mathbf{u} = \lambda^{-1}\mathbf{u}. \quad (10)$$

Hence when  $\lambda$  is an eigenvalue of  $\mathbf{A}$ , then  $\lambda^k$  is an eigenvalue of  $\mathbf{A}^k$  where  $k$  is positive if  $\mathbf{A}$  is singular, and  $k$  is positive or negative if  $\mathbf{A}$  is nonsingular. In particular, when  $\mathbf{A}$  is nonsingular with eigenvalue  $\lambda$ , the inverse  $\mathbf{A}^{-1}$  has  $1/\lambda$  as an eigenvalue. (Recall that  $k = 0$  gives  $\mathbf{A}^k = \mathbf{A}^0 = \mathbf{I}$ , as in Section 2.7i.)

#### b. Eigenvalues of a scalar-by-matrix product

Multiplying  $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$  of (2) by a scalar  $c$  gives

$$c\mathbf{A}\mathbf{u} = c\lambda\mathbf{u}. \quad (11)$$

Hence when  $\lambda$  is an eigenvalue of  $\mathbf{A}$ , we have  $c\lambda$  being an eigenvalue of  $c\mathbf{A}$ .

Rewriting (11) as  $\mathbf{A}(c\mathbf{u}) = \lambda(c\mathbf{u})$  shows that when  $\mathbf{u}$  is an eigenvector of  $\mathbf{A}$  so also is  $c\mathbf{u}$ , both vectors corresponding to the same eigenvalue  $\lambda$ . This is one form of the arbitrariness evident in (4).

When  $\mathbf{A}$  has eigenvalue  $\lambda$  then  $\mathbf{A} + c\mathbf{I}$  for scalar  $c$  has eigenvalue  $\lambda + c$ . This is so because

$$(\mathbf{A} + c\mathbf{I})\mathbf{u} = \mathbf{A}\mathbf{u} + c\mathbf{u} = \lambda\mathbf{u} + c\mathbf{u} = (\lambda + c)\mathbf{u}.$$

Combining this with (10) shows that when  $(\mathbf{A} + c\mathbf{I})^{-1}$  exists it has  $1/(\lambda + c)$  as an eigenvalue.

#### c. Eigenvalues of polynomials

A consequence of the two preceding sections is that when  $\mathbf{A}$  has an eigenvalue  $\lambda$  then a polynomial in  $\mathbf{A}$ , say  $f(\mathbf{A})$ , has an eigenvalue  $f(\lambda)$ . For example, consider the polynomial  $f(\mathbf{A}) = \mathbf{A}^3 + 7\mathbf{A}^2 + \mathbf{A} + 5\mathbf{I}$ . If  $\lambda$  and  $\mathbf{u}$  are a corresponding eigenvalue and vector of  $\mathbf{A}$ ,

$$\mathbf{A}^3\mathbf{u} = \lambda^3\mathbf{u}, \quad \mathbf{A}^2\mathbf{u} = \lambda^2\mathbf{u}, \quad \mathbf{A}\mathbf{u} = \lambda\mathbf{u} \quad \text{and} \quad 5\mathbf{I}\mathbf{u} = 5\mathbf{u}.$$

Hence

$$\begin{aligned} (\mathbf{A}^3 + 7\mathbf{A}^2 + \mathbf{A} + 5\mathbf{I})\mathbf{u} &= \mathbf{A}^3\mathbf{u} + 7\mathbf{A}^2\mathbf{u} + \mathbf{A}\mathbf{u} + 5\mathbf{I}\mathbf{u} \\ &= \lambda^3\mathbf{u} + 7\lambda^2\mathbf{u} + \lambda\mathbf{u} + 5\mathbf{u} \\ &= (\lambda^3 + 7\lambda^2 + \lambda + 5)\mathbf{u}. \end{aligned}$$

Thus  $f(\lambda) = \lambda^3 + 7\lambda^2 + \lambda + 5$  is an eigenvalue. Extension to any polynomial  $f(\mathbf{A})$  gives the result that  $f(\mathbf{A})$  has eigenvalue  $f(\lambda)$ . An important special case is that  $e^{\mathbf{A}} = \sum_{i=0}^{\infty} \mathbf{A}^i/i!$  has eigenvalue  $e^\lambda$ .

**Example 1** (continued).

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

has eigenvalues  $-5$  and  $7$ . The eigenvalues of

$$f(\mathbf{A}) = 2\mathbf{A}^2 + 2\mathbf{A} - 12\mathbf{I}$$

$$= 2 \begin{bmatrix} 37 & 8 \\ 18 & 37 \end{bmatrix} + 2 \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - 12 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 64 & 24 \\ 54 & 64 \end{bmatrix}$$

are given by  $\begin{vmatrix} 64 - \lambda & 24 \\ 54 & 64 - \lambda \end{vmatrix} = 0$ , i.e.,  $\lambda^2 - 128\lambda + 2800 = 0$ , so that  $\lambda$  is  $100$  or  $28$ . And  $100 = 2(7)^2 + 2(7) - 12 = f(7)$ , and  $28 = f(-5)$ .

#### d. The sum and product of eigenvalues

The characteristic equation (5) of order 3,

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0, \quad (12)$$

reduces to

$$-\lambda^3 + (-\lambda)^2 \text{tr}_1(\mathbf{A}) + (-\lambda) \text{tr}_2(\mathbf{A}) + |\mathbf{A}| = 0 \quad (13)$$

using diagonal expansion of determinants and the  $\text{tr}_i(\mathbf{A})$  notation defined in Section 4.6. For  $\mathbf{A}$  of order  $n$  this takes the form

$$\begin{aligned} & (-\lambda)^n + (-\lambda)^{n-1} \text{tr}_1(\mathbf{A}) + (-\lambda)^{n-2} \text{tr}_2(\mathbf{A}) \\ & + \cdots + (-\lambda) \text{tr}_{n-1}(\mathbf{A}) + \text{tr}_n(\mathbf{A}) = 0. \end{aligned} \quad (14)$$

If  $\lambda_1, \lambda_2, \dots, \lambda_n$  are roots of this equation, then it is equivalent to

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) = 0,$$

which expands to

$$(-\lambda)^n + (-\lambda)^{n-1} \sum \lambda_i + (-\lambda)^{n-2} \sum_{i \neq j} \lambda_i \lambda_j + \cdots + \prod_{i=1}^n \lambda_i = 0. \quad (15)$$

Equating the coefficients of  $(-\lambda)^{n-1}$  in (14) and (15), and also equating the final terms, gives

$$\sum_{i=1}^n \lambda_i = \text{tr}_1(\mathbf{A}) = \text{tr}(\mathbf{A}) \quad \text{and} \quad \prod_{i=1}^n \lambda_i = \text{tr}_n(\mathbf{A}) = |\mathbf{A}|. \quad (16)$$

[11.4]

Hence the sum of the eigenvalues of a matrix equals its trace, and their product equals its determinant.

**Example 2** (continued).

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \quad \text{has determinant} \quad \begin{vmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -1 & 5 & 0 \end{vmatrix} = -12$$

and its eigenvalues are  $1, 3$  and  $-4$ . Their sum is  $1 + 3 - 4 = 0$  as is  $\text{tr}(\mathbf{A}) = 2 + 1 - 3 = 0$ ; and their product is  $1(3)(-4) = -12 = |\mathbf{A}|$ .

## 4. CALCULATING EIGENVECTORS

### a. A general method

Suppose  $\lambda_k$  is an eigenvalue of  $\mathbf{A}$ . It is a solution of the characteristic equation  $|\mathbf{A} - \lambda \mathbf{I}| = 0$ . Calculating an eigenvector corresponding to  $\lambda_k$  requires finding a non-null  $\mathbf{u}$  to satisfy  $\mathbf{A}\mathbf{u} = \lambda_k \mathbf{u}$ , which is equivalent to solving

$$(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{u} = \mathbf{0} \quad (17)$$

similar to (3). Since  $\lambda_k$  is such that  $|\mathbf{A} - \lambda_k \mathbf{I}| = 0$ ,  $\mathbf{A} - \lambda_k \mathbf{I}$  is singular and so at least one non-null solution to (17) always exists (see Section 9.7). However many solutions there are, they can be found from

$$\mathbf{u}_k = [(\mathbf{A} - \lambda_k \mathbf{I})^{-1} (\mathbf{A} - \lambda_k \mathbf{I}) - \mathbf{I}] \mathbf{z} \quad (18)$$

similar to (4), using a generalized inverse of  $\mathbf{A} - \lambda_k \mathbf{I}$  and an arbitrary  $\mathbf{z}$ . Furthermore, (17) has

$$n - r(\mathbf{A} - \lambda_k \mathbf{I}) \text{ LIN solutions,} \quad (19)$$

all obtainable from (18).

Suppose in (18) we make the partitioning

$$\mathbf{A} - \lambda_k \mathbf{I} = \begin{bmatrix} \mathbf{R}_k & \mathbf{C}_k \\ \mathbf{D} & \mathbf{E} \end{bmatrix} \quad (20)$$

for each eigenvalue  $\lambda_k$ , where  $\mathbf{R}_k$  is nonsingular with the same rank as  $\mathbf{A} - \lambda_k \mathbf{I}$ . Then using the algorithm of Section 8.4a we have

$$(\mathbf{A} - \lambda_k \mathbf{I})^{-1} = \begin{bmatrix} \mathbf{R}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$



and so (18), on partitioning the arbitrary  $\mathbf{z}$  as  $\mathbf{z}' = [-\mathbf{v} \quad -\mathbf{w}]'$ , becomes

$$\mathbf{u}_k = \begin{bmatrix} -\mathbf{R}_k^{-1}\mathbf{C}_k\mathbf{w} \\ \mathbf{w} \end{bmatrix} \text{ for arbitrary } \mathbf{w} \text{ of order } n - r(\mathbf{A} - \lambda_k\mathbf{I}). \quad (21)$$

### b. Simple roots

Results (18), (20) and (21) must be used for each eigenvalue  $\lambda_k$  of  $\mathbf{A}$ , including  $\lambda_k = 0$  if it occurs. Since  $\lambda_k$  is a solution to a polynomial equation it can be a solution more than once, in which case it is called a *multiple root*. We deal first with  $\lambda_k$  being a solution only once, in which case we call  $\lambda_k$  a *simple root*.

Whenever  $\lambda_k$  is a simple root it will be found [see Section 11A.1] that  $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1$ . Hence by (19) there is just one LIN eigenvector corresponding to  $\lambda_k$ . This is reflected in (21) where  $\mathbf{w}$  is then a scalar,  $w$ , indicating that all values of  $\mathbf{u}_k$  for a given  $\lambda_k$  are multiples of one another. (That they are all eigenvectors is further attested to by Section 3b where we saw that if  $\mathbf{u}$  is an eigenvector of  $\mathbf{A}$  so is  $c\mathbf{u}$  for any scalar  $c$ .)

#### Example 2 (continued).

$$\text{Eigenvalues of } \mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \text{ are } \lambda_1 = 1, \lambda_2 = 3 \text{ and } \lambda_3 = -4.$$

For  $\lambda_1 = 1$ , (20) and (21) give  $\mathbf{A} - \lambda_1\mathbf{I} =$

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -7 & 2 & -4 \end{bmatrix} \text{ and } \mathbf{u}_1 = \begin{bmatrix} \frac{1}{4} \begin{pmatrix} 0 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} w \\ w \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}w \\ \frac{1}{4}w \\ w \end{bmatrix}. \quad (22)$$

Similarly, with  $\lambda_2 = 3$ , (20) and (21) give  $\mathbf{A} - \lambda_2\mathbf{I} =$

$$\begin{bmatrix} -1 & 2 & 0 \\ 2 & -2 & 1 \\ -7 & 2 & -6 \end{bmatrix} \text{ and } \mathbf{u}_2 = \begin{bmatrix} \frac{1}{2} \begin{pmatrix} -2 & -2 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} w \\ w \end{bmatrix} = \begin{bmatrix} -w \\ -\frac{1}{2}w \\ w \end{bmatrix} \quad (23)$$

and  $\lambda_3 = -4$  yields  $\mathbf{A} - \lambda_3\mathbf{I} =$

$$\begin{bmatrix} 6 & 2 & 0 \\ 2 & 5 & 1 \\ -7 & 2 & 1 \end{bmatrix} \text{ and } \mathbf{u}_3 = \begin{bmatrix} -\frac{1}{26} \begin{pmatrix} 5 & -2 \\ -2 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} w \\ w \end{bmatrix} = \begin{bmatrix} w/13 \\ -3w/13 \\ w \end{bmatrix}. \quad (24)$$

In each of (22), (23) and (24) the  $w$  is arbitrary and can be chosen at will, e.g., to make elements of the  $\mathbf{u}$ 's be integers. For example, using  $w = 4, 2$  and  $13$  in (22), (23) and (24), respectively, gives one set of eigenvectors as

$$\mathbf{u}_1 = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \text{ and } \mathbf{u}_3 = \begin{bmatrix} 1 \\ -3 \\ 13 \end{bmatrix}. \quad (25)$$

Verification that  $\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i$  in each case is easy.

### c. Multiple roots

We now deal with eigenvalues  $\lambda_k$  that are solutions of the characteristic equation  $|\mathbf{A} - \lambda\mathbf{I}| = 0$  more than one time. Any  $\lambda_k$  for which this occurs is called a *multiple eigenvalue* and the number of times that it is a solution is called its *multiplicity*. In general, we formulate  $\mathbf{A}$  of order  $n$  as having  $s$  distinctly different eigenvalues  $\lambda_1, \dots, \lambda_s$  with  $\lambda_k$  having multiplicity  $m_k$  for  $k = 1, 2, \dots, s$  and, of course,  $\sum_{k=1}^s m_k = n$ . Note two features of this formulation. If zero is an eigenvalue (as it can be), it is one of the  $\lambda_k$ 's; and simple eigenvalues are also included, their multiplicities each being 1.

As already noted, each eigenvalue  $\lambda_k$  having multiplicity  $m_k = 1$  leads to  $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1$  and hence, from (19), correspondingly just one LIN eigenvector. (This is so whether all eigenvalues have multiplicity 1, as in Example 2, or whether only some of them do.) In contrast, for any multiple eigenvalue  $\lambda_k$ , the rank of  $\mathbf{A} - \lambda_k\mathbf{I}$  must be ascertained for each value because, as shall be seen, this rank value plays an important role in subsequent developments.

#### Example 3. The characteristic equation of

$$\mathbf{A} = \begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix} \text{ reduces to } (\lambda - 1)(\lambda^2 - 1) = 0$$

so that the eigenvalues are  $\lambda_1 = 1$  with  $m_1 = 2$ , and  $\lambda_2 = -1$  with  $m_2 = 1$ . For  $\lambda_1 = 1$

$$(\mathbf{A} - \lambda_1\mathbf{I}) = \begin{bmatrix} -2 & -2 & -2 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \text{ has rank 1,} \quad (26)$$

so that by (19) there are  $3 - 1 = 2$  LIN eigenvectors of  $\mathbf{A}$  corresponding to  $\lambda_1 = 1$ . They are given by (21), with  $\mathbf{w}$  of order 2, as

$$\mathbf{u}_1 = \begin{bmatrix} \frac{1}{2} \begin{pmatrix} -2 & -2 \end{pmatrix} \mathbf{w} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} -w_1 - w_2 \\ w_1 \\ w_2 \end{bmatrix}. \quad (27)$$

Similarly,  $\lambda_2 = -1$  in (20) and (21) gives

$$\mathbf{A} - \lambda_2 \mathbf{I} = \begin{bmatrix} 0 & -2 & -2 \\ 1 & 3 & 1 \\ -1 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} 2w \\ -w \\ w \end{bmatrix}.$$

Again the  $w$ 's are arbitrary, and they can be chosen to yield the following:

$$\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{u}_1^* = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}. \quad (28)$$

Note that these three vectors are LIN.

**Example 4.** The characteristic equation of

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 2 \\ 2 & 3 & 0 \\ 0 & 4 & 5 \end{bmatrix} \quad \text{reduces to} \quad (\lambda - 1)^2(\lambda - 6) = 0,$$

so that eigenvalues are  $\lambda_1 = 1$  with  $m_1 = 2$ , and  $\lambda_2 = 6$  with  $m_2 = 1$ . Then

$$\mathbf{A} - \lambda_1 \mathbf{I} = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 2 & 0 \\ 0 & 4 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{A} - \lambda_2 \mathbf{I} = \begin{bmatrix} -6 & 1 & 2 \\ 2 & -3 & 0 \\ 0 & 4 & -1 \end{bmatrix} \quad (29)$$

both have rank 2, so that (19) indicates there are  $3 - 2 = 1$  LIN eigenvectors corresponding to each of  $\lambda_1$  and  $\lambda_2$ . Using the matrices of (29) in (20) and (21) gives these vectors as

$$\mathbf{u}'_1 = [w_1 \quad -w_1 \quad w_1] \quad \text{and} \quad \mathbf{u}'_2 = [3w_2/8 \quad 2w_2/8 \quad w_2]. \quad (30)$$

Numerical examples of these are

$$\mathbf{u}'_1 = [1 \quad -1 \quad 1] \quad \text{and} \quad \mathbf{u}'_2 = [3 \quad 2 \quad 8]. \quad (31)$$

Note in this example that although  $n = 3$  there are only 2 ( $< n$ ) LIN eigenvectors. We refer to this subsequently.

## 5. THE SIMILAR CANONICAL FORM

### a. Derivation

Every eigenvalue  $\lambda_i$  has a corresponding eigenvector  $\mathbf{u}_i$  for which

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{for} \quad i = 1, 2, \dots, n. \quad (32)$$

This is true for all  $n$  roots of the characteristic equation, which means that for  $\lambda_k$  of multiplicity  $m_k$  there are  $m_k$  equations like (32) each involving the same  $\lambda_k$ ; if there are  $m_k$  LIN eigenvectors  $\mathbf{u}_k$  corresponding to  $\lambda_k$  (as in Example 3), there will be one equation for each vector, whereas if there are not  $m_k$  LIN eigenvectors  $\mathbf{u}_k$  (as in Example 4), there can still be  $m_k$  equations with some of the  $\mathbf{u}_k$ 's being repetitions of others. In all cases the array of equations represented by (32) can be written as

$$\mathbf{A}[\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n] = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Then, on defining  $\mathbf{U}$  as the matrix of  $n$  eigenvectors and  $\mathbf{D}$  as the diagonal matrix of eigenvalues, namely

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n] \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}, \quad (33)$$

we have

$$\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{D}. \quad (34)$$

$\mathbf{D}$  is known as the *canonical form under similarity* or equivalently as the *similar canonical form*.

The repetition of  $\mathbf{u}_k$ 's is illustrated for Example 4 by

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 3 \\ -1 & -1 & 2 \\ 1 & 1 & 8 \end{bmatrix} \quad \text{with} \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}. \quad (35)$$

Through permitting this repetition whenever necessary, the form (34) can always be made to exist. However, in cases involving repeated  $\mathbf{u}_k$ 's it is clear that  $\mathbf{U}$  is singular. But for the cases when  $\mathbf{U}$  is nonsingular, i.e., when all  $n$  eigenvectors are LIN, (34) can be expressed as

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}, \quad (36)$$

using the notation for a diagonal matrix given in Section 1.6. The product on the left is known as reduction to the similar canonical form;  $\mathbf{D}$  on the right is the diagonal matrix of eigenvalues.

The existence of (36) depends on  $\mathbf{U}$  being nonsingular.  $\mathbf{AU} = \mathbf{UD}$  of (34) always exists, but rearranging it as  $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$  of (36) requires  $\mathbf{U}$  to be nonsingular. Within the context of eigenvalues and eigenvectors, the existence of nonsingular  $\mathbf{U}$  is sometimes referred to as  $\mathbf{A}$  being *diagonalizable*, since  $\mathbf{D}$  is diagonal. An important theorem provides us with conditions for ascertaining whether  $\mathbf{U}$  is singular or not. We refer to it as the diagonalizability theorem.

**Diagonalizability Theorem.**  $\mathbf{A}_{n \times n}$ , having eigenvalues  $\lambda_k$  with multiplicity  $m_k$  for  $k = 1, 2, \dots, s$  and  $\sum_{k=1}^s m_k = n$ , has  $n$  eigenvectors that are LIN if and only if  $r(\mathbf{A} - \lambda_k \mathbf{I}) = n - m_k$  for all  $k = 1, 2, \dots, s$ ; whereupon  $\mathbf{U}$  of (33) is nonsingular and  $\mathbf{A}$  is diagonalizable as  $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$  of (36).

Proof of this theorem is lengthy and is given in Section 11A.1c, together with prerequisite lemmas.

Existence of  $\mathbf{U}^{-1}$  depends upon ascertaining if

$$r(\mathbf{A} - \lambda_k \mathbf{I}) = n - m_k. \quad (37)$$

Each eigenvalue satisfying (37) is called a *regular* eigenvalue. When every eigenvalue is regular,  $\mathbf{U}^{-1}$  exists and  $\mathbf{A}$  is called a *regular matrix*. Whenever one or more eigenvalues are not regular,  $\mathbf{U}^{-1}$  does not exist. Thus a single violation of (37) ensures the nonexistence of  $\mathbf{U}^{-1}$ , whereupon  $\mathbf{U}^{-1}\mathbf{AU}$  does not exist and  $\mathbf{A}$  is said to be a *deficient* or *defective matrix*. Note, though, that there is no need to check (37) for simple eigenvalues (multiplicity  $m_k = 1$ ) because, as already alluded to and proved in Section 11A.1c, (37) is always satisfied for such values.

**Example 2 (continued).** Each eigenvalue has  $m_k = 1$ , we know that (37) is satisfied, and so  $\mathbf{U}^{-1}$  exists. Assembling the eigenvectors as given in (25) into a matrix  $\mathbf{U}$  we get

$$\mathbf{U} = \begin{bmatrix} -2 & -2 & 1 \\ 1 & -1 & -3 \\ 4 & 2 & 13 \end{bmatrix} \text{ and } \mathbf{U}^{-1} = \frac{1}{70} \begin{bmatrix} -7 & 28 & 7 \\ -25 & -30 & -5 \\ 6 & -4 & 4 \end{bmatrix}. \quad (38)$$

With these values

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D} = \text{diag}\{1, 3, -4\} \text{ and } \mathbf{A} = \mathbf{UDU}^{-1}. \quad (39)$$

The reader should check these calculations.

**Example 3 (continued).** Equation (26) has  $\lambda_1 = 1$  with  $m_1 = 2$ ; and  $r(\mathbf{A} - \lambda_1 \mathbf{I}) = 1 = 3 - 2 = n - m_1$  satisfies (37). The other eigenvalue is

$\lambda_2 = -1$  with  $m_2 = 1$  which we know satisfies (37); i.e., (37) is satisfied by all  $\lambda_k$ . Hence  $\mathbf{U}^{-1}$  exists: from (28)

$$\mathbf{U} = \begin{bmatrix} 0 & -2 & 2 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix}, \text{ with } \mathbf{U}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

and it is easily checked that

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D} = \text{diag}\{1, 1, -1\} \text{ and } \mathbf{A} = \mathbf{UDU}^{-1}.$$

**Example 4 (continued).** Here we have  $\lambda_1 = 2$  with  $m_1 = 2$  and from (29)

$$r(\mathbf{A} - \lambda_1 \mathbf{I}) = 2 \neq n - m_1 = 3 - 2 = 1.$$

Hence (37) is not satisfied,  $\mathbf{U}^{-1}$  does not exist, and so  $\mathbf{A}$  is not diagonalizable as  $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ ; but of course  $\mathbf{AU} = \mathbf{UD}$  exists, with  $\mathbf{U}$  and  $\mathbf{D}$  of (35) wherein the singularity of  $\mathbf{U}$  is clearly apparent.

Since (37) is satisfied for every simple eigenvalue, the diagonalizability theorem is satisfied for any matrix having all  $n$  eigenvalues distinct. It is also satisfied for symmetric matrices, as discussed in Section 11A.1d. Further properties of the eigenvalues and vectors of symmetric matrices are given in Section 11A.2.

#### b. Uses

Uses of the similar canonical form are many and varied. One of the most important in applications is the easy opportunity provided by the diagonalizability of  $\mathbf{A}$  for calculating powers of  $\mathbf{A}$ . This is because  $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$  implies

$$\mathbf{A} = \mathbf{UDU}^{-1} \quad (40)$$

and hence  $\mathbf{A}^2 = \mathbf{UDU}^{-1}\mathbf{UDU}^{-1} = \mathbf{UD}^2\mathbf{U}^{-1}$  and in general, for positive integers  $p$ ,

$$\mathbf{A}^p = \mathbf{UD}^p\mathbf{U}^{-1}. \quad (41)$$

Similarly, for  $\mathbf{A}$  nonsingular

$$\mathbf{A}^{-1} = \mathbf{UD}^{-1}\mathbf{U}^{-1}, \quad (42)$$

in which case (41) also holds for negative integers  $p$ . Because  $\mathbf{D}$  is diagonal, (41) is easy to calculate:  $\mathbf{D}^p$  is just  $\mathbf{D}$  with its (diagonal) nonzero elements raised to the  $p$ th power.

**Example 2** (continued). From (38) and (39), equation (41) is

$$\begin{aligned} \mathbf{A}^p &= \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix}^p \\ &= \begin{bmatrix} -2 & -2 & 1 \\ 1 & -1 & -3 \\ 4 & 2 & 13 \end{bmatrix} \begin{bmatrix} 1^p & 0 & 0 \\ 0 & 3^p & 0 \\ 0 & 0 & (-4)^p \end{bmatrix} \frac{1}{70} \begin{bmatrix} -7 & 28 & 7 \\ -25 & -30 & -5 \\ 6 & -4 & 4 \end{bmatrix}. \end{aligned} \quad (43)$$

The reader should verify this for  $p = -1, 1$  and  $2$ .

We give four illustrations involving powers of a matrix, where procedures like (43) can be used to calculate them.

**Illustration 1.** Li (1955, p. 107) discusses a genetics problem in which genotypes can be put into one of three groups  $T_0, T_1$  and  $T_2$ . The probabilities, under the breeding plan known as selfing, that an individual of type  $T_i$  produces a progeny of type  $T_j$  for  $i, j = 0, 1, 2$  can then be arrayed as a transition probability matrix  $\mathbf{P}$  having the value

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{18} & \frac{8}{18} & \frac{9}{18} \end{bmatrix}.$$

One question of interest to geneticists is "what is the probability that an individual of type  $T_i$  will have descendants of type  $T_j$  in the  $k$ th generation of selfing?" An extension of the question is to find the value of this probability when  $k$  is infinitely large. The complete problem therefore involves deriving the  $k$ th power of  $\mathbf{P}$  because its elements are, for  $k$  generations of selfing, the required probabilities. Calculations for obtaining  $\mathbf{P}^k$  from (41) are as follows.

On observing that elements in each row of  $\mathbf{P}$  sum to unity we can write  $\mathbf{P}\mathbf{1} = \mathbf{1}$ , which means that  $\lambda_1 = 1$  is one eigenvalue, for which

$$\mathbf{u}_1 = w\mathbf{1} = [w_1 \quad w_1 \quad w_1]' \quad (44)$$

is an eigenvector, with  $w$  being any scalar. More generally, for the other two eigenvalues,  $1/6$  and  $5/6$ , we use

$$\mathbf{P} - \lambda_k \mathbf{I} = \begin{bmatrix} 1 - \lambda_k & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} - \lambda_k & \frac{1}{4} \\ \frac{1}{18} & \frac{8}{18} & \frac{9}{18} - \lambda_k \end{bmatrix}. \quad (45)$$

Applying (20) and (21) to (45), first for  $\lambda_2 = \frac{1}{6}$  and then for  $\lambda_3 = \frac{5}{6}$ , gives

$$\mathbf{u}'_2 = [0 \quad -\frac{3}{4}w_2 \quad w_2] \quad \text{and} \quad \mathbf{u}'_3 = [0 \quad \frac{3}{4}w_3 \quad w_3]. \quad (46)$$

Hence, with appropriate choice of  $w$ 's in (44) and in (46),

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 3 \\ 1 & -4 & 4 \end{bmatrix} \quad \text{with} \quad \mathbf{U}^{-1} = \frac{1}{24} \begin{bmatrix} 24 & 0 & 0 \\ -1 & 4 & -3 \\ -7 & 4 & 3 \end{bmatrix}.$$

Thus

$$\mathbf{P}^k = \frac{1}{24} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 3 \\ 1 & -4 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & (1/6)^k & 0 \\ 0 & 0 & (5/6)^k \end{bmatrix} \begin{bmatrix} 24 & 0 & 0 \\ -1 & 4 & -3 \\ -7 & 4 & 3 \end{bmatrix},$$

which reduces to

$$\mathbf{P}^k = \begin{bmatrix} 1 & 0 & 0 \\ 1 - r_k/2 - 3s_k/8 & r_k/2 & 3s_k/8 \\ 1 - r_k/2 - 2s_k/3 & 2s_k/3 & r_k/2 \end{bmatrix} \quad (47)$$

where  $r_k = (5/6)^k + (1/6)^k$  and  $s_k = (5/6)^k - (1/6)^k$ .

From these expressions the elements of  $\mathbf{P}^k$  are readily obtainable for any value of  $k$ . The probability of a genotype of type  $T_i$  having a descendant of type  $T_j$  after  $k$  generations of selfing is then given by the appropriate element of  $\mathbf{P}^k$ . In particular, when  $k$  is infinitely large,  $r_k$  and  $s_k$  tend to zero

and the value of  $\mathbf{P}^k$  becomes  $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ , showing that ultimately all

genotypes become of type  $T_0$ .

Not all matrices used in this manner have a limiting form, as seen in the following illustration.

**Illustration 2.** The matrix appropriate to Bernadelli's beetles (*loc. cit.*) is

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix} \quad \text{with} \quad \mathbf{A}^2 = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 3 \\ \frac{1}{6} & 0 & 0 \end{bmatrix}$$

and  $\mathbf{A}^3 = \mathbf{I}$ . Hence  $\mathbf{A}^4 = \mathbf{A}$ ,  $\mathbf{A}^5 = \mathbf{A}^2$ ,  $\mathbf{A}^6 = \mathbf{I}$ ,  $\mathbf{A}^7 = \mathbf{A}$  and so on, with regular periodicity. Thus, as Bernadelli shows, if the beetle population is initially 3600, divided equally between its three stages of life, the sequence of age-distribution vectors from then on is as shown in Table 11.1.

TABLE 11.1. AGE-DISTRIBUTION VECTORS FOR BERNADELLI'S BEETLES

Year	0	1	2	3	4	...
Vector	$n_0$	$n_1$ $= An_0$	$n_2$ $= A^2 n_0$	$n_3$ $= A^3 n_0$	$n_4$ $= A^4 n_0$	...
Elements of vector	1200	7200 1200 1200	2400 600 400	1200 1200 1200	7200 600 400	...

The regular periodicity, with a three-year cycle, is evident. As shown earlier, the population is stable only if its initial age distribution is in the proportion 6:3:1. In general, if the initial proportion is  $6x:3y:z$ , the series of age-distribution vectors will be

$$\begin{bmatrix} 6x \\ 3y \\ z \end{bmatrix}, \begin{bmatrix} 6z \\ 3x \\ y \end{bmatrix}, \begin{bmatrix} 6y \\ 3z \\ x \end{bmatrix}, \begin{bmatrix} 6x \\ 3y \\ z \end{bmatrix}, \dots \text{and so on.}$$

This cyclic pattern can be established by deriving  $A^k = \mathbf{U}\mathbf{D}^k\mathbf{U}^{-1}$ ; however, two eigenvalues of  $\mathbf{A}$  are complex numbers, so we pursue the matter no further here.

**Illustration 3.** We see in (47) how the similar canonical form permits calculating powers of a matrix,  $\mathbf{P}^k$  and then, when it exists, the limiting value of  $\mathbf{P}^k$  as  $k \rightarrow \infty$ . Corresponding values of  $\mathbf{P}^k \mathbf{x}_0$  or  $\mathbf{x}'_0 \mathbf{P}^k$  (for any given  $\mathbf{x}_0$ ) are then readily obtainable. A particular case of this is a Markov chain where  $\mathbf{x}'_0$  is a vector of probabilities that add to unity:  $\mathbf{x}'_0 \mathbf{1} = 1$ . Then the successive vectors are  $\mathbf{x}'_k = \mathbf{x}'_0 \mathbf{P}^k$  for  $k = 1, 2, \dots$ . Suppose

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \begin{bmatrix} \mathbf{r}' \\ \mathbf{r}' \\ \vdots \\ \mathbf{r}' \end{bmatrix} = \mathbf{1}\mathbf{r}'.$$

All rows in this limiting form are the same,  $\mathbf{r}'$ , whereupon  $\mathbf{r}'$  is called the *vector of steady-state probabilities*. Then  $\lim_{k \rightarrow \infty} \mathbf{x}'_k = \mathbf{x}'_0 \mathbf{1}\mathbf{r}' = \mathbf{r}'$ ; i.e., the limiting form of  $\mathbf{x}'_k$  is the row  $\mathbf{r}'$  of steady-state probabilities, independent of the initial vector  $\mathbf{x}_0$ . In Illustration 1, the value of  $\mathbf{r}'$  is  $[1 \ 0 \ 0]$ .

**Illustration 4.** Economists and others often have occasion to use systems of linear equations known as linear difference equations. An example of a

system of first-order linear difference equations is

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{d} \tag{48}$$

for  $t = 1, 2, 3, \dots$ , and where  $\mathbf{A}$ ,  $\mathbf{x}_0$  and  $\mathbf{d}$  are known. The problem is to calculate values of  $\mathbf{x}_t$ . From (48) we have

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}(\mathbf{A}\mathbf{x}_{t-2} + \mathbf{d}) + \mathbf{d} = \mathbf{A}^2\mathbf{x}_{t-2} + (\mathbf{A} + \mathbf{I})\mathbf{d} = \dots \\ &= \mathbf{A}^t\mathbf{x}_0 + (\mathbf{A}^{t-1} + \mathbf{A}^{t-2} + \dots + \mathbf{A} + \mathbf{I})\mathbf{d}. \end{aligned}$$

Then, providing that  $\mathbf{A}^k \rightarrow \mathbf{0}$  as  $k \rightarrow \infty$  and that  $(\mathbf{I} - \mathbf{A})^{-1}$  exists, use can be made of (34) in Section 5.8b to yield

$$\mathbf{x}_t = \mathbf{A}^t\mathbf{x}_0 + (\mathbf{I} - \mathbf{A}^t)(\mathbf{I} - \mathbf{A})^{-1}\mathbf{d}. \tag{49}$$

The similar canonical form provides a method of calculating  $\mathbf{A}^t$ ; and if, as has already been assumed,  $\mathbf{A}^t \rightarrow \mathbf{0}$  as  $t \rightarrow \infty$ , then we see from (49) that the corresponding limit for  $\mathbf{x}_t$  is  $(\mathbf{I} - \mathbf{A})^{-1}\mathbf{d}$ .

An extension of (49) as the solution to (48) is to adapt linear difference equations of higher than the first order to being first-order equations. For example, suppose we wish to solve the third-order difference equation

$$y_t = \alpha y_{t-1} + \beta y_{t-2} + \gamma y_{t-3} + \delta, \tag{50}$$

for  $t = 3, 4, \dots$ , knowing  $\alpha, \beta, \gamma$  and  $\delta$  and the initial values  $y_0, y_1$  and  $y_2$ . Define  $z_t = y_{t-1}$  for  $t = 1, 2, \dots$  and  $w_t = y_{t-2} = z_{t-1}$  for  $t = 2, 3, \dots$ , and observe that using these and (50)

$$\begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \\ w_{t-1} \end{bmatrix} + \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix}. \tag{51}$$

Hence on defining

$$\mathbf{x}_t = \begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \alpha & \beta & \gamma \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{d} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix},$$

(51) can be expressed in the form of (48) and so (49) yields its solution, which is the solution of (50).

## 6. SYMMETRIC MATRICES

Symmetric matrices have sufficient notable properties in regard to eigenvalues and eigenvectors as to warrant special attention. Furthermore, the widespread use of symmetric matrices in statistics, through their involvement in quadratic forms (see Section 3.5), make it worthwhile to discuss these properties in detail. Only real, symmetric matrices are considered.

## a. Eigenvalues all real

The eigenvalues of a matrix of order  $n$  are roots of a polynomial equation of degree  $n$  and so are not necessarily real numbers. Some may be complex numbers, occurring in pairs as  $a + ib$  and its complex conjugate  $a - ib$ , where  $i = \sqrt{-1}$  and  $a$  and  $b$  are real. However, when  $\mathbf{A}$  is symmetric (and has elements that are real numbers) then all its eigenvalues are real. We state this formally as a lemma, with proof.

**Lemma.** The eigenvalues of every real symmetric matrix are real.

**Proof**<sup>1</sup>. Suppose  $\lambda$  is a complex eigenvalue of the symmetric matrix  $\mathbf{M}_{n \times n}$  with  $\mathbf{u}$  being a corresponding eigenvector. If  $\lambda = \alpha + i\beta$ , define  $\bar{\lambda} = \alpha - i\beta$ , and for  $\mathbf{u} = \{u_k\} = \mathbf{a} + i\mathbf{b}$  define  $\bar{\mathbf{u}} = \{\bar{u}_k\} = \mathbf{a} - i\mathbf{b}$ . Then by definition,  $\mathbf{M}\mathbf{u} = \lambda\mathbf{u}$  so that  $\bar{\mathbf{u}}'\mathbf{M}\mathbf{u} = \bar{\mathbf{u}}'\lambda\mathbf{u} = \lambda\bar{\mathbf{u}}'\mathbf{u}$ . But  $\mathbf{M}\bar{\mathbf{u}} = \bar{\lambda}\bar{\mathbf{u}}$  so that we also have  $\bar{\mathbf{u}}'\mathbf{M}\mathbf{u} = (\mathbf{M}\bar{\mathbf{u}})'\mathbf{u} = (\bar{\lambda}\bar{\mathbf{u}})'\mathbf{u} = \bar{\lambda}\bar{\mathbf{u}}'\mathbf{u}$ . Equating these two expressions for  $\bar{\mathbf{u}}'\mathbf{M}\mathbf{u}$  gives  $\lambda\bar{\mathbf{u}}'\mathbf{u} = \bar{\lambda}\bar{\mathbf{u}}'\mathbf{u}$ , and since  $\bar{\mathbf{u}}'\mathbf{u}$  is a sum of squares of real numbers it is nonzero and so  $\lambda = \bar{\lambda}$ , i.e.,  $a + ib = a - ib$  and hence  $b = 0$ . Thus every  $\lambda$  is real. QED

This lemma means that when dealing with symmetric matrices, all eigenvalues are real. Corresponding to each eigenvalue is an eigenvector  $\mathbf{u}$ , say, that is real and although  $\mathbf{u}\sqrt{-1}$ , which is not real, is also an eigenvector (Section 2b), attention can usually be confined to real eigenvectors.

## b. Symmetric matrices are diagonalizable

For every eigenvalue of a symmetric matrix  $\mathbf{A}$ , condition (37) is satisfied:  $r(\mathbf{A} - \lambda_k \mathbf{I}) = n - m_k$  for every  $\lambda_k$ . (Proof is given in Section 11A.1d.) Therefore  $\mathbf{A} = \mathbf{A}'$  is diagonalizable. Hence for any symmetric matrix  $\mathbf{A}$ , we have  $\mathbf{U}^{-1}$  of  $\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{D}$  existing, and  $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$  and  $\mathbf{A}^p = \mathbf{U}\mathbf{D}^p\mathbf{U}^{-1}$ . Moreover, as is now developed,  $\mathbf{U}$  is orthogonal.

## c. Eigenvectors are orthogonal

Symmetric matrices have eigenvectors that are orthogonal to each other. We establish this in two steps: first for eigenvectors corresponding to

<sup>1</sup>Thanks go to G. P. H. Styan for the brevity of this proof.

different eigenvalues and second for those corresponding to a multiple eigenvalue.

(i) *Different eigenvalues.* Consider two different eigenvalues  $\lambda_1 \neq \lambda_2$  with  $\mathbf{u}_1, \mathbf{u}_2$  being corresponding eigenvectors. Then with  $\mathbf{A} = \mathbf{A}'$  and  $\mathbf{A}\mathbf{u}_k = \lambda_k \mathbf{u}_k$  we have

$$\lambda_1 \mathbf{u}_2' \mathbf{u}_1 = \mathbf{u}_2' \lambda_1 \mathbf{u}_1 = \mathbf{u}_2' \mathbf{A}\mathbf{u}_1 = \mathbf{u}_1' \mathbf{A}' \mathbf{u}_2 = \mathbf{u}_1' \mathbf{A}\mathbf{u}_2 = \mathbf{u}_1' \lambda_2 \mathbf{u}_2 = \lambda_2 \mathbf{u}_1' \mathbf{u}_2 = \lambda_2 \mathbf{u}_2' \mathbf{u}_1;$$

i.e.,  $\lambda_1 \mathbf{u}_2' \mathbf{u}_1 = \lambda_2 \mathbf{u}_2' \mathbf{u}_1$ . But  $\lambda_1 \neq \lambda_2$ . Therefore  $\mathbf{u}_2' \mathbf{u}_1 = 0$ ; i.e.,  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are orthogonal. Hence eigenvectors corresponding to different eigenvalues are orthogonal.

(ii) *Multiple eigenvalues.* If  $\mathbf{A} = \mathbf{A}'$  has  $\lambda_k$  as an eigenvalue with multiplicity  $m_k$ , then  $\mathbf{A} - \lambda_k \mathbf{I}$  has rank  $n - m_k$  and is singular; and there are  $m_k$  eigenvectors corresponding to  $\lambda_k$  which are LIN solutions to  $(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{u} = \mathbf{0}$ . Such an equation can be made to yield LIN solutions that are orthogonal; see Section 9.7b wherein the equations are typified as  $\mathbf{A}\mathbf{x} = \mathbf{0}$  for real, singular  $\mathbf{A}$ . Hence if  $r(\mathbf{A} - \lambda_k \mathbf{I}) = n - m_k$ , there exist  $m_k$  LIN orthogonal eigenvectors of  $\mathbf{A}$  corresponding to  $\lambda_k$ . (For any eigenvalue  $\lambda_k$  that is not a complex number, this result holds for any real matrix  $\mathbf{A}$ , whether symmetric or not.)

(iii) *The canonical form under orthogonal similarity.* We now have that  $\mathbf{A} = \mathbf{A}'$  is diagonalizable, that eigenvectors corresponding to different eigenvalues are orthogonal, and that  $m_k$  LIN eigenvectors corresponding to any eigenvalue  $\lambda_k$  of multiplicity  $m_k$  can be obtained such that they are orthogonal. Thus the  $m_k$  eigenvectors corresponding to  $\lambda_k$  are orthogonal not only to each other but also to the  $m_i$  eigenvectors corresponding to each other eigenvalue  $\lambda_i$ ; and this is true for every eigenvalue  $\lambda_k$ . Hence eigenvectors for a symmetric matrix can always be found such that they are all orthogonal to each other. On normalizing each vector (see Section 3.4a) by changing  $\mathbf{u}$  to  $(1/\sqrt{\mathbf{u}'\mathbf{u}})\mathbf{u}$  and arraying the normalized vectors in a matrix  $\mathbf{U}$ , we then have  $\mathbf{U}$  as an orthogonal matrix and hence

$$\mathbf{U}'\mathbf{A}\mathbf{U} = \mathbf{D} \quad \text{with} \quad \mathbf{U}\mathbf{U}' = \mathbf{I}. \quad (52)$$

This is the *canonical form under orthogonal similarity* and it exists for all symmetric matrices.

## Example

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

has characteristic equation which reduces to  $(\lambda + 1)^2(\lambda - 5) = 0$ . Hence

$\lambda_1 = 5$  with  $m_1 = 1$ , and  $\lambda_2 = -1$  with  $m_2 = 2$ . From (20) and (21) for  $\lambda_1 = 5$

$$\mathbf{A} - \lambda_1 \mathbf{I} = \begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix}$$

with

$$\mathbf{u}_1 = \left[ \frac{1}{\sqrt{2}} \begin{pmatrix} -4 & -2 \\ -2 & -4 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} \mathbf{w} \right] = \begin{bmatrix} w \\ w \\ w \end{bmatrix}$$

as a corresponding eigenvector. Similarly, for  $\lambda_2 = -1$

$$\mathbf{A} - \lambda_2 \mathbf{I} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \quad \text{with} \quad \mathbf{u}_2 = \begin{bmatrix} -\frac{1}{2}[2 \ 2] \mathbf{w} \\ \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} -(w_1 + w_2) \\ w_1 \\ w_2 \end{bmatrix}$$

Clearly one value for  $\mathbf{u}_2$  is  $[-2 \ 1 \ 1]' = \mathbf{v}_2$ , say. Because  $m_2 = 2$  there will be two possible values of  $\mathbf{u}_2$  and by subsection (ii) they can be orthogonal. Hence a second value can be found by solving  $\mathbf{v}_2' \mathbf{u}_2 = 0$  for  $\mathbf{u}_2$ . Since  $\mathbf{v}_2' \mathbf{u}_2 = 0$  is  $2(w_1 + w_2) + w_1 + w_2 = 0$ , having  $w_1 = -1$  and  $w_2 = 1$  as a solution, a second possible value for  $\mathbf{u}_2$  is  $\mathbf{v}_3 = [0 \ -1 \ 1]$ . We therefore have three LIN orthogonal eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Arraying the normalized forms of these vectors as a matrix gives

$$\mathbf{U} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -2 & 0 \\ \sqrt{2} & 1 & -\sqrt{3} \\ \sqrt{2} & 1 & \sqrt{3} \end{bmatrix}$$

The reader should verify that  $\mathbf{U}'\mathbf{A}\mathbf{U} = \mathbf{D} = \text{diag}\{5, -1, -1\}$  and  $\mathbf{U}\mathbf{U}' = \mathbf{I}$

#### d. Rank equals number of nonzero eigenvalues

Define  $z_A$  as the number of zero eigenvalues of the matrix  $\mathbf{A}$ ; then  $n - z_A$  is the number of nonzero eigenvalues. Since, for  $\mathbf{A}$  being symmetric,

$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}'$  for nonsingular (orthogonal)  $\mathbf{U}$ , the ranks of  $\mathbf{A}$  and  $\mathbf{D}$  are equal,  $r_A = r_D$  (Lemma 1, Section 7.5). But the only nonzero elements in the diagonal matrix  $\mathbf{D}$  are the nonzero eigenvalues, and so its rank is the number of such eigenvalues,  $n - z_A$ . Hence

$$r_A = n - z_A, \quad \text{for} \quad \mathbf{A} = \mathbf{A}'; \quad (53)$$

i.e., for symmetric matrices rank equals the number of nonzero eigenvalues.

This result is true not only for all symmetric matrices (because they are diagonalizable), but also for all diagonalizable matrices. Other than diagonalizability there is nothing inherent in the development of (53) that uses the symmetry of  $\mathbf{A}$ . Nevertheless, (53) is of importance because it applies to all symmetric matrices.

## 7. DOMINANT EIGENVALUES

Several methods are available for calculating a numerically dominant eigenvalue of a matrix (e.g., Stewart, 1973, Chapter 7). They provide opportunity for finding the largest eigenvalue of a matrix without having to solve the characteristic equation, dealing, *inter alia*, with matters such as multiple roots, multiple dominant roots and roots involving complex numbers, as well as considering the problem of the actual existence of a dominant root. We discuss only one of these topics here: an easy method for calculating the largest eigenvalue *when it exists*. The existence of  $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$  and of a real-valued dominant root are assumed throughout; on this basis a method for obtaining that root is discussed.

The equation  $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$  has been established. Define

$$\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_n], \quad \mathbf{U}^{-1} \equiv \mathbf{V} = \begin{bmatrix} \mathbf{v}'_1 \\ \mathbf{v}'_2 \\ \vdots \\ \mathbf{v}'_n \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \text{diag}\{\lambda_i\}$$

for  $\lambda_i$  with  $i = 1, \dots, n$  being the eigenvalues of  $\mathbf{A}$ . Then  $\mathbf{A}^k = \mathbf{U}\mathbf{D}^k\mathbf{U}^{-1}$  can be written as

$$\mathbf{A}^k = [\lambda_1^k \mathbf{u}_1 \ \lambda_2^k \mathbf{u}_2 \ \cdots \ \lambda_n^k \mathbf{u}_n] \mathbf{V}$$

and from (11) of Section 3.1d this is

$$\mathbf{A}^k = \sum_{j=1}^n \lambda_j^k \mathbf{u}_j \mathbf{v}'_j \quad (54)$$

Suppose that  $\lambda_1$  is the dominant root, such that  $\lambda_1^k$  for some sufficiently large value of  $k$  is numerically so much greater than any of the values  $\lambda_2^k, \dots, \lambda_n^k$  that the latter may be taken as zero. Then (54) becomes

$$\mathbf{A}^k = \lambda_1^k \mathbf{u}_1 \mathbf{v}_1'. \quad (55)$$

Postmultiplying (55) by a non-null vector  $\mathbf{x}$  gives, on defining  $\theta = \mathbf{v}_1' \mathbf{x}$ ,

$$\mathbf{w}_k = \mathbf{A}^k \mathbf{x} = \lambda_1^k \mathbf{u}_1 \mathbf{v}_1' \mathbf{x} = \theta \lambda_1^k \mathbf{u}_1, \quad (56)$$

so defining  $\mathbf{w}_k$ . Then the ratio of the  $i$ th element of  $\mathbf{w}_k$  to that of  $\mathbf{w}_{k-1}$  is

$$\frac{w_{i,k}}{w_{i,k-1}} = \frac{\theta \lambda_1^k u_{i1}}{\theta \lambda_1^{k-1} u_{i1}} = \lambda_1. \quad (57)$$

Thus, on assuming that  $\lambda_1$  is the largest eigenvalue, and that for some large value of  $k$  the value of  $\lambda_1^k$  predominates sufficiently over  $\lambda_2^k, \dots, \lambda_n^k$  that the latter can be ignored in (54)—on the basis of these two assumptions we see in (56), for some arbitrary non-null vector  $\mathbf{x}$ , that  $\mathbf{A}^k \mathbf{x}$  and  $\mathbf{A}^{k-1} \mathbf{x}$  are vectors  $\mathbf{w}_k$  and  $\mathbf{w}_{k-1}$ , respectively, and in (57) that the ratio of their elements is  $\lambda_1$ . This means that repeated premultiplication of  $\mathbf{x}$  by  $\mathbf{A}$  will, after  $k$  multiplications, lead to a vector  $\mathbf{w}_k$  such that the ratio of each of its elements to the corresponding element of  $\mathbf{w}_{k-1}$  is the same for all elements, this ratio being the largest eigenvalue of  $\mathbf{A}$ . And furthermore,  $\mathbf{w}_{k-1}$  is the corresponding eigenvector, because

$$\mathbf{A} \mathbf{w}_{k-1} = \mathbf{A}^k \mathbf{x} = \mathbf{w}_k = \lambda_1 \mathbf{w}_{k-1}.$$

Although these results have been obtained by writing (55) as an exact equality, they are really only approximations. The degree of approximation depends on  $k$  and the extent to which  $\lambda_1^k$  exceeds  $\lambda_2^k, \dots, \lambda_n^k$  that the latter can be assumed zero in (54) in the presence of  $\lambda_1^k$ . Increasing  $k$  increases the accuracy of the approximation, so that in carrying out this process it is possible to find the largest eigenvalue to any required degree of accuracy by increasing  $k$ . The essential result, (57), is true for all values of  $i$ , that is, for all elements of  $\mathbf{w}_k$  and  $\mathbf{w}_{k-1}$ , namely of  $\mathbf{A}^k \mathbf{x}$  and  $\mathbf{A}^{k-1} \mathbf{x}$ . In any numerical case it is unlikely that these ratios will be exactly equal for any particular value of  $k$ ; they will differ, but the differences between them will decrease as the value of  $k$  is increased.

**Example.** For the matrix  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$  and the arbitrary vector  $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  Table 11.2 shows the vector  $\mathbf{w}_k = \mathbf{A}^k \mathbf{x}$  for  $k = 0, 1, \dots, 6$ , together with the ratios of corresponding elements,  $w_{i,k}/w_{i,k-1}$ .

TABLE 11.2. EXAMPLE OF ITERATIVE VALUES OF A DOMINANT EIGENVALUE

A	k						
	0	1	2	3	4	5	6
$\begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 7 \\ 13 \end{bmatrix}$	$\begin{bmatrix} 20 \\ 41 \end{bmatrix}$	$\begin{bmatrix} 61 \\ 121 \end{bmatrix}$	$\begin{bmatrix} 182 \\ 365 \end{bmatrix}$	$\begin{bmatrix} 547 \\ 1093 \end{bmatrix}$
	$\mathbf{w}_k = \mathbf{A}^k \mathbf{x}$						
	$w_{i,k}/w_{i,k-1}$						
$i = 1$	—	2	3.5	2.86	3.05	2.98	3.005
$i = 2$	—	5	2.6	3.15	2.95	3.02	2.995

We see at once the decrease in the differences between the ratios of corresponding elements as  $k$  increases—and at  $k = 6$  it is concluded that the largest eigenvalue is 3. This is confirmed by solving the characteristic equation. Also, by studying the vector  $\mathbf{w}_k$  as  $k$  increases, we see that its elements are approaching the ratio 1 : 2; thus an eigenvector associated with the dominant root of 3 is  $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ . This may be verified in the form of the equation  $\mathbf{A} \mathbf{u} = \lambda \mathbf{u}$ :

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

The vector  $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$  is approximately  $\mathbf{w}_6$  with 547 factored out.

Demonstrating this method with a  $2 \times 2$  matrix has kept the arithmetic to a minimum, but it can be applied to matrices of any order. The calculations become quite substantial for large-sized matrices, but with modern-day computers they are usually not a limiting problem. Discussion of some of the numerical analysis difficulties involved is given in Stewart (1973, pp. 340–348). Although the true test for ending the iterative procedure is that  $w_{i,k}/w_{i,k-1}$  should be the same for all  $i$ , within the limits of accuracy desired, two alternative procedures can be used that are not correct theoretically but are frequently satisfactory in practice. They demand continuing the iterative procedure until

$$\text{either } \sum_{i=1}^n \frac{w_{i,k}}{w_{i,k-1}} = \sum_{i=1}^n \frac{w_{i,k-1}}{w_{i,k-2}} \quad \text{or} \quad \frac{\sum_{i=1}^n w_{i,k}}{\sum_{i=1}^n w_{i,k-1}} = \frac{\sum_{i=1}^n w_{i,k-1}}{\sum_{i=1}^n w_{i,k-2}}.$$



Choice of which of these alternatives to use depends on the problem at hand and on matters of computing efficiency and numerical analysis which are beyond our scope here. Nevertheless, as alternative stopping procedures they are worth noting: one involves equating sums of ratios, the other involves ratios of sums.

**Illustration.** Dominant eigenvalues have direct application in problems relating to age-distribution vectors and to transition probability matrices discussed in Sections 2 and 3. It was seen there that if  $\mathbf{n}_0$  is an age-distribution vector at time  $t = 0$ , the corresponding vector at time  $t = k$  is  $\mathbf{n}_k = \mathbf{M}^k \mathbf{n}_0$ . Suppose that  $\lambda_1$  is the dominant eigenvalue of  $\mathbf{M}$  and that  $k$  is sufficiently large that  $\lambda_1^k$  is so much greater than the corresponding powers of other eigenvalues that they may be overlooked. Then, similar to (56),  $\mathbf{n}_k$  can be expressed as

$$\mathbf{n}_k = \lambda_1^k \mathbf{u}_1 \mathbf{v}_1' \mathbf{n}_0 = \lambda_1^k (\mathbf{v}_1' \mathbf{n}_0) \mathbf{u}_1.$$

Because  $\mathbf{u}_1$  is the first column of  $\mathbf{U}$  it is the eigenvector of  $\mathbf{M}$  corresponding to the eigenvalue  $\lambda_1$ ; and because  $\mathbf{v}_1' \mathbf{n}_0$  is a scalar  $(\mathbf{v}_1' \mathbf{n}_0) \mathbf{u}_1 = \mathbf{u}^*$  say, is also an eigenvector of  $\mathbf{M}$  corresponding to  $\lambda_1$ . Hence

$$\mathbf{n}_k = \mathbf{M}^k \mathbf{n}_0 = \lambda_1^k \mathbf{u}^*.$$

Likewise  $\mathbf{n}_{k+1} = \mathbf{M}^{k+1} \mathbf{n}_0 = \lambda_1^{k+1} \mathbf{u}^*$ , and so  $\mathbf{n}_{k+1} = \lambda_1 \mathbf{n}_k$ . But this is the very situation discussed at the beginning of the chapter, that of the age distribution being stable from one point in time to the next, apart from a constant, which is now seen to be  $\lambda_1$ , the dominant eigenvalue of  $\mathbf{M}$ . Therefore, if and when stability in the age distribution is reached, the ratio of successive age-distribution vectors is the dominant eigenvalue of  $\mathbf{M}$ , with the vectors being proportional to the corresponding eigenvector of  $\mathbf{M}$ .

These results mean that if stability is the only matter for investigation in a problem relating to age distributions, probability transition matrices or similar situations, the ratio of successive vectors and a vector proportional to them may be found, provided that a dominant eigenvalue exists, by carrying out the iterative process for deriving the dominant eigenvalue and an associated eigenvector. For example, in Illustration 1 of Section 5b the probability transition matrix for a problem in genetics is

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{18} & \frac{8}{18} & \frac{9}{18} \end{bmatrix}.$$

With the arbitrary vector  $\mathbf{x}' = [1 \ 1 \ 1]$  it is easily seen that  $\mathbf{P}\mathbf{x} = \mathbf{x}$ , and at once we conclude that the largest eigenvalue is unity, with  $\mathbf{x}$  being the associated eigenvector.

If there is no eigenvalue that can be termed dominant in the sense used here, the iterative procedure breaks down and tends to no limit. The matrix involved in Bernadelli's beetles is just such a case.

A large-scale use of this procedure for studying different hunting policies to be directed against the rabbit pest on New Zealand farmlands is reported by Darwin and Williams (1964). There, for a time interval of four weeks,  $\mathbf{M}$ -matrices of order 39 were used, of exactly the form described at the beginning of this chapter. Thirteen such matrices were available, representing the fertility and survival rates in each of the four-week periods of a calendar year. They differed because breeding habits and natural mortality vary with seasonal changes throughout the year. From these matrices,  $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_{13}$  say, the age-distribution vector at the end of a year,  $\mathbf{n}_E$ , is obtained from that at the beginning of the year,  $\mathbf{n}_0$ , as the product of the  $\mathbf{M}$ 's multiplied by  $\mathbf{n}_0$ :  $\mathbf{n}_E = \mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_{13} \mathbf{n}_0$ . And, assuming a stable age distribution, the dominant eigenvalue of the product  $\mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_{13}$  gives the effective rate of annual increase in rabbit population under natural conditions. But the point of interest was to compare the effect of different hunting policies, each designed to reduce the rabbit population. Comparison was made on the basis of minimum effect: in applying any policy it would be used only once a year. Each policy was represented by a diagonal matrix of order 39, the diagonal elements being the percentages of rabbits of each age (measured in four-week units) that might be expected to survive the hunting. Interspersing this matrix between any two of the  $\mathbf{M}$  matrices in the product of all 13 of them (or at the beginning or end of the product) represents in this scheme of things the 14 possible times of year at which the hunting policy could be waged on the rabbits. The dominant eigenvalue for each of these 14 possible products was then computed, this value representing the effective rate of annual increase in the rabbit population taking into account age-differential fertility and natural death (survival) rates, with hunting policy superimposed. By this means, the effectiveness of 11 different hunting policies was compared, each at 14 possible times of the year.

The results showed that it is better to use a hunting policy that kills more old than young rabbits just before the old start a new breeding season, and to use the opposite sort of hunting policy when the young have had a chance to be killed (naturally) but not much of a chance to breed, i.e., about six months away from the other optimum. While this result qualitatively agrees with general reasoning, its value was in showing the *size* of effect that can be achieved. Differences of 3 to 4 percent annual kill showed appreciable

differences in terms of overall effect in a population that is naturally increasing but at not too fast a rate.

### 8. FACTORING THE CHARACTERISTIC EQUATION

The examples of obtaining eigenvalues that have been used so far have all involved solving either a cubic or a quadratic equation in  $\lambda$ . This is not difficult, but in finding eigenvalues of large matrices there arises the problem of having to solve polynomial equations of large degree, for example, a  $20 \times 20$  matrix leads to a polynomial of degree 20. Should the dominant eigenvalue be obtainable by the procedure just discussed, the second largest value can be found by a similar process after factoring out the dominant value by a method now to be described.

Operationally the procedure is as follows. Suppose that  $\lambda_1$  is the dominant eigenvalue of  $\mathbf{A}$ , and that  $\mathbf{u}$  is an eigenvector corresponding to  $\lambda_1$ . Choose any vector  $\mathbf{v}$  such that  $\mathbf{v}'\mathbf{u} = 1$ . (A simple form for  $\mathbf{v}$  is the null vector with one zero element replaced by the reciprocal of the corresponding element of  $\mathbf{u}$ .) Then the second dominant root of  $\mathbf{A}$  is obtained as the dominant root of  $\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}'$ .

The following theorem is the basis of this procedure.

**Theorem.** If  $\lambda_1$  is an eigenvalue of  $\mathbf{A}$  with  $\mathbf{u}$  being a corresponding eigenvector, then eigenvalues of  $\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}'$  are eigenvalues of  $\mathbf{A}$  except that  $\lambda_1(1 - \mathbf{v}'\mathbf{u})$  replaces  $\lambda_1$ .

**Proof.**<sup>2</sup> Let  $\lambda$  (no subscript) be something other than an eigenvalue of  $\mathbf{A}$ . Then  $\mathbf{A} - \lambda\mathbf{I}$  is nonsingular and

$$\begin{aligned} |\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}' - \lambda\mathbf{I}| &= |(\mathbf{A} - \lambda\mathbf{I})[\mathbf{I} - (\mathbf{A} - \lambda\mathbf{I})^{-1}\lambda_1\mathbf{u}\mathbf{v}']| \\ &= |\mathbf{A} - \lambda\mathbf{I}| |1 - \lambda_1\mathbf{v}'(\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{u}|, \end{aligned}$$

on using (11) in Section 4.6, and Exercise 10.5a. But  $(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{A}\mathbf{u} - \lambda\mathbf{u} = (\lambda_1 - \lambda)\mathbf{u}$ , and so  $(\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{u} = \mathbf{u}/(\lambda_1 - \lambda)$ . Therefore

$$\begin{aligned} |\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}' - \lambda\mathbf{I}| &= |\mathbf{A} - \lambda\mathbf{I}| [1 - \lambda_1\mathbf{v}'\mathbf{u}/(\lambda_1 - \lambda)] \\ &= \prod_{i=1}^n (\lambda_i - \lambda) [\lambda_1(1 - \mathbf{v}'\mathbf{u}) - \lambda]/(\lambda_1 - \lambda) \\ &= \prod_{i=2}^n (\lambda_i - \lambda) [\lambda_1(1 - \mathbf{v}'\mathbf{u}) - \lambda]. \end{aligned} \quad (58)$$

<sup>2</sup>With grateful acknowledgment to G. P. H. Styan.

Hence, since the eigenvalues of  $\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}'$  are the solutions for  $\lambda$  obtained by equating the left-hand side of (58) to zero, they are, by scrutiny of the right-hand side, the same as those of  $\mathbf{A}$  except that  $\lambda_1$  is replaced by  $\lambda_1(1 - \mathbf{v}'\mathbf{u})$ . QED

**Corollary.** If  $\mathbf{v}'\mathbf{u} = 1$ , then the eigenvalues of  $\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}'$  are those of  $\mathbf{A}$ , with  $\lambda_1$  replaced by zero. Thus the dominant eigenvalue of  $\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}'$  is the second most dominant eigenvalue of  $\mathbf{A}$ .

**Example.** The characteristic equation of

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 2 \\ -10 & 9 & 1 \end{bmatrix}$$

reduces to

$$(\lambda - 2)(\lambda - 5)(\lambda + 4) = 0. \quad (59)$$

For the root  $\lambda_1 = 5$  it will be found that an eigenvector is  $\mathbf{u}' = [2 \ 8 \ 13]$  and the amended characteristic equation  $|\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}' - \lambda\mathbf{I}| = 0$  is

$$\begin{vmatrix} -4 - \lambda & 1 & 0 \\ -17 & 1 - \lambda & 2 \\ -42\frac{1}{2} & 9 & 1 - \lambda \end{vmatrix} = 0.$$

This becomes  $\lambda(\lambda - 2)(\lambda + 4) = 0$ , which is (59) with  $\lambda - 5$  replaced by  $\lambda$ .

In applying this technique to finding the second largest eigenvalue of a matrix one would usually employ the methods of the previous section to find the dominant roots of  $\mathbf{A}$  and  $\mathbf{A} - \lambda_1\mathbf{u}\mathbf{v}'$ .

### EXERCISES

- Find the eigenvalues and eigenvectors of the following matrices. In each case combine the eigenvectors into a matrix  $\mathbf{U}$  and verify that  $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D}$  where  $\mathbf{D}$  is the diagonal matrix of the eigenvalues.

$$\mathbf{B} = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 2 & -2 & 3 \\ 10 & -4 & 5 \\ 5 & -4 & 6 \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} 7 & 4 & -1 \\ 4 & 7 & -1 \\ -4 & -4 & 4 \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} 9 & 15 & 3 \\ 6 & 10 & 2 \\ 3 & 5 & 1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 4 \\ 2 & -4 & 1 \end{bmatrix}$$

$$\mathbf{K} = \begin{bmatrix} -1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & -1 \end{bmatrix} \quad \mathbf{L} = \begin{bmatrix} 1 & 4 & 2 \\ 4 & 1 & 2 \\ 8 & 8 & 1 \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} -9 & 2 & 6 \\ 2 & -9 & 6 \\ 6 & 6 & 7 \end{bmatrix} \quad \mathbf{N} = \begin{bmatrix} 2 & 7 & 1 \\ 2 & 3 & 8 \\ 9 & 4 & 4 \end{bmatrix}$$

- Use the iterative procedure to establish the dominant eigenvalue for each of the matrices in Exercise 1, and by using the factoring procedure of Section 8, find the other two eigenvalues.
- Of the matrices in Exercise 1, show that  $\mathbf{F} = \mathbf{B}^2 + \mathbf{B}$ , and that if  $\lambda$  is an eigenvalue of  $\mathbf{B}$  then  $\lambda^2 + \lambda$  is an eigenvalue of  $\mathbf{F}$ .
- Show that the eigenvectors of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ are of the form } \begin{bmatrix} -b \\ a - \lambda_1 \end{bmatrix} \text{ and } \begin{bmatrix} -b \\ a - \lambda_2 \end{bmatrix}$$

where  $\lambda_1$  and  $\lambda_2$  are the eigenvalues. Verify that  $2a = \lambda_1 + \lambda_2$  when  $a = d$ .

- Factorize the characteristic equation of

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

using the eigenvalue  $\lambda = a + 2b$ . Carry out the same process using the root  $\lambda = a - b$ .

- Express  $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ u & v & 0 \\ x & y & z \end{bmatrix}$  in the form  $\mathbf{A} = \mathbf{UDU}^{-1}$ . In doing so,

$$\text{notice that } \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ ac - b & -c & 1 \end{bmatrix}.$$

- In a selfing series with a pair of linked genes, the transition probability matrix for the states of heterogeneity at 0, 1 and 2 loci is

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} - c(1-c) & 2c(1-c) & \frac{1}{2} - c(1-c) \end{bmatrix},$$

where  $c$  is the proportion of crossovers. Use the result obtained in Exercise 6 to show that for  $z = \frac{1}{2} - c(1-c)$

$$\mathbf{P}^k = \begin{bmatrix} 1 & 0 & 0 \\ 1 - (1/2)^k & (1/2)^k & 0 \\ 1 + z^k - (1/2)^{k-1} & (1/2)^{k-1} - 2z^k & z^k \end{bmatrix}.$$

- The genetic variances and covariances in a diploid series of recurrent backcrosses to the  $F_1$ 's of two homozygotes is obtained from the power series

$$\mathbf{T}^k = \begin{bmatrix} \frac{1}{4} + (\frac{1}{2})^{k+1} & \frac{1}{2} & \frac{1}{4} - (\frac{1}{2})^{k+1} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} - (\frac{1}{2})^{k+1} & \frac{1}{2} & \frac{1}{4} + (\frac{1}{2})^{k+1} \end{bmatrix} \text{ for } \mathbf{T} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Find the eigenvalues of  $\mathbf{T}$  and verify  $\mathbf{T}^k$ .

- Investigation of the fate of a mutant recessive lethal gene under a series of brother-sister matings is based on the transition probability matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} \text{ with } \mathbf{S}^k = \begin{bmatrix} 1 & 0 & 0 \\ 1 - a_k - b_k & a_k & b_k \\ 1 - \alpha_k - \beta_k & \alpha_k & \beta_k \end{bmatrix},$$

where, with  $a = \sqrt{145}$ ,

$$s_k = \frac{1}{2} \left[ \left( \frac{17+a}{36} \right)^k + \left( \frac{17-a}{36} \right)^k \right]$$

and

$$d_k = \frac{1}{2a} \left[ \left( \frac{17+a}{36} \right)^k - \left( \frac{17-a}{36} \right)^k \right],$$

the elements of  $\mathbf{S}^k$  have  $a_k = s_k + d_k$ ,  $b_k = 9d_k$ ,  $\alpha_k = 16d_k$  and  $\beta_k = s_k - d_k$ . Verify  $\mathbf{S}^k$ .

- A simplified version of a theoretical model for the demand for capital goods uses  $\mathbf{d}_{2 \times 1}$  to represent desired levels of capital and

labor and  $x_n$  the actual levels at time  $n$ , with the relationship between the two being  $x_n = (\mathbf{I} - \mathbf{B})x_{n-1} + \mathbf{B}d$ . For  $\mathbf{B} = \begin{bmatrix} 0.1 & -0.5 \\ 0.3 & 0.9 \end{bmatrix}$ , demonstrate that as  $n \rightarrow \infty$  the actual levels of capital and labor do tend to the desired levels.

11. (a) Prove that unity is one eigenvalue of a transition probability matrix (see Section 1.6); find the corresponding eigenvector.
- (b) For a transition probability matrix of order 2, prove that each eigenvalue is equal to or less than unity.
12. Find the eigenvalues of  $\mathbf{I} - \mathbf{x}\mathbf{x}'$  for  $\mathbf{x}$  being a vector of order  $n$ . Find orthogonal eigenvectors of  $\mathbf{I} - \mathbf{x}\mathbf{x}'$  for  $\mathbf{x}' = [2 \ 4 \ 1 \ 3]$ .
13. Using the characteristic equation, show for  $\lambda$  being an eigenvalue of  $\mathbf{A}$  that  $1/(1 + \lambda)$  is an eigenvalue of  $(\mathbf{I} + \mathbf{A})^{-1}$ .
14. When eigenvalues of  $\mathbf{A}$  are positive, prove that those of  $\mathbf{A} + \mathbf{A}^{-1}$  are equal or greater than 2.
15. For  $\mathbf{A}$  being symmetric of order  $n$ , with eigenvalues  $\lambda_i$ , prove that

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 = \sum_{k=1}^n \lambda_k^2.$$

16. Use the iterative procedure, to at least five rounds of iteration, to suggest that the dominant eigenvalue of

$$\mathbf{G} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix}$$

is 6. Use the factorization procedure to suggest that the next largest eigenvalue is 3. Verify your results by solving the characteristic equation for  $\mathbf{G}$ .

17. Show that if  $\lambda$  is an eigenvalue of an orthogonal matrix, then so is  $1/\lambda$ . What is the numerical value of one root when the matrix is of odd order?
18. Suppose  $\mathbf{A}$  is idempotent and symmetric. Prove the following:
  - (a) Eigenvalues are 0 or 1.
  - (b) Trace equals rank.
  - (c) Diagonability theorem is satisfied.
  - (d)  $\mathbf{A} = \mathbf{B}\mathbf{B}'$  where  $\mathbf{B}'\mathbf{B} = \mathbf{I}$ .

19. Prove that the nonzero eigenvalues of the Kronecker product (Section 10.7) of  $\mathbf{A}$  and  $\mathbf{B}$  are products of those of  $\mathbf{A}$  and  $\mathbf{B}$ .
20. When  $\mathbf{Q} = \mathbf{R}\mathbf{R}'$  and  $\mathbf{Y}$  has the same order as  $\mathbf{Q}$ , prove that eigenvalues of  $\mathbf{Y}\mathbf{Q}$  are also eigenvalues of  $\mathbf{R}'\mathbf{Y}\mathbf{R}$ .
21. Prove that eigenvalues of  $\mathbf{B}\mathbf{A}$  are the same as those of  $\mathbf{A}\mathbf{B}\mathbf{A}$  when  $\mathbf{A}$  is idempotent.
22. Suppose  $\mathbf{T}$  and  $\mathbf{K}$  commute in multiplication. For  $\mathbf{x}$  being an eigenvector of  $\mathbf{T}$ , show that  $\mathbf{K}\mathbf{x}$  is also.
23. When  $\mathbf{G}$  is a generalized inverse of  $\mathbf{A}$ , with  $\mathbf{A}$  and  $\mathbf{G}\mathbf{A}$  being symmetric, show that the reciprocal of a nonzero eigenvalue of  $\mathbf{A}$  is an eigenvalue of  $\mathbf{G}$ .
24. For  $\mathbf{U}_1$  being the matrix of latent vectors corresponding to the nonzero eigenvalues of the symmetric matrix  $\mathbf{A}$ , show that  $\mathbf{U}_1(\mathbf{U}_1'\mathbf{A}\mathbf{U}_1)^{-1}\mathbf{U}_1'$  is a generalized inverse of  $\mathbf{A}$ .
25. The eigenvalues of  $\mathbf{A}$  with respect to  $\mathbf{V}$  are defined as the solutions for  $\lambda$  to  $|\mathbf{A} - \lambda\mathbf{V}| = 0$ ; and then  $\mathbf{A}\mathbf{t} = \lambda\mathbf{V}\mathbf{t}$  defines  $\mathbf{t}$  as an eigenvector of  $\mathbf{A}$  with respect to  $\mathbf{V}$ . If  $\mathbf{V}$  is positive definite, symmetric
  - (a) Show that  $\lambda$  is an eigenvalue of  $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}$ .
  - (b) Find  $\mathbf{t}$  in terms of an eigenvector of  $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}$ .
26. (a) Prove that the only nonsingular idempotent matrix is the identity matrix.
- (b) Suppose that for  $i = 1, 2, \dots, t$  the matrix  $\mathbf{X}_i$  is of order  $m \times k_i$  with rank  $r_i$ . If  $\sum_{i=1}^t r_i = m$  and if  $\mathbf{X}_i'\mathbf{X}_j = 0$  for  $i \neq j$ , with  $i, j = 1, 2, \dots, t$ , prove that

$$\sum_{i=1}^t \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' = \mathbf{I}_m.$$

Carefully state the reason for each step in your proof.

27. For any idempotent matrix  $\mathbf{A}$ , prove that
  - (a)  $\mathbf{A}^k$  has the same eigenvalues as  $\mathbf{A}$ .
  - (b)  $\mathbf{A}^k$  has rank  $r_{\mathbf{A}}$  and
  - (c)  $r(\mathbf{I} - \mathbf{A}) = n - r_{\mathbf{A}}$  for  $\mathbf{A}$  of order  $n$ .
28. Find the eigenvalues of  $a\mathbf{I} + b\mathbf{J}$  where  $\mathbf{J}$  has order  $r$  and every element unity. Find eigenvectors for the case of  $r = 4$  and conjecture a generalization to order  $r$ . Observe for order  $r$ , and for  $\mathbf{h}'$  and

- $\mathbf{K}$ , defined as in Exercise 3.20, that  $\mathbf{U}$  for  $\mathbf{U}' = [\mathbf{h} \ \mathbf{Q} \ \mathbf{K}]$  is a matrix of eigenvectors of  $a\mathbf{I} + b\mathbf{J}$ , where  $\mathbf{Q}$  is any orthogonal matrix of order  $r - 1$ .
29. Prove property (x) of direct products, in Section 10.7, and use it to easily prove properties (viii) and (ix).
  30. It might seem "obvious" that all eigenvalues of a matrix being zero implies the matrix is null.
    - (a) Prove this, for real, symmetric matrices.
    - (b) Create a numerical example (e.g., a  $2 \times 2$  matrix) illustrating that the statement is not true for nonsymmetric matrices. Note that your example also illustrates that Section 6d is not valid for nonsymmetric matrices.
  31. When  $\mathbf{A}$  and  $\mathbf{B}$  are real and  $\mathbf{B}$  is p.d., prove that solutions for  $\theta$  to  $|\mathbf{A} - \theta\mathbf{B}| = 0$  are real.
  32. For  $\mathbf{P}$  being a transition probability matrix, prove the following:
    - (a) The dominant eigenvalue of  $\mathbf{P}$  is unity.
    - (b) The subdominant eigenvalue of  $\mathbf{P}$  is the dominant eigenvalue of  $\mathbf{P} - \bar{\mathbf{J}}$ .
    - (c) Aside from (a), the subdominant eigenvalue of  $\mathbf{P} - \bar{\mathbf{J}}$  is less than unity.
  33. If the full rank factorization of  $\mathbf{A}$  having order  $n \times n$  and rank  $r$  is  $\mathbf{BC}$ , show that the characteristic equation of  $\mathbf{A}$  is  $\lambda^{n-r} |\mathbf{CB} - \lambda\mathbf{I}| = 0$ .
  34. If  $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$  and  $\mathbf{B}\mathbf{v} = \theta\mathbf{v}$  define eigenvalues and eigenvectors of  $\mathbf{A}$  and  $\mathbf{B}$ , use property (v) of Section 10.7 to prove property (x) in that same section. Also establish that an eigenvector of the Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$  is a Kronecker product of eigenvectors of  $\mathbf{A}$  and  $\mathbf{B}$ . Hence establish the similar canonical form of  $\mathbf{A} \otimes \mathbf{B}$ .

Use

$$\mathbf{A} = \begin{bmatrix} 4 & -2 \\ 5 & -3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & -1 \\ 2 & 5 \end{bmatrix}$$

to illustrate your results.

Consider also the case of  $\mathbf{A}$  and  $\mathbf{B}$  rectangular but  $\mathbf{A} \otimes \mathbf{B}$  square. How many eigenvalues of  $\mathbf{A} \otimes \mathbf{B}$  will be zero?

## CHAPTER 11A

### APPENDIX TO CHAPTER 11

Eigenvalues and eigenvectors are a foundation for extending both applications and theory far beyond the horizons of this book. A few indications of this are given here. First, with two pre-requisite lemmas, comes proof of the diagonalability theorem, together with proof that all symmetric matrices are regular, i.e., diagonalable. Second are some results for symmetric matrices; third is the Cayley-Hamilton theorem and finally, the very useful and important singular value decomposition of a matrix.

**Notation.** Most of the following notation comes from Chapter 11, but is summarized here for convenience.

$\mathbf{A}$  is square, of order  $n$  and rank  $r_A = r$ .

$t_i = \text{tr}_i(\mathbf{A}) =$  sum of principal minors of order  $i$  of  $|\mathbf{A}|$  (Section 4.5).

$t_0 = 1$ ,  $t_1 = \text{tr}(\mathbf{A})$  and  $t_n = |\mathbf{A}|$ .

$s =$  number of distinct eigenvalues.

$m_k =$  multiplicity of the eigenvalue  $\lambda_k$ , with  $\sum_{k=1}^s m_k = n$ .

$z_A =$  the number of zero eigenvalues of  $\mathbf{A}$ .

$\mathbf{D} =$  diagonal matrix of all  $n$  eigenvalues.

$\mathbf{U} =$  a matrix of eigenvectors used in  $\mathbf{AU} = \mathbf{UD}$ .

$\mathbf{B} = \mathbf{A} - \lambda_k \mathbf{I}$  for  $\lambda_k$  an eigenvalue of  $\mathbf{A}$ .

#### 1. PROVING THE DIAGONABILITY THEOREM

##### a. The number of nonzero eigenvalues never exceeds rank

It is shown in (53) of Section 11.6d that for symmetric matrices rank equals the number of nonzero eigenvalues,  $r_A = n - z_A$ ; and we remarked there that this is so for any diagonalable matrix  $\mathbf{A}$ . This is part of a more general result given in the following lemma applicable to any square matrix, whether diagonalable or not.

**Lemma 1.** For  $A_{n \times n}$  of rank  $r_A$ , with  $z_A$  zero eigenvalues,

$$r_A \geq n - z_A; \quad (1)$$

i.e., the number of nonzero eigenvalues never exceeds rank.

**Proof.** Let  $r_A = r$ . Then, when the full-rank factorization of  $A$  is  $XY$ , its characteristic equation can be expressed (Exercise 11.33) as

$$\lambda^{n-r} |YX - \lambda I| = 0. \quad (2)$$

Therefore  $\lambda = 0$  is certainly a root  $n - r$  times. But to the extent that  $YX$  has zero eigenvalues,  $\lambda = 0$  can be an eigenvalue of  $A$  more than  $n - r$  times. Therefore  $z_A \geq n - r_A$ . QED

**Example**

$$A = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 2 & 0 \\ 0 & 4 & 4 \end{bmatrix} \quad (3)$$

has rank 2 (row 3 = twice row 1 + row 2); and its characteristic equation reduces to  $\lambda^2(\lambda - 5) = 0$ . Hence  $r_A = 2 > n - z_A = 3 - 2 = 1$ , so illustrating the inequality in (1).

**b. A lower bound on  $r(A - \lambda_k I)$**

A matrix  $A$  is diagonalizable only if  $r(A - \lambda_k I) = n - m_k$  for every eigenvalue  $\lambda_k$  of multiplicity  $m_k$ . Example 4 at the end of Section 11.5a illustrates this condition not being satisfied. For cases like this, the following lemma shows that  $r(A - \lambda_k I) \geq n - m_k$ , i.e., that when  $r(A - \lambda_k I)$  does not equal  $n - m_k$ , it always exceeds  $n - m_k$ .

**Lemma 2.** When  $\lambda_k$  is an eigenvalue of  $A_{n \times n}$  with multiplicity  $m_k$ , then

$$r(A - \lambda_k I) \geq n - m_k. \quad (4)$$

**Proof.** Define  $B = A - \lambda_k I$  and  $r_B = n - p_k$ . Then  $n - r_B = p_k$  and so from Lemma 1

$$z_B \geq p_k. \quad (5)$$

For  $\theta$  being an eigenvalue of  $B$ , the corresponding eigenvalue of  $A = B + \lambda_k I$  is  $\theta + \lambda_k$ . Since by (5) we have  $\theta = 0$  not fewer than  $p_k$  times,  $0 + \lambda_k = \lambda_k$  is an eigenvalue of  $A$  not less than  $p_k$  times; i.e.,  $m_k \geq p_k$ . Hence  $r_B = r(A - \lambda_k I) = n - p_k \geq n - m_k$ , and (4) is established. QED

**Corollary.** For  $\lambda_k$  being a simple root,  $r(A - \lambda_k I) = n - 1$ .

**Proof.** Lemma 2 gives  $r(A - \lambda_k I) \geq n - 1$ . But  $|A - \lambda_k I| = 0$  and so  $r(A - \lambda_k I) < n$ . Therefore  $r(A - \lambda_k I) = n - 1 = n - m_k$ . QED

**c. Proof of the diagonalizability theorem**

For convenience, the theorem is restated.

**Theorem.**  $A_{n \times n}$ , having eigenvalues  $\lambda_k$  with multiplicity  $m_k$  for  $k = 1, 2, \dots, s$  and  $\sum_{k=1}^s m_k = n$ , has  $n$  eigenvectors that are LIN if and only if  $r(A - \lambda_k I) = n - m_k$  for all  $k = 1, 2, \dots, s$ ; whereupon  $U$  of  $AU = UD$  is nonsingular and  $A$  is diagonalizable as  $U^{-1}AU = D$ .

**Proof.**<sup>1</sup> Sufficiency: if  $r(A - \lambda_k I) = n - m_k$ , then  $U$  is nonsingular.

Since  $r(A - \lambda_k I) = n - m_k$ , the equation  $(A - \lambda_k I)x = 0$  has exactly  $n - (n - m_k) = m_k$  linearly independent non-null solutions. (See Section 9.7.) By definition these solutions are eigenvectors of  $A$ . Hence associated with each  $\lambda_k$  there is a set of  $m_k$  LIN eigenvectors.

To show that the sets are LIN of each other, suppose they are not, and that one vector of the second set of vectors,  $y_2$  say, is a linear combination of vectors of the first set,  $z_1, z_2, \dots, z_{m_1}$ . Then  $y_2 = \sum_{i=1}^{m_1} c_i z_i$  for some scalars  $c_i$  not all zero. Multiplying this equation by  $A$  leads to  $Ay_2 = \sum c_i Az_i$ . Because  $y_2$  and the  $z$ 's are eigenvectors corresponding respectively to the different eigenvalues  $\lambda_2$  and  $\lambda_1$ , this means  $\lambda_2 y_2 = \sum c_i \lambda_1 z_i = \lambda_1 \sum c_i z_i = \lambda_1 y_2$ , which cannot be true because  $\lambda_2 \neq \lambda_1$  and they are not both zero. Therefore the supposition is wrong, and we conclude that all  $s$  sets of  $m_k$  eigenvectors, for  $k = 1, 2, \dots, s$ , are LIN; i.e.,  $U$  is nonsingular.

Necessity: that if  $U^{-1}AU = D$  exists, then  $r(A - \lambda_k I) = n - m_k$ .

Because  $D$  is a diagonal matrix of all  $n$  eigenvalues,  $(D - \lambda_k I)$  has exactly  $m_k$  zeros in its diagonal and hence  $r(D - \lambda_k I) = n - m_k$ . But  $U^{-1}AU = D$  so that  $A = UDU^{-1}$ . Therefore  $(A - \lambda_k I) = (UDU^{-1} - \lambda_k I) = U(D - \lambda_k I)U^{-1}$ , and since multiplication by nonsingular matrices does not affect rank (Lemma 1, Section 7.5),  $r(A - \lambda_k I) = r(D - \lambda_k I) = n - m_k$ . QED

**d. All symmetric matrices are diagonalizable**

A symmetric matrix is diagonalizable because, as shown by the following lemma, the condition  $r(A - \lambda_k I) = n - m_k$  of the diagonalizability theorem is satisfied for every eigenvalue of  $A = A'$ .

**Lemma 3.** If  $\lambda_k$  is an eigenvalue of  $A = A'$ , having multiplicity  $m_k$ , then  $r(A - \lambda_k I) = n - m_k$ .

<sup>1</sup>I am indebted to Dr. B. L. Raktoc for the bulk of this proof.

The following proof of this lemma does not, of course, utilize any results in Chapter 11 that follow Section 11.6b where the diagonability of symmetric matrices is stated and where, for proof of that statement, the reader is referred to this section.

**Proof.** Use  $\mathbf{B} = \mathbf{A} - \lambda_k \mathbf{I}$ . With  $\mathbf{A}$  being symmetric,  $\lambda_k$  is real, and so  $\mathbf{B}$  is real; and  $\mathbf{B}$  is symmetric. Using Exercise 7.13 we therefore have  $r_{\mathbf{B}} = r_{\mathbf{B}'\mathbf{B}}$ , and so  $r_{\mathbf{B}} = r_{\mathbf{B}^2} = n - p = t$ , say, so defining  $p$  and  $t$ . Therefore [see Section 6.7, item (iv)] at least one submatrix of order  $t$  in  $\mathbf{B}$  is nonsingular, and so the corresponding principal minor in  $\mathbf{B}^2 = \mathbf{B}'\mathbf{B}$  is nonzero. And because (by the theorem in Section 7.7e)  $\mathbf{B}'\mathbf{B}$  is n.n.d., that principal minor of  $\mathbf{B}^2 = \mathbf{B}'\mathbf{B}$  (see Section 7.7a) is positive. Therefore the sum of such principal minors,  $\text{tr}_t(\mathbf{B}^2) \neq 0$ . Also, because  $r_{\mathbf{B}^2} = t$ , all minors of order greater than  $t$  are zero, and so  $\text{tr}_i(\mathbf{B}^2) = 0$  for  $i > t$ . Hence the characteristic equation of  $\mathbf{B}^2$  is

$$(-\lambda)^n + (-\lambda)^{n-1} \text{tr}_1(\mathbf{B}^2) + \cdots + (-\lambda)^{n-t} \text{tr}_t(\mathbf{B}^2) = 0.$$

Because the last term of this equation is nonzero,  $\lambda^{n-t}$  factors out; i.e.,  $\lambda = 0$  is a root  $n - t$  times. Thus  $z_{\mathbf{B}^2} = n - t = p$  and so  $z_{\mathbf{B}} = p$ . But for  $\theta$  being an eigenvalue of  $\mathbf{B}$ , that of  $\mathbf{A} = \mathbf{B} + \lambda_k \mathbf{I}$  is  $\theta + \lambda_k$ , and so  $0 + \lambda_k$  is an eigenvalue of  $\mathbf{A}$  with multiplicity  $p$ . Hence  $p = m$ . Thus  $r_{\mathbf{B}} = r(\mathbf{A} - \lambda_k \mathbf{I}) = n - p = n - m_k$ . QED

This lemma assures the diagonability of symmetric matrices; i.e., that for  $\mathbf{A} = \mathbf{A}'$  the  $\mathbf{U}$  in  $\mathbf{AU} = \mathbf{UD}$  is nonsingular and so  $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ ; and as is further shown in Section 11.6c,  $\mathbf{U}$  is an orthogonal matrix so that for  $\mathbf{A}$  being symmetric,  $\mathbf{U}'\mathbf{AU} = \mathbf{D}$  and  $\mathbf{A} = \mathbf{UDU}'$ . Equivalent to diagonability is the fact that  $r(\mathbf{A} - \lambda_k \mathbf{I}) = n - m_k$ ; this means that  $(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{u} = \mathbf{0}$  has  $n - (n - m_k) = m_k$  LIN solutions for  $\mathbf{u}$ ; i.e., the number of LIN eigenvectors corresponding to  $\lambda_k$  equals its multiplicity.

## 2. OTHER RESULTS FOR SYMMETRIC MATRICES

### a. Spectral decomposition

At the end of Section 3.1d, it is shown that  $\mathbf{AA}' = \sum_j \mathbf{a}_j \mathbf{a}_j'$ , where  $\mathbf{a}_j$  is the  $j$ th column of  $\mathbf{A}$ . Applied to columns of  $\mathbf{U}$  where  $\mathbf{U}$  is orthogonal,  $\mathbf{I} = \mathbf{UU}'$ , this gives  $\mathbf{I} = \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i'$ . If the  $\mathbf{u}_i$ 's are eigenvectors of  $\mathbf{A} = \mathbf{A}'$ , premultiplication by  $\mathbf{A}$  gives

$$\mathbf{A} = \mathbf{A} \sum \mathbf{u}_i \mathbf{u}_i' = \sum \mathbf{A} \mathbf{u}_i \mathbf{u}_i' = \sum \lambda_i \mathbf{u}_i \mathbf{u}_i'. \quad (6)$$

This is the *spectral decomposition* of  $\mathbf{A}$ . It also applies to powers of  $\mathbf{A}$ :

$$\mathbf{A}^k = \sum \lambda_i^k \mathbf{u}_i \mathbf{u}_i'$$

for any positive integer  $k$  (and any negative integer if  $\mathbf{A}$  is nonsingular).

**Example.** Corresponding to eigenvalues 5, -1 and -1, eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \quad \text{are} \quad 1/\sqrt{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad 1/\sqrt{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad 1/\sqrt{2} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix},$$

as in Section 11.6c. Then from (6)

$$\mathbf{A} = \frac{5}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [1 \ 1 \ 1] - \frac{1}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} [-2 \ 1 \ 1] - \frac{1}{2} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} [0 \ -1 \ 1].$$

**Illustration.** For a transition probability matrix  $\mathbf{P}$  and a known vector  $\mathbf{t}$  we often want to evaluate  $\mathbf{P}^k \mathbf{t}$  for some  $k$ . If  $\mathbf{P}$  is symmetric,

$$\mathbf{P}^k \mathbf{t} = \sum \lambda_i^k \mathbf{u}_i \mathbf{u}_i' \mathbf{t},$$

and if  $k$  is sufficiently large that the largest eigenvalue  $\lambda_L$  is such that  $\lambda_L^k$  greatly exceeds  $\lambda_i^k$  for any other eigenvalue, then an asymptotic value for  $\mathbf{P}^k \mathbf{t}$  is  $(\lambda_L^k \mathbf{u}_L' \mathbf{t}) \mathbf{u}_L$  where  $\mathbf{u}_L$  is the normalized eigenvector corresponding to  $\lambda_L$ .

### b. Non-negative definite (n.n.d.) matrices

Definition of n.n.d. matrices is given in Section 3.6, and elementary properties are discussed in Section 7.7. Since the definition is in terms of quadratic forms, n.n.d. matrices are usually taken as being symmetric, and so also have the following properties:

- (i) All eigenvalues are real (Section 11.6a).
- (ii) They are diagonable (Section 11.6b).
- (iii) Rank equals the number of nonzero eigenvalues (Section 11.6d).

These lead, in turn, to further results, as follows.

**Theorem.** The eigenvalues of a symmetric matrix are all non-negative if and only if the matrix is n.n.d.

**Proof<sup>1</sup>.** Let  $\mathbf{A} = \mathbf{A}'$  be real. Then  $\mathbf{A} = \mathbf{UDU}'$  for orthogonal  $\mathbf{U}$  and  $\mathbf{D} = \text{diag}\{\lambda_1, \dots, \lambda_n\}$  for  $\lambda_i, i = 1, \dots, n$ , being the eigenvalues of  $\mathbf{A}$ . Hence  $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{UDU}'\mathbf{x} = \mathbf{y}'\mathbf{D}\mathbf{y}$  for  $\mathbf{y} = \mathbf{U}'\mathbf{x}$ . Therefore  $\mathbf{A}$  is n.n.d. if, and only if,  $\mathbf{D}$  is n.n.d. But  $\mathbf{D}$  is diagonal and so is n.n.d. if and only if  $\lambda_i \geq 0$  for  $i = 1, \dots, n$ . QED

**Corollary.** The eigenvalues of a symmetric matrix are all positive if and only if the matrix is positive definite (p.d.).

**Proof<sup>1</sup>.** In the last two sentences of the preceding proof, n.n.d. and  $\lambda_i \geq 0$  are replaced by p.d. and  $\lambda_i > 0$ , respectively. QED

The consequences of this theorem are that when a symmetric matrix has all  $\lambda \geq 0$  it is n.n.d.; if all  $\lambda > 0$  it is p.d., and if some  $\lambda$ 's  $> 0$  and others (at least one) equal 0, then it is positive semidefinite (p.s.d.).

**Example.** The characteristic equation for the p.s.d. matrix of Section 3.6

$$\mathbf{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} \text{ is } \lambda(\lambda - 14)(\lambda - 53) = 0, \quad (7)$$

so that the eigenvalues are 0, 14 and 53.

A further feature of n.n.d. matrices is that each is the square of some other matrix of real elements. For if  $\mathbf{A}$  is n.n.d. we have  $\mathbf{U}'\mathbf{AU} = \mathbf{D}$  where

$$\mathbf{D} = \text{diag}\{\lambda_i\} = \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} \text{ for } \mathbf{D}^{\frac{1}{2}} = \text{diag}\{\sqrt{\lambda_i}\}$$

with  $\sqrt{\lambda_i}$  being the positive square root. Hence

$$\mathbf{A} = \mathbf{UDU}' = \mathbf{UD}^{\frac{1}{2}}\mathbf{U}'\mathbf{UD}^{\frac{1}{2}}\mathbf{U}' = \mathbf{H}^2 \text{ for } \mathbf{H} = \mathbf{UD}^{\frac{1}{2}}\mathbf{U}'. \quad (8)$$

For p.d. matrices no  $\lambda_i$  is zero and so  $\mathbf{D}$ ,  $\mathbf{D}^{\frac{1}{2}}$  and  $\mathbf{H}$  are all of full rank giving  $\mathbf{A}^{-1} = (\mathbf{H}^{-1})^2$ .

Also, if  $\mathbf{U}'\mathbf{AU} = \mathbf{D} = \begin{bmatrix} \Delta^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$  where  $\Delta = \text{diag}\{\sqrt{\lambda_i}\}$  for just the  $r_A$  nonzero eigenvalues, then with

$$\mathbf{S} = \begin{bmatrix} \Delta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r_A} \end{bmatrix} \text{ and } \mathbf{T} = \mathbf{US}, \text{ we have } \mathbf{T}'\mathbf{AT} = \begin{bmatrix} \mathbf{I}_{r_A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (9)$$

where  $\mathbf{T}$  is nonsingular.

<sup>1</sup>Thanks go to J. C. Berry for this proof.

**Example (continued).** Eigenvectors of  $\mathbf{A}$  in (7) corresponding to eigenvalues 14, 53 and 0 are

$$\begin{bmatrix} 2 \\ -13 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \quad (10)$$

respectively. Hence  $\mathbf{H}$  of (8) is

$$\mathbf{H} = \begin{bmatrix} 2 & -3 & 2 \\ -13 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{182} & 0 & 0 \\ 0 & 1/\sqrt{13} & 0 \\ 0 & 0 & 1/\sqrt{14} \end{bmatrix} \\ \times \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{53} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{182} & 0 & 0 \\ 0 & 1/\sqrt{13} & 0 \\ 0 & 0 & 1/\sqrt{14} \end{bmatrix} \begin{bmatrix} 2 & -13 & 3 \\ -3 & 0 & 2 \\ 2 & 1 & 3 \end{bmatrix},$$

with the second and fourth matrices in this product representing the normalization of the eigenvectors in (10). After simplification,  $\mathbf{H}$  can be written as

$$\mathbf{H} = \frac{1}{13\sqrt{14}} \begin{bmatrix} 4 + 9a & -26 & 6 - 6a \\ -26 & 169 & -39 \\ 6 - 6a & -39 & 9 + 4a \end{bmatrix} \text{ with } a = \sqrt{14(53)}.$$

The reader can verify that  $\mathbf{H}^2 = \mathbf{A}$ , as in (8). With

$$\mathbf{S} = \begin{bmatrix} 1/\sqrt{14} & 0 & 0 \\ 0 & 1/\sqrt{53} & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and}$$

$$\mathbf{T} = \mathbf{US} = \begin{bmatrix} 2 & -3 & 2 \\ -13 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{14(182)} & 0 & 0 \\ 0 & 1/\sqrt{53(13)} & 0 \\ 0 & 0 & \sqrt{1/14} \end{bmatrix},$$

(9) can also be verified.



c. **Simultaneous diagonalization of two symmetric matrices**

There are at least three situations in which it is possible to find a matrix  $\mathbf{P}$  that will simultaneously diagonalize two symmetric matrices (of the same order, obviously). They are as follows.

- (I)  $\mathbf{A}$  and  $\mathbf{B}$  symmetric, with  $\mathbf{AB} = \mathbf{BA}$ : then there exists an orthogonal  $\mathbf{P}$  such that  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{D}$ , a diagonal matrix of the eigenvalues of  $\mathbf{A}$ , and  $\mathbf{P}'\mathbf{B}\mathbf{P} = \mathbf{\Delta}$ , some other diagonal matrix.
- (II)  $\mathbf{A}$  being positive definite as well as symmetric: then there exists a  $\mathbf{P}$ , nonsingular but not necessarily orthogonal, such that  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{I}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P} = \mathbf{D}$  where  $\mathbf{D}$  is a diagonal matrix whose diagonal elements are solutions for  $\lambda$  to  $|\mathbf{B} - \lambda\mathbf{A}| = 0$ . The latter equation is sometimes called the characteristic equation of  $\mathbf{B}$  with respect to  $\mathbf{A}$ .
- (III)  $\mathbf{A}$  and  $\mathbf{B}$  both non-negative definite: then there exists a non-singular  $\mathbf{P}$  such that  $\mathbf{P}'\mathbf{A}\mathbf{P}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P}$  are both diagonal.

The three cases are stated as theorems and proved.

**Theorem 1.** For symmetric  $\mathbf{A}$  and  $\mathbf{B}$  of the same order, there exists an orthogonal matrix  $\mathbf{P}$  such that  $\mathbf{P}'\mathbf{A}\mathbf{P}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P}$  are both diagonal if and only if  $\mathbf{AB} = \mathbf{BA}$ .

**Proof.** This proof follows that of Graybill (1969).  
Sufficiency: that if  $\mathbf{AB} = \mathbf{BA}$ , an orthogonal  $\mathbf{P}$  exists.  
Because  $\mathbf{A}$  is symmetric there exists  $\mathbf{R}$  such that

$$\mathbf{R}'\mathbf{A}\mathbf{R} = \mathbf{D} = \text{diag}\{\lambda_i \mathbf{I}_{m_i}\} \quad \text{for } \mathbf{R} \text{ orthogonal,} \quad (11)$$

where  $\lambda_i$  is one of the  $s$  distinct eigenvalues of  $\mathbf{A}$  of multiplicity  $m_i$ . Let

$$\mathbf{R}'\mathbf{B}\mathbf{R} = \mathbf{C} = \mathbf{C}' = \{\mathbf{C}_{ij}\} \quad (12)$$

where  $\mathbf{C}$  is partitioned conformably with  $\mathbf{D}$  in (11).

With  $\mathbf{AB} = \mathbf{BA}$ , the orthogonality of  $\mathbf{R}$ , and the symmetry of  $\mathbf{A}$  and  $\mathbf{B}$ , we have  $\mathbf{CD} = \mathbf{R}'\mathbf{B}\mathbf{R}\mathbf{R}'\mathbf{A}\mathbf{R} = \mathbf{R}'\mathbf{B}\mathbf{A}\mathbf{R} = \mathbf{R}'\mathbf{A}\mathbf{B}\mathbf{R} = \mathbf{R}'\mathbf{A}\mathbf{R}\mathbf{R}'\mathbf{B}\mathbf{R} = \mathbf{DC}$ . Hence, because  $\mathbf{D}$  of (11) is diagonal,  $\mathbf{C}_{ij}\lambda_j = \lambda_i\mathbf{C}_{ij}$  for  $i \neq j$  and  $\lambda_i \neq \lambda_j$ , and so  $\mathbf{C}_{ij} = \mathbf{0}$  for  $i \neq j$ . Thus  $\mathbf{C}$  of (12) is block diagonal:

$$\mathbf{C} = \text{diag}\{\mathbf{C}_{ii}\}. \quad (13)$$

Symmetry of  $\mathbf{C}$  implies that  $\mathbf{C}_{ii}$  is symmetric, so that there exists  $\mathbf{Q}_i$  such that

$$\mathbf{Q}'_i\mathbf{C}_{ii}\mathbf{Q}_i = \mathbf{\Delta}_i \text{ is diagonal, with } \mathbf{Q}_i \text{ orthogonal.} \quad (14)$$

Let

$$\mathbf{Q} = \text{diag}\{\mathbf{Q}_i\}, \quad \text{also orthogonal,} \quad (15)$$

and define  $\mathbf{P} = \mathbf{R}\mathbf{Q}$ , which is orthogonal because  $\mathbf{R}$  and  $\mathbf{Q}$  are. Then  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{Q}'\mathbf{R}'\mathbf{A}\mathbf{R}\mathbf{Q} = \mathbf{Q}'\mathbf{D}\mathbf{Q} = \text{diag}\{\lambda_i \mathbf{Q}'_i\mathbf{Q}_i\} = \text{diag}\{\lambda_i \mathbf{I}_{m_i}\} = \mathbf{D}$ , using (11) and (15); and  $\mathbf{P}'\mathbf{B}\mathbf{P} = \mathbf{Q}'\mathbf{R}'\mathbf{B}\mathbf{R}\mathbf{Q} = \mathbf{Q}'\mathbf{C}\mathbf{Q} = \text{diag}\{\mathbf{\Delta}_i\} = \mathbf{\Delta}$ , using (12) and (14). Thus with  $\mathbf{P}$  being orthogonal,  $\mathbf{P}'\mathbf{A}\mathbf{P}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P}$  are diagonal.

Necessity: that if  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{D}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P} = \mathbf{\Delta}$  are diagonal, then  $\mathbf{AB} = \mathbf{BA}$ .

This is easily proved: with  $\mathbf{D}$  and  $\mathbf{\Delta}$  being diagonal,  $\mathbf{D}\mathbf{\Delta} = \mathbf{\Delta}\mathbf{D}$  and so  $\mathbf{A}\mathbf{B} = \mathbf{P}\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{P}'\mathbf{B}\mathbf{P} = \mathbf{P}\mathbf{D}\mathbf{\Delta}\mathbf{P}' = \mathbf{P}\mathbf{\Delta}\mathbf{D}\mathbf{P}' = \mathbf{P}\mathbf{P}'\mathbf{B}\mathbf{P}\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{B}\mathbf{A}$ . QED

**Theorem 2.** For symmetric matrices  $\mathbf{A}$  and  $\mathbf{B}$  of the same order, with  $\mathbf{A}$  being positive definite, there exists a nonsingular matrix  $\mathbf{P}$  (not necessarily orthogonal), such that  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{I}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P}$  is a diagonal matrix of the solutions for  $\lambda$  to  $|\mathbf{B} - \lambda\mathbf{A}| = 0$ .

**Proof.** When  $\mathbf{A}$  is positive definite, it can be expressed as  $\mathbf{A} = \mathbf{L}'^{-1}\mathbf{L}^{-1}$  for  $\mathbf{L}^{-1}$  existing and hence  $|\mathbf{L}| \neq 0$  (Section 7.7c). Also,  $\mathbf{L}'\mathbf{A}\mathbf{L} = \mathbf{I}$ . Therefore solutions  $\lambda_i$ ,  $i = 1, \dots, n$ , for  $\lambda$  to  $|\mathbf{L}'\mathbf{B}\mathbf{L} - \lambda\mathbf{I}| = 0$  are also solutions to  $|\mathbf{L}'||\mathbf{B} - \lambda\mathbf{A}||\mathbf{L}| = 0$ , i.e., to  $|\mathbf{B} - \lambda\mathbf{A}| = 0$ . Further, since  $\mathbf{L}'\mathbf{B}\mathbf{L}$  is symmetric there exists an orthogonal  $\mathbf{Q}$  such that  $\mathbf{Q}'\mathbf{L}'\mathbf{B}\mathbf{L}\mathbf{Q} = \text{diag}\{\lambda_i\}$ . Let  $\mathbf{P} = \mathbf{L}\mathbf{Q}$ . Then  $\mathbf{P}'\mathbf{B}\mathbf{P} = \text{diag}\{\lambda_i\}$  and  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{Q}'\mathbf{L}'\mathbf{A}\mathbf{L}\mathbf{Q} = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$ . QED

**Theorem 3.** For real symmetric matrices  $\mathbf{A}$  and  $\mathbf{B}$  of the same order and non-negative definite, there exists a nonsingular matrix  $\mathbf{P}$  such that  $\mathbf{P}'\mathbf{A}\mathbf{P}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P}$  are both diagonal.

**Proof.**<sup>2</sup> Let the order of  $\mathbf{A}$  and  $\mathbf{B}$  be  $n$ , with ranks  $a$  and  $b$ , respectively, and without loss of generality suppose  $b \geq a$ . Then because  $\mathbf{A}$  is symmetric and n.n.d., there exists, as in (9), a nonsingular matrix  $\mathbf{T}$  such that on defining  $\gamma = n - a$ ,

$$\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{D} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} \quad (16)$$

where, for clarity, a subscripted matrix is square, the subscript being its order. Denote  $\mathbf{T}'\mathbf{B}\mathbf{T}$  by

$$\mathbf{T}'\mathbf{B}\mathbf{T} = \begin{bmatrix} \mathbf{E}_a & \mathbf{F} \\ \mathbf{F}' & \mathbf{G}_\gamma \end{bmatrix};$$

It is n.n.d., of rank  $b$ . Suppose  $\mathbf{E}_a$  has rank  $r \leq a \leq b$ . Then, using the last  $\gamma$  rows and columns of  $\mathbf{T}'\mathbf{B}\mathbf{T}$ , perform elementary operations, to be represented by  $\mathbf{S}'$  and  $\mathbf{S}$

<sup>2</sup>This proof is based on Newcombe (1960).

respectively, so that for  $\beta = b - r$  and  $\delta = \gamma - \beta$  we reduce  $\mathbf{T}'\mathbf{B}\mathbf{T}$  to

$$\mathbf{S}'\mathbf{T}'\mathbf{B}\mathbf{T}\mathbf{S} = \begin{bmatrix} \mathbf{C}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_\delta \end{bmatrix}, \quad (17)$$

where  $\mathbf{C}_a$  has rank  $r$ . The operations are to be done in such a way that the first  $a$  rows (columns) of  $\mathbf{T}'\mathbf{B}\mathbf{T}$  are involved only by adding multiples of the other  $\gamma$  rows (columns) to them. Therefore performing the operations represented by  $\mathbf{S}$  on  $\mathbf{T}'\mathbf{A}\mathbf{T}$  of (16) will have no effect on (16), since all the last  $\gamma$  rows and columns of (16) are null; i.e.,

$$\mathbf{S}'\mathbf{T}'\mathbf{A}\mathbf{T}\mathbf{S} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix}. \quad (18)$$

Since  $\mathbf{C}_a$  of (17) is symmetric of rank  $r$ , there is a  $\mathbf{Q}_a$  such that

$$\mathbf{Q}'_a \mathbf{C}_a \mathbf{Q}_a = \mathbf{D}_a = \mathbf{D} \{ \lambda_1 \cdots \lambda_r, 0 \cdots 0 \} \quad \text{for } \mathbf{Q}_a \text{ orthogonal.} \quad (19)$$

Therefore on defining

$$\mathbf{R} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_\delta \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\gamma \end{bmatrix} \quad (20)$$

we have from (17), (19) and (20)

$$\mathbf{R}'\mathbf{S}'\mathbf{T}'\mathbf{B}\mathbf{T}\mathbf{S}\mathbf{R} = \begin{bmatrix} \mathbf{D}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_\delta \end{bmatrix} = \mathbf{\Delta}, \quad \text{diagonal;}$$

and from (18) and (20) we have

$$\mathbf{R}'\mathbf{S}'\mathbf{T}'\mathbf{A}\mathbf{T}\mathbf{S}\mathbf{R} = \begin{bmatrix} \mathbf{Q}'_a \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} = \mathbf{D}, \quad \text{diagonal.}$$

Hence for  $\mathbf{P} = \mathbf{T}\mathbf{S}\mathbf{R}$  we have  $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{D}$  and  $\mathbf{P}'\mathbf{B}\mathbf{P} = \mathbf{\Delta}$ , both diagonal. QED

### 3. THE CAYLEY-HAMILTON THEOREM

This is stated very simply.

**Theorem.** A matrix satisfies its own characteristic equation.

The theorem is true for any matrix, although the proof given here is for diagonalizable matrices only. Proof of the general case is available, for example, in Frazer, Duncan and Collar (1938).

**Proof** (for diagonalizable matrices). Writing  $t_1$  for  $-\text{tr}_1(\mathbf{A})$ ,  $t_2$  for  $\text{tr}_2(\mathbf{A})$ ,  $t_3$  for  $-\text{tr}_3(\mathbf{A})$  and so on, the characteristic equation for  $\mathbf{A}$  of order  $n$  is, from (14) of Chapter 11,

$$\lambda^n + t_1 \lambda^{n-1} + t_2 \lambda^{n-2} + \cdots + t_{n-1} \lambda + t_n = 0. \quad (21)$$

This is a scalar equation and can be multiplied by any vector,  $\mathbf{u}$  say, to give

$$\lambda^n \mathbf{u} + t_1 \lambda^{n-1} \mathbf{u} + \cdots + t_{n-1} \lambda \mathbf{u} + t_n \mathbf{u} = \mathbf{0}.$$

If  $\mathbf{u}$  is an eigenvector corresponding to an eigenvalue  $\lambda$ , use can be made of (9) of Chapter 11 to rewrite the foregoing as

$$\mathbf{A}^n \mathbf{u} + t_1 \mathbf{A}^{n-1} \mathbf{u} + \cdots + t_{n-1} \mathbf{A} \mathbf{u} + t_n \mathbf{u} = \mathbf{0}. \quad (22)$$

This will be true for all eigenvectors. But since they are of order  $n$ , are linearly independent (because  $\mathbf{A}$  is being assumed diagonalizable) and are  $n$  in number, any other  $n$ -order vector can be expressed as a linear combination of them. Therefore (22) holds for all  $n$ -order vectors and in particular for the columns of  $\mathbf{I}_n$ . Hence

$$\mathbf{A}^n + t_1 \mathbf{A}^{n-1} + \cdots + t_{n-1} \mathbf{A} + t_n \mathbf{I} = \mathbf{0};$$

i.e.,  $\mathbf{A}$  satisfies its own characteristic equation. QED

A useful application of this result is that when the characteristic equation is known, even if the eigenvalues are not, the  $n$ th and successive powers of  $\mathbf{A}$  can be obtained as polynomials of  $\mathbf{A}$  of order  $n - 1$ . This is also true for the inverse.

**Example.** For

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix}, \quad \mathbf{A}^2 = \begin{bmatrix} 22 & 27 \\ 18 & 31 \end{bmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{14} \begin{bmatrix} 5 & -3 \\ -2 & 4 \end{bmatrix}$$

and the characteristic equation of  $\mathbf{A}$  is

$$\lambda^2 - 9\lambda + 14 = 0.$$

The Cayley-Hamilton theorem is satisfied because

$$\mathbf{A}^2 - 9\mathbf{A} + 14\mathbf{I} = \begin{bmatrix} 22 & 27 \\ 18 & 31 \end{bmatrix} - 9 \begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix} + 14 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Polynomials for obtaining powers of  $\mathbf{A}$  of order 2 or more come from rewriting  $\mathbf{A}^2 - 9\mathbf{A} + 14\mathbf{I} = \mathbf{0}$  as  $\mathbf{A}^2 = 9\mathbf{A} - 14\mathbf{I}$ . Hence

$$\mathbf{A}^3 = 9\mathbf{A}^2 - 14\mathbf{A} = 9(9\mathbf{A} - 14\mathbf{I}) - 14\mathbf{A} = 67\mathbf{A} - 126\mathbf{I}$$

and

$$A^4 = 67A^2 - 126A = 67(9A - 14I) - 126A = 477A - 938I.$$

By this means recurrence relations can be established between the coefficients in successive powers of  $A$ , so enabling  $A^k$ , for  $k \geq n$ , to be written as a polynomial in  $A$  of order  $n - 1$ , the coefficients being functions of  $k$ . Thus for the above example it can be shown that, for  $k \geq 2$ ,

$$A^k = \left(\frac{1}{3}\right)(7^k - 2^k)A - \left(\frac{14}{3}\right)(7^{k-1} - 2^{k-1})I.$$

#### 4. THE SINGULAR-VALUE DECOMPOSITION

An analogy of the similar canonical form of square matrices is the singular-value decomposition of any matrix, square or rectangular. For a matrix  $A$ , it is based on the reductions to canonical form under orthogonal similarity of the symmetric matrices  $AA'$  and  $A'A$ , as given by the following theorem.

**Theorem.**  $A_{p \times q}$  of rank  $r$  can be expressed as

$$A = L \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} M', \quad (23)$$

where  $L$  and  $M$  are orthogonal and where  $\Delta^2$  is the diagonal matrix of the nonzero (positive) eigenvalues of  $AA'$  and of  $A'A$ ; and

$$L'AA'L = \begin{bmatrix} \Delta^2 & 0 \\ 0 & 0 \end{bmatrix}_{p \times p} \quad \text{and} \quad M'A'AM = \begin{bmatrix} \Delta^2 & 0 \\ 0 & 0 \end{bmatrix}_{q \times q}. \quad (24)$$

#### Comments

- (i) Both  $AA'$  and  $A'A$  are n.n.d. (Section 7.7e), and they have the same nonzero eigenvalues, which are all positive (Section 2b).
- (ii) Not all matrices in (24) have the same order:  $\Delta^2$  is  $r \times r$ ,  $L$  is  $p \times p$  and  $M$  is  $q \times q$ , and the matrices with leading submatrix  $\Delta^2$  are  $p \times p$  in one case and  $q \times q$  in the other.
- (iii) Matrices  $L$  and  $M$  of (24) always exist (Section 1d) but they are not unique, and not all pairs of them yield (23). However, for a particular choice of  $L$ , one is always able to construct an  $M$  (as shown in the following proof) such that (23) and (24) are satisfied.<sup>3</sup>

<sup>3</sup>I am grateful to G. P. H. Styan for bringing this point to my attention.

**Proof.**  $AA'$  and  $A'A$  are symmetric, and so  $L$  and  $M$  of (24) exist, each is orthogonal, and  $\Delta^2$  is a diagonal matrix of the  $r$  real, positive eigenvalues of  $AA'$  and of  $A'A$ . Partition  $L$  as  $[L_1 \ L_2]$  with  $L_1$  having  $r$  columns, and from the orthogonality of  $L$  get

$$L_1'L_1 = I_r \quad \text{and} \quad L_1'L_1 + L_2'L_2 = I_p. \quad (25)$$

Then from (24)

$$L_1'AA'L_1 = \Delta^2; \quad L_1'AA'L_2 = 0; \quad \text{and} \quad L_2'AA'L_2 = 0 \quad \text{implying} \quad L_2'A = 0, \quad (26)$$

the latter from using (7) of Section 3.1b. Define

$$M_1 = A'L_1\Delta^{-1}, \quad \text{of order } q \times r, \quad (27)$$

and choose  $M_2$  to be any matrix of order  $q \times (q - r)$  such that  $M = [M_1 \ M_2]$  for  $M_1$  of (27) is orthogonal. This can be done because, first, from (26) and (27)

$$M_1'M_1 = \Delta^{-1}L_1'AA'L_1\Delta^{-1} = \Delta^{-1}\Delta^2\Delta^{-1} = I_r; \quad (28)$$

and second, columns for  $M_2$  can be calculated as orthogonal solutions to  $M_1'x = 0$ , as in Lemma 2 of Section 9.2b. Then, using (25) and (27)

$$L \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} M' = L_1\Delta M_1' = L_1\Delta\Delta^{-1}L_1'A = L_1L_1'A = A,$$

which is (23). QED

This proof follows Stewart (1973, p. 319).

The diagonal elements of  $\Delta$  are known as the *singular values* of  $A$ . They are positive square roots of the (positive) eigenvalues of  $AA'$  (or equivalently, of  $A'A$ ).

Extensions of this theorem are discussed in Van Loan (1976).

**Example.** For

$$A = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad AA' = \begin{bmatrix} 6 & 2 & 4 \\ 2 & 6 & 4 \\ 4 & 4 & 4 \end{bmatrix} \quad \text{and} \quad A'A = \begin{bmatrix} 5 & 1 & 3 & 3 \\ 1 & 5 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix}$$

have nonzero eigenvalues 12 and 4. Also,  $AA'U = UD$  is

$$\begin{bmatrix} 6 & 2 & 4 \\ 2 & 6 & 4 \\ 4 & 4 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

and  $\mathbf{A}'\mathbf{A}\mathbf{U}^* = \mathbf{U}^*\mathbf{D}'$  is

$$\begin{bmatrix} 5 & 1 & 3 & 3 \\ 1 & 5 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 0 & -2 & 1 \\ 1 & 0 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 0 & -2 & 1 \\ 1 & 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

It is clear that  $\Delta^2 = \begin{bmatrix} 12 & 0 \\ 0 & 4 \end{bmatrix}$  and that  $\mathbf{U}$  and  $\mathbf{U}^*$  are un-normalized Helmer matrices [Section 3.4b(i)]. Normalizing them to get  $\mathbf{L}$  and  $\mathbf{M}$ , respectively, then gives  $\mathbf{A} =$

$$\frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \frac{1}{\sqrt{12}} \begin{bmatrix} \sqrt{3} & \sqrt{3} & \sqrt{3} & 3 \\ \sqrt{6} & -\sqrt{6} & \sqrt{0} & 0 \\ \sqrt{2} & \sqrt{2} & -2\sqrt{2} & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix}$$

as the singular value decomposition of  $\mathbf{A}$ .

An important and useful consequence of the singular-value decomposition is that for

$$\mathbf{A} = \mathbf{L} \begin{bmatrix} \Delta_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{M}, \quad \mathbf{A}^+ = \mathbf{M} \begin{bmatrix} \Delta_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{L}' \quad (29)$$

is the Moore–Penrose inverse.

### EXERCISES

1. Prove that idempotent matrices are diagonalizable.
2. Discuss the singular value decomposition of (a) symmetric idempotent matrices and (b) orthogonal matrices.

3. Find the singular-value decomposition of  $\begin{bmatrix} 10 & -5 \\ 2 & -11 \\ 6 & -8 \end{bmatrix}$ .

4. (a) Find the rank of  $\mathbf{M} = \begin{bmatrix} 1 & -3 & 5 \\ 0 & 1 & 7 \\ 2 & -4 & 28 \\ 3 & -13 & 0 \\ 4 & -14 & -3 \\ 5 & -9 & 30 \\ 6 & 2 & 173 \end{bmatrix}$ .

Without any further arithmetic, for  $\mathbf{M}$  given in part (a):

- (b) State an important feature of  $\mathbf{S}$  such that  $\mathbf{M}'\mathbf{M}\mathbf{S}\mathbf{M}'\mathbf{M} = \mathbf{M}'\mathbf{M}$ .
- (c) Find the rank of  $\mathbf{M}\mathbf{S}\mathbf{M}'$ .
- (d) Give the numerical values of the eigenvalues of  $\mathbf{M}\mathbf{S}\mathbf{M}'$ .
- (e) If  $\mathbf{T}$  is a generalized inverse of  $\mathbf{M}\mathbf{M}'$ , prove that  $\mathbf{M}'\mathbf{T}\mathbf{M}$  is an identity matrix.
- (f) Suggest two different generalized inverses of  $\mathbf{M}'$ .

5. Suppose

$$\mathbf{K} = \mathbf{K}', \quad \mathbf{K} = \mathbf{K}^3, \quad \mathbf{K}\mathbf{1} = \mathbf{0}, \quad \text{and} \quad \mathbf{K} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}.$$

Calculate the following values, in each case giving reasons as to why the value can be calculated *without calculating*  $\mathbf{K}$ .

- (a) The order of  $\mathbf{K}$ .
  - (b) The rank of  $\mathbf{K}$ .
  - (c) The trace of  $\mathbf{K}$ .
  - (d) The determinant of  $\mathbf{K}$ .
  - (e) The trace of  $\mathbf{K}^{26}$ .
  - (f) The determinant of  $\mathbf{K}^{26}$ .
  - (g) The trace of  $6\mathbf{K}^{60} - 7\mathbf{K}^{37} + 3\mathbf{I}$ .
  - (h) The determinant of  $6\mathbf{K}^{60} - 7\mathbf{K}^{37} + 3\mathbf{I}$ .
6. For  $\mathbf{M}_1$  of (27), prove that  $\mathbf{M}'_1\mathbf{A}'\mathbf{M}_1 = \Delta^2$ .
  7. (a) Prove that  $\mathbf{A}^+$  of (29) is the Moore–Penrose inverse of  $\mathbf{A}$ .  
(b) Show that a solution of  $\mathbf{A}\mathbf{x} = \mathbf{y}$  is  $\mathbf{A}^+\mathbf{y}$ .

## CHAPTER 12

### MISCELLANEA

This chapter contains miscellaneous topics useful in applications of matrix algebra. Some sections summarize topics scattered throughout earlier chapters; other sections introduce new topics.

#### 1. ORTHOGONAL MATRICES—A SUMMARY

- a. The matrix  $A$  is orthogonal when any two of the statements (i)  $A$  is square, (ii)  $AA' = I$  and (iii)  $A'A = I$  are true; and any two imply the third (Section 5.7d).
- b. Rows of an orthogonal matrix are orthonormal (as are columns); see Section 3.4.
- c. Products of orthogonal matrices are orthogonal (Exercise 3.13).
- d. Determinants of orthogonal matrices are  $+1$  or  $-1$  (Section 4.3e). Hence orthogonal matrices are nonsingular; and  $A' = A^{-1}$  (Section 5.7d).
- e. If  $\lambda$  is an eigenvalue of an orthogonal matrix, so is  $1/\lambda$  (Exercise 11.17).
- f. When  $A$  is skew-symmetric (Section 3.1f),  $(I - A)(I + A)^{-1}$  is orthogonal (Exercise 5.13).
- g. Helmert, Givens and Householder matrices are special forms of orthogonal matrices (Section 3.4b and Exercises 3.20, 3.21 and 3.22).

#### 2. IDEMPOTENT MATRICES—A SUMMARY

The following properties hold for an idempotent matrix  $A$ .

- a. The definition of idempotency is  $A^2 = A$  (Section 3.3).
- b. All idempotent matrices (except  $I$ ) are singular (Exercise 5.11).
- c. The rank and trace of  $A$  are equal:  $r_A = \text{tr}(A)$ .

[12.2]

- d.  $I_n - A$  is idempotent, with  $r(I_n - A) = n - r_A$ .
- e. All symmetric, idempotent matrices (except  $I$ ) are p.s.d. (because of the theorem in Section 7.7e).
- f. The product of two idempotent matrices is idempotent if the matrices commute in multiplication (Exercise 3.14).
- g. Eigenvalues of an idempotent matrix are all 0 or 1. This is so because  $Au = \lambda u$  and  $A^2u = \lambda^2u$ , so that when  $A^2 = A$ , we have  $\lambda u = \lambda^2u$ , and since  $u \neq 0$  this means  $\lambda = \lambda^2$ , i.e.,  $\lambda = 0$  or 1.

Although idempotent matrices have eigenvalues that are 0 or 1, the converse is not generally true. That is, eigenvalues of 0 and 1 do not always imply idempotency; they do for symmetric matrices (proof of which is left for the reader in Exercise 8) but not necessarily for nonsymmetric matrices. For example,

$$B = \frac{1}{8} \begin{bmatrix} 11 & 3 & 3 \\ 1 & 1 & 1 \\ -12 & -4 & -4 \end{bmatrix}, \text{ with } B^2 = \frac{1}{8} \begin{bmatrix} 11 & 3 & 3 \\ 0 & 0 & 0 \\ -11 & -3 & -3 \end{bmatrix} \neq B$$

has eigenvalues 1, 0 and 0 but is not idempotent.

- h. There are  $r_A$  eigenvalues of unity and  $z_A = n - r_A$  of zero. This result comes from Section 11A.1b. For the eigenvalue  $\lambda = 0$ , that section gives  $r(A - \lambda I) \geq n - z_A$ , i.e.,  $r_A \geq n - z_A$ . Hence  $n - r_A \leq z_A$ . And for  $\lambda = 1$  it gives  $r(A - I) \geq n - (n - z_A)$ , and so, using property d,  $n - r_A \geq z_A$ . Hence  $z_A = n - r_A$ .
- i. All idempotent matrices are diagonalizable in the form

$$U^{-1}AU = D = \begin{bmatrix} I_{r_A} & 0 \\ 0 & 0 \end{bmatrix}.$$

(From item h, the conditions of the diagonalizability theorem of Section 11.5a are seen to be satisfied.)

- j. A symmetric idempotent matrix  $A$  can be expressed as  $A = LL'$  for  $L'L = I_{r_A}$  (Exercise 11.18).
- k. A special form of idempotent matrix which has many properties useful in statistical applications and in general matrix theory is  $P = I - X(X'X)^{-1}X'$  with properties shown in Exercise 8.37.
- l. Matrices  $A$  satisfying  $A^2 = A$  are idempotent; similarly, those satisfying  $A^2 = 0$  are called *nilpotent* matrices, and those for which  $A^2 = I$  could be called *unipotent*.

#### Examples

$$A_1 = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \text{ is nilpotent; } B = \begin{bmatrix} I & X \\ 0 & -I \end{bmatrix} \text{ is unipotent.}$$

Variations on these definitions are  $A^k = A$ ,  $A^k = \mathbf{0}$ , or  $A^k = \mathbf{I}$  for some positive integer  $k > 2$ . An example is the matrix of Bernadelli's beetles (Section 11.2) where  $A^3 = \mathbf{I}$  but  $A^2 \neq \mathbf{I}$ .

### 3. THE MATRIX $a\mathbf{I} + b\mathbf{J}$ —A SUMMARY

The summing vector  $\mathbf{1}_n$  and allied matrix  $\mathbf{J}_{r \times s} = \mathbf{1}_r \mathbf{1}'_s$ , both having all elements unity, have been introduced in Section 2.2. A useful variation of  $\mathbf{J}_n \equiv \mathbf{J}_{n \times n}$  is that of combining it with an identity matrix in the form  $a\mathbf{I} + b\mathbf{J}$  for scalars  $a$  and  $b$ . An example is

$$a\mathbf{I}_3 + b\mathbf{J}_3 = \begin{bmatrix} a+b & b & b \\ b & a+b & b \\ b & b & a+b \end{bmatrix}.$$

Useful properties are as follows.

- a.  $(a\mathbf{I}_n + b\mathbf{J}_n)$ -matrices have products of the same form:

$$(a\mathbf{I} + b\mathbf{J})(\alpha\mathbf{I} + \beta\mathbf{J}) = \alpha a\mathbf{I} + (a\beta + \alpha b + nb\beta)\mathbf{J}.$$

- b.  $(a\mathbf{I}_n + b\mathbf{J}_n)^{-1} = \frac{1}{a} \left( \mathbf{I}_n - \frac{b}{a+nb} \mathbf{J}_n \right)$  for  $a \neq 0$  and  $a \neq -nb$ .

- c.  $|a\mathbf{I}_n + b\mathbf{J}_n| = a^{n-1}(a+nb)$ .

- d. Eigenvalues of  $a\mathbf{I}_n + b\mathbf{J}_n$  are  $a$  and  $a+nb$ , with multiplicities  $n-1$  and 1, respectively.

- e.  $a\mathbf{I}_n + b\mathbf{J}_n$  is diagonalizable, with matrix of eigenvectors  $\mathbf{H}'_n$ , for  $\mathbf{H}_n$  being the Helmert matrix of Section 3.4b.

- f.  $\bar{\mathbf{J}}_n = \mathbf{J}_n/n$  is idempotent.

The centering matrix [(17), Chapter 2]

$$\mathbf{C}_n \equiv \mathbf{I} - \bar{\mathbf{J}}_n \quad \text{with} \quad \sum_{i=1}^n (x_i - \bar{x})^2 = \mathbf{x}'\mathbf{C}_n\mathbf{x}$$

has many uses in statistics. It is symmetric and idempotent (with rank  $n-1$ ),  $\mathbf{C} = \mathbf{C}' = \mathbf{C}^2$ , with  $\mathbf{C}_n\mathbf{J}_n = \mathbf{0}$ .

### 4. NON-NEGATIVE DEFINITE MATRICES—A SUMMARY

Only if a matrix is square can it have an inverse; and it does have an inverse only if its determinant is nonzero (Section 5.5). A rectangular matrix

$A_{r \times c}$  can have a left inverse only if  $r \geq c$  (Theorem, Section 5.10) and it does have a left inverse only if it has full column rank,  $r_A = c$  (Lemma 8, Section 7.5); and one form of left inverse is then  $(A'A)^{-1}A'$ , as in Corollary 2, Section 7.7e. The existence of  $(A'A)^{-1}$  in this case is assured by the full column rank property of  $A$  (Theorem, Section 7.7e).

Non-negative definite (n.n.d.) matrices, i.e., positive definite and positive-semidefinite matrices, are defined in Section 3.6. As indicated there, they are usually taken as being symmetric.

- For real  $\mathbf{X}$ , the product  $\mathbf{X}'\mathbf{X}$  is n.n.d.; it is p.d. if  $\mathbf{X}$  has full column rank (Theorem, Section 7.7e).
- For  $\mathbf{N}$  being n.n.d. (and symmetric),  $|\mathbf{N}| \geq 0$ , its principal leading minors, its diagonal elements and its eigenvalues are all zero or positive; and all are strictly positive for p.d. matrices. (See Sections 7.7a and 11A.2b.) Also,  $\mathbf{N} = \mathbf{A}\mathbf{A}'$  for some  $\mathbf{A}$  of full column rank (Section 7.7c) for which, by property a,  $\mathbf{A}'\mathbf{A}$  is p.d. and  $(\mathbf{A}'\mathbf{A})^{-1}$  exists; and the Moore-Penrose inverse of  $\mathbf{N}$  is  $\mathbf{N}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-2}\mathbf{A}'$  (Section 8.6a).
- All n.n.d. matrices are diagonalizable—because they are symmetric (Section 11.6b).

Finally, any symmetric matrix  $\mathbf{S}$  can be expressed as  $\mathbf{S} = \mathbf{K}\mathbf{K}'$  for  $\mathbf{K}$  of full column rank equal to the rank of  $\mathbf{S}$ ; but  $\mathbf{K}$  may not be real (Section 7.6e). When the congruent diagonal form of  $\mathbf{S}$  contains no negatives,  $\mathbf{K}$  is real and then  $\mathbf{S}$  is n.n.d. by property a; and if, further,  $\mathbf{S}$  is nonsingular,  $\mathbf{K}$  will be also and by property b,  $\mathbf{S}$  is then p.d.

### 5. CANONICAL FORMS AND OTHER DECOMPOSITIONS— A SUMMARY

We summarize here the four canonical forms, the singular value decomposition of a matrix and the spectral decomposition of a symmetric matrix.

**Equivalent canonical form.** The equivalent canonical form of any matrix  $\mathbf{A}$  of rank  $r$  is

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

for  $\mathbf{P}$  and  $\mathbf{Q}$  nonsingular (Section 7.4c).

**Congruent canonical form.** The diagonal form of a symmetric matrix of rank  $r$  is

$$\mathbf{PAP}' = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{for } \mathbf{A} = \mathbf{A}'$$

with  $\mathbf{P}$  being nonsingular  $\mathbf{P}$  (Section 7.6b), and with  $\mathbf{D}_r$  diagonal, of order  $r$ . The congruent canonical form (Section 7.6c) is

$$\mathbf{RAR}' = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{for } \mathbf{A} = \mathbf{A}'.$$

$\mathbf{P}$  is always real, but  $\mathbf{R}$  will not be real when any element of  $\mathbf{D}_r$  is negative, ( $\mathbf{R}$  is always real for n.n.d. matrices.)

**Similar canonical form.** The similar canonical form of any matrix  $\mathbf{A}$  is  $\mathbf{D}(\lambda)$ , a diagonal matrix of the eigenvalues of  $\mathbf{A}$ , given (Section 11.5a) by

$$\mathbf{AU} = \mathbf{UD}(\lambda)$$

where  $\mathbf{U}$  is the matrix of eigenvectors corresponding to the eigenvalues of  $\mathbf{D}(\lambda)$ . When the diagonalizability theorem (Section 11.5a) is satisfied,  $\mathbf{U}$  is nonsingular and

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}(\lambda).$$

**Orthogonal similar canonical form.** With the same notation as in the preceding paragraph, the canonical form under orthogonal similarity for any symmetric matrix  $\mathbf{A}$  is (Section 11.6c)

$$\mathbf{U}'\mathbf{AU} = \mathbf{D}(\lambda), \quad \text{with } \mathbf{A} = \mathbf{A}' \text{ and orthogonal } \mathbf{U}.$$

$\mathbf{U}$  always exists. The eigenvalues are always real and for n.n.d. matrices they are either positive or zero (all positive for p.s.d. matrices).

**Spectral decomposition.** The spectral decomposition of a symmetric matrix  $\mathbf{A}$  is (Section 11A.2a)

$$\mathbf{A} = \sum \lambda_i \mathbf{u}_i \mathbf{u}_i' \quad \text{for } \mathbf{A} = \mathbf{A}' \text{ and } \mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{u}_i,$$

i.e., for  $\lambda_i$  and  $\mathbf{u}_i$  being an eigenvalue and corresponding eigenvector.

**Singular-value decomposition.** The singular-value decomposition of any matrix  $\mathbf{A}$  is (Section 11A.4)

$$\mathbf{A} = \mathbf{L} \begin{bmatrix} \mathbf{\Delta}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{M}'$$

where  $\mathbf{L}$  and  $\mathbf{M}$  are orthogonal, and  $\mathbf{\Delta}_r$  is the diagonal matrix of positive square roots of the (positive) eigenvalues of  $\mathbf{A}'\mathbf{A}$  (or equivalently of  $\mathbf{A}\mathbf{A}'$ ):  $\mathbf{\Delta}_r = \sqrt{\mathbf{\Delta}^2}$  where

$$\mathbf{L}'\mathbf{A}\mathbf{A}'\mathbf{L} = \begin{bmatrix} \mathbf{\Delta}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{M}'\mathbf{A}'\mathbf{A}\mathbf{M} = \begin{bmatrix} \mathbf{\Delta}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

## 6. MATRIX FUNCTIONS

### a. Function of matrices

We have mentioned (Section 5.8b) the infinite (when convergent) sum  $\sum_{i=0}^{\infty} \mathbf{A}^i$  for  $(\mathbf{I} - \mathbf{A})^{-1}$ . Other functions of matrices which seem at first sight to have no meaning can, in fact, be defined in a similar manner. For example,  $e^{\mathbf{A}}$  is defined by adapting the power series for a scalar,

$$e^x = 1 + x + x^2/2! + x^3/3! + \dots,$$

to define  $e^{\mathbf{A}}$  as

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2/2! + \mathbf{A}^3/3! + \dots = \sum_{i=0}^{\infty} \mathbf{A}^i/i!, \quad (1)$$

provided the series is convergent. This definition requires  $\mathbf{A}$  to be square, in which case  $e^{\mathbf{A}}$  has the same order as  $\mathbf{A}$ . Turner et al. (1963) use this formulation in studying concentrations of an injected substance in the bloodstream, urine and kidney, where  $\mathbf{A}$  has elements relating to the transport of the substance between various parts of the body. Pielou (1964), in studying patch-and-gap vegetation in a forest, uses transition probability matrices  $\mathbf{R}$  and  $\mathbf{P}$  in the relationship  $\mathbf{R} = \log_e \mathbf{P}$ . Rewriting this as  $\mathbf{P} = e^{\mathbf{R}}$ , equation (1) can be used for  $e^{\mathbf{R}}$  to establish relationships between  $p_{ij}$  and  $r_{ij}^{(k)}$ , the  $ij$ th elements of  $\mathbf{P}$  and  $\mathbf{R}^k$ , respectively:

$$p_{ij} = \delta_{ij} + \sum_{k=1}^{\infty} r_{ij}^{(k)}/k!,$$

with  $\delta_{ij} = 1$  for  $i = j$  and  $\delta_{ij} = 0$  otherwise. Pielou's matrices are of order 2, so that this relationship has a relatively simple form (see Exercise 17).

### b. Matrices of functions

This book is concerned primarily with matrices that have real numbers as elements. But matrices can have elements that are functions of one variable  $x$ , or of several variables  $x_1, x_2, \dots, x_r$ , which can be represented as  $\mathbf{x}$ . Using

$F$  to represent matrices of such elements, their functional relationship to  $x$  or  $\mathbf{x}$  can be represented by  $F(x)$  and  $F(\mathbf{x})$ , respectively. Simple examples are

$$F(x) = \begin{bmatrix} x^3 & e^x \\ 2^x & 3 + 4x^2 \end{bmatrix}$$

and, for

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad F(\mathbf{x}) = \begin{bmatrix} x_1^2 & x_1 x_2 & x_2^2 \\ x_1 x_2 & x_2^3 & x_1 + x_2 \end{bmatrix}.$$

This notation is used in the next section.

## 7. ITERATIVE SOLUTION OF NONLINEAR EQUATIONS

The solving of linear equations using matrix techniques is considered at length in Chapter 9. Matrices can also be used in formulating iterative solutions of nonlinear equations, as is illustrated by the following elementary development of one particular method.

Suppose equations to be solved are

$$f_i(x_1, x_2, \dots, x_n) = 0 \quad \text{for } i = 1, \dots, n$$

where  $f_i(x_1, x_2, \dots, x_n)$  is a nonlinear function of  $n$  unknowns  $x_1, x_2, \dots, x_n$ . Denoting the  $x$ 's by  $\mathbf{x}$  and the vector of functions  $f_i(\mathbf{x})$  by  $\mathbf{f}(\mathbf{x})$ , the equations can be written succinctly as

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}. \quad (2)$$

Define the  $n \times n$  matrix

$$G(\mathbf{x}) = \{g_{ij}(\mathbf{x})\} = \left\{ \frac{\partial}{\partial x_j} f_i(\mathbf{x}) \right\} \quad \text{for } i, j = 1, 2, \dots, n.$$

We suppose the functions  $f_i(x)$  are sufficiently complicated that (2) cannot be solved explicitly for  $\mathbf{x}$ . Let  $\mathbf{x}_0$  be some first guess at a solution. Then a succession of improvements on that guess (i.e., successive iterative solutions), with  $\mathbf{x}_r$  being the  $r$ th improvement, can be derived using the equation

$$\mathbf{f}(\mathbf{x}_{r+1}) = \mathbf{f}(\mathbf{x}_r) + G(\mathbf{x}_r)\Delta_r \quad (3)$$

where

$$\Delta_r = \mathbf{x}_{r+1} - \mathbf{x}_r. \quad (4)$$

To a first approximation, (3) is a multivariable extension of Taylor's expansion.

Starting with  $\mathbf{x}_0$ , i.e., with  $r = 0$ , we use (3) to derive  $\Delta_r$  and thence obtain  $\mathbf{x}_{r+1}$  from (4) as  $\mathbf{x}_{r+1} = \mathbf{x}_r + \Delta_r$ . This is done by noting that were  $\mathbf{x}_{r+1}$  a solution to (2), then (3) would be  $\mathbf{0} = \mathbf{f}(\mathbf{x}_r) + G(\mathbf{x}_r)\Delta_r$ , so giving

$$\Delta_r = -[G(\mathbf{x}_r)]^{-1}\mathbf{f}(\mathbf{x}_r), \quad (5)$$

providing  $G(\mathbf{x}_r)$  is nonsingular. With this,

$$\mathbf{x}_{r+1} = \mathbf{x}_r + \Delta_r \quad (6)$$

Equation (5) is first used with  $\mathbf{x}_r$  being given the value  $\mathbf{x}_0$ , and then (6) gives  $\mathbf{x}_1$ ; using this for  $\mathbf{x}_r$  in (5) and then (6) again gives  $\mathbf{x}_2$ ; and so on until, for some integer  $r$ , a value for  $\Delta_r$  is obtained that is sufficiently close to  $\mathbf{0}$  for us to be satisfied that  $\mathbf{x}_r$  is, to that same degree of approximation, a solution of (2). Consideration must of course be given to matters such as convergence to a solution, the rate of convergence, multiple solutions, arithmetic stability and so on. These are problems for the numerical analyst and are beyond the scope of this book. Development of (5) and (6) are given solely as illustration of the uses of matrix notation.

## 8. VECTORS OF DIFFERENTIAL OPERATORS

This section and the next make simple use of, and should be omitted by readers not familiar with, the differential calculus.

### a. Scalars

Sometimes a function of several variables has to be differentiated with respect to each of the variables concerned. In such cases it is convenient to write the derivatives as a column vector. A common use of this involves only linear and quadratic functions of variables, so we confine ourselves to these. For example, if

$$\lambda = 3x_1 + 4x_2 + 9x_3$$

where  $\lambda$ ,  $x_1$ ,  $x_2$  and  $x_3$  are scalars, we write the derivatives of  $\lambda$  with respect to  $x_1$ ,  $x_2$  and  $x_3$  as

$$\frac{\partial \lambda}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \lambda}{\partial x_1} \\ \frac{\partial \lambda}{\partial x_2} \\ \frac{\partial \lambda}{\partial x_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{bmatrix} \lambda = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}.$$



The second vector shows how the symbol  $\frac{\partial}{\partial \mathbf{x}}$  can represent a whole vector of differential operators. Noting that

$$\lambda = 3x_1 + 4x_2 + 9x_3 = [3 \quad 4 \quad 9] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{a}'\mathbf{x},$$

so defining vectors  $\mathbf{a}$  and  $\mathbf{x}$ , we therefore see that with  $\frac{\partial}{\partial \mathbf{x}}$  representing a column of operators the differential of the scalar  $\mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$  is

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}'\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{a}) = \mathbf{a}. \quad (7)$$

### b. Vectors<sup>1</sup>

Suppose the principle of (7) is applied to each element of  $\mathbf{y}' = \mathbf{x}'\mathbf{A}$  expressed as

$$\mathbf{y}' = [y_1 \quad y_2 \quad \cdots \quad y_n] = [\mathbf{x}'\mathbf{a}_1 \quad \mathbf{x}'\mathbf{a}_2 \quad \cdots \quad \mathbf{x}'\mathbf{a}_n] \quad (8)$$

where the  $i$ th element of  $\mathbf{y}$  is  $y_i = \mathbf{x}'\mathbf{a}_i$  for  $\mathbf{a}_i$  being the  $i$ th column of  $\mathbf{A}$ , with  $i = 1, \dots, n$ . Applying (7) to each element of (8) gives

$$\begin{aligned} \frac{\partial \mathbf{y}'}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} & \frac{\partial y_2}{\partial \mathbf{x}} & \cdots & \frac{\partial y_n}{\partial \mathbf{x}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial \mathbf{x}'\mathbf{a}_1}{\partial \mathbf{x}} & \frac{\partial \mathbf{x}'\mathbf{a}_2}{\partial \mathbf{x}} & \cdots & \frac{\partial \mathbf{x}'\mathbf{a}_n}{\partial \mathbf{x}} \end{bmatrix} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n] = \mathbf{A}. \end{aligned}$$

Thus

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}) = \mathbf{A}. \quad (9)$$

In differentiating the row vector  $\mathbf{y}' = \mathbf{x}'\mathbf{A}$ , the columns  $\partial y_i / \partial \mathbf{x}$  are, in deriving (9), set alongside one another. We do the same in differentiating the column vector  $\mathbf{y} = \mathbf{A}\mathbf{x}$ ; otherwise, the logical procedure of putting the columns  $\partial y_i / \partial \mathbf{x}$  one under the other would yield a vector having the same order as  $\mathbf{x} \otimes \mathbf{y}$ . The convention is therefore adopted that by  $\partial(\mathbf{A}\mathbf{x}) / \partial \mathbf{x}$  we

<sup>1</sup>Thanks go to G. P. H. Styan for suggesting improved notation for this section.

mean  $\partial(\mathbf{A}\mathbf{x})' / \partial \mathbf{x}$  and so, by (9)

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) \equiv \frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x})' = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}') = \mathbf{A}'. \quad (10)$$

For example, for  $\mathbf{y} = \mathbf{A}\mathbf{x}$  with  $\mathbf{A}$  of order  $3 \times 3$ ,

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \frac{\partial y_3}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_3}{\partial x_2} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_2}{\partial x_3} & \frac{\partial y_3}{\partial x_3} \end{bmatrix} = \mathbf{A}'. \quad (11)$$

**Example.** Applying (11) to

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} 2 & 6 & -1 \\ 3 & -2 & 4 \\ 3 & 4 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 6x_2 - x_3 \\ 3x_1 - 2x_2 + 4x_3 \\ 3x_1 + 4x_2 + 7x_3 \end{bmatrix}$$

gives  $\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x})$

$$\begin{aligned} &= \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}}(2x_1 + 6x_2 - x_3) & \frac{\partial}{\partial \mathbf{x}}(3x_1 - 2x_2 + 4x_3) & \frac{\partial}{\partial \mathbf{x}}(3x_1 + 4x_2 + 7x_3) \end{bmatrix} \\ &= \begin{bmatrix} 2 & 3 & 3 \\ 6 & -2 & 4 \\ -1 & 4 & 7 \end{bmatrix} = \mathbf{A}'. \end{aligned}$$

### c. Quadratic forms

Adapting the principle of differentiation by parts, the differential of  $\mathbf{x}'\mathbf{A}\mathbf{x}$  with respect to  $\mathbf{x}$  is

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{P}) + \frac{\partial}{\partial \mathbf{x}}(\mathbf{Q}\mathbf{x}) \quad \text{for } \mathbf{P} = \mathbf{A}\mathbf{x} \quad \text{and} \quad \mathbf{Q} = \mathbf{x}'\mathbf{A} \\ &= \mathbf{P} + \mathbf{Q}', \quad \text{from (9) and (10)} \\ &= \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x}. \end{aligned} \quad (12)$$

An alternative derivation is

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{Ax}) &= \left\{ \frac{\partial}{\partial x_i} \sum_{r=1}^n \sum_{s=1}^n a_{rs} x_r x_s \right\}, \text{ for } i = 1, \dots, n \\ &= \sum_{r=1}^n \sum_{s=1}^n (a_{rs} \delta_{ir} x_s + a_{rs} x_r \delta_{is})\end{aligned}$$

for  $\delta_{ir} = 1$  when  $r = i$  and  $\delta_{ir} = 0$  when  $r \neq i$ . Hence

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{Ax}) &= \left\{ \sum_{s=1}^n a_{is} x_s + \sum_{r=1}^n a_{ri} x_r \right\}, \text{ for } i = 1, \dots, n \\ &= \mathbf{Ax} + \mathbf{A}'\mathbf{x}.\end{aligned}$$

Usually  $\mathbf{A}$  is taken as symmetric,  $\mathbf{A} = \mathbf{A}'$ , in which case

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{Ax}) = 2\mathbf{Ax}. \quad (13)$$

**Example.** For

$$\begin{aligned}q = \mathbf{x}'\mathbf{Ax} &= [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 1 & 3 & 5 \\ 3 & 4 & 7 \\ 5 & 7 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= x_1^2 + 6x_1x_2 + 10x_1x_3 + 4x_2^2 + 14x_2x_3 + 9x_3^2, \\ \frac{\partial q}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial q}{\partial x_1} \\ \frac{\partial q}{\partial x_2} \\ \frac{\partial q}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 + 6x_2 + 10x_3 \\ 6x_1 + 8x_2 + 14x_3 \\ 10x_1 + 14x_2 + 18x_3 \end{bmatrix} = 2\mathbf{Ax}.\end{aligned}$$

**Illustration.** The method of least squares in statistics involves minimizing the sum of squares of the elements of the vector  $\mathbf{e}$  in the equation

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}. \quad (14)$$

If  $\mathbf{e} = \{e_i\}$  for  $i = 1, 2, \dots, n$ , the sum of squares is

$$\begin{aligned}S &= \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{Xb} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}\end{aligned}$$

and because  $\mathbf{y}'\mathbf{Xb}$  is a scalar it equals its transpose,  $\mathbf{b}'\mathbf{X}'\mathbf{y}$ , and hence

$$S = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}.$$

Minimizing this with respect to elements of  $\mathbf{b}$  involves equating to zero the expression  $\partial S / \partial \mathbf{b}$  where  $\partial / \partial \mathbf{b}$  is a vector of differential operators of the nature just discussed. Using results (9) and (13) gives

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{Xb},$$

and equating this to a null vector leads to the consistent equations

$$\mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{y} \quad (15)$$

with solution

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}. \quad (16)$$

**Illustration.** The selection index method of genetics as proposed by Hazel (1943) can be formulated as follows. Let  $\mathbf{x}$  be a vector of observations (phenotypes) on certain traits in an animal and  $\mathbf{b}$  a vector of coefficients by which these observations get combined into an index  $I = \mathbf{b}'\mathbf{x}$ . Corresponding to the traits represented in  $\mathbf{x}$ , let  $\mathbf{g}$  be a vector of (unknown) genetic values and let  $\mathbf{a}$  be a vector of the relative economic values of these traits. Then the coefficients represented by  $\mathbf{b}$  are determined by maximizing, for values of  $\mathbf{b}$ , the correlation between  $I = \mathbf{b}'\mathbf{x}$  and the combined genetic value  $H = \mathbf{a}'\mathbf{g}$ . This correlation is

$$r_{IH} = \frac{\text{cov}(IH)}{\sigma_I \sigma_H} = \frac{\mathbf{b}'\mathbf{Ga}}{\sqrt{\mathbf{b}'\mathbf{Pb}} \sqrt{\mathbf{a}'\mathbf{Ga}}},$$

where  $\mathbf{P}$  is the matrix of phenotypic variances and covariances, namely the variance-covariance matrix appropriate to the vector of phenotypic values  $\mathbf{x}$ ; and  $\mathbf{G}$  is the matrix of genetic variances and covariances, the variance-

covariance matrix corresponding to the genetic values  $\mathbf{g}$ . Maximizing this with respect to the elements in  $\mathbf{b}$  is achieved by maximizing  $\mathbf{b}'\mathbf{G}\mathbf{a}/\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}$ . Equating to zero the derivative of this with respect to  $\mathbf{b}$ , using results (9) and (13) to carry out the differentiation,  $\mathbf{P}$  and  $\mathbf{G}$  being symmetric, gives

$$(\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}})\mathbf{G}\mathbf{a} = \frac{1}{2}(\mathbf{b}'\mathbf{P}\mathbf{b})^{-\frac{1}{2}}2\mathbf{P}\mathbf{b}(\mathbf{b}'\mathbf{G}\mathbf{a});$$

i.e.,

$$(\mathbf{b}'\mathbf{G}\mathbf{a})\mathbf{P}\mathbf{b} = (\mathbf{b}'\mathbf{P}\mathbf{b})\mathbf{G}\mathbf{a}.$$

Apart from a scalar, all solutions to this equation are proportional to solutions to  $\mathbf{P}\mathbf{b} = \mathbf{G}\mathbf{a}$ . The required value of  $\mathbf{b}$  is therefore taken as the solution of the latter equation:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}. \quad (17)$$

The corresponding (maximized) value of  $r_{IH}$  is then  $r_{IH} = \sqrt{\frac{\mathbf{b}'\mathbf{G}\mathbf{a}}{\mathbf{a}'\mathbf{G}\mathbf{a}}}$ .

Estimates of  $\mathbf{b}$  and  $r_{IH}$  are obtained by using a predetermined value of  $\mathbf{a}$  and estimates of the phenotypic and genetic variances and covariances involved in  $\mathbf{P}$  and  $\mathbf{G}$ .

## 9. VEC AND VECH OPERATORS

### a. Definitions

A matrix operation dating back nearly a century, but in which there is currently great renewed interest (see Henderson and Searle, 1979), is that of stacking the columns of a matrix one under the other to form a single column. Over the years it has had a variety of names, the most recent being "vec." Thus for

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \text{vec } \mathbf{X} = \begin{bmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{bmatrix}.$$

The notations  $\text{vec } \mathbf{X}$  and  $\text{vec}(\mathbf{X})$  are used interchangeably, the latter only when parentheses are deemed necessary.

An extension of  $\text{vec } \mathbf{X}$  is  $\text{vech } \mathbf{X}$ , defined in the same way that  $\text{vec } \mathbf{X}$  is, except that for each column of  $\mathbf{X}$  only that part of it which is on or below the diagonal of  $\mathbf{X}$  is put into  $\text{vech } \mathbf{X}$  (vector-half of  $\mathbf{X}$ ). In this way, for  $\mathbf{X}$

symmetric,  $\text{vech } \mathbf{X}$  contains only the distinct elements of  $\mathbf{X}$ ; e.g., for

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 6 \\ 7 & 3 & 8 \\ 6 & 8 & 2 \end{bmatrix} = \mathbf{X}', \quad \text{vech } \mathbf{X} = \begin{bmatrix} 1 \\ 7 \\ 6 \\ 3 \\ 8 \\ 2 \end{bmatrix}.$$

Henderson and Searle (1979) give history, properties and many applications of these operators. Only a few highlights are summarized here.

### b. Properties of vec

The following three theorems give useful properties of the vec operator. Proofs of the first two depend on the elementary vector  $\mathbf{e}_i$ , the  $i$ th column of an identity matrix, discussed in Section 3.1e.

**Theorem 1.**  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec } \mathbf{B}$ .

**Proof.** The  $j$ th column of  $\mathbf{ABC}$ , and hence the  $j$ th subvector of  $\text{vec}(\mathbf{ABC})$  is, for  $\mathbf{C}$  having  $r$  rows,

$$\mathbf{ABCe}_j = \mathbf{AB} \sum_i \mathbf{e}_i \mathbf{e}_i' \mathbf{C}\mathbf{e}_j = \sum_i c_{ij} \mathbf{ABe}_i = [c_{1j}\mathbf{A} \quad c_{2j}\mathbf{A} \quad \cdots \quad c_{rj}\mathbf{A}] \begin{bmatrix} \mathbf{Be}_1 \\ \mathbf{Be}_2 \\ \vdots \\ \mathbf{Be}_r \end{bmatrix},$$

which is the  $j$ th subvector of  $(\mathbf{C}' \otimes \mathbf{A})\text{vec } \mathbf{B}$ .

QED

**Theorem 2.**  $\text{tr}(\mathbf{AB}) = (\text{vec } \mathbf{A}')'\text{vec } \mathbf{B}$ .

$$\text{Proof. } \text{tr}(\mathbf{AB}) = \sum_i \mathbf{e}_i' \mathbf{A}\mathbf{B}\mathbf{e}_i = [\mathbf{e}_1' \mathbf{A} \quad \cdots \quad \mathbf{e}_r' \mathbf{A}] \begin{bmatrix} \mathbf{Be}_1 \\ \vdots \\ \mathbf{Be}_r \end{bmatrix} = (\text{vec } \mathbf{A}')'\text{vec } \mathbf{B}.$$

QED

**Theorem 3.**  $\text{tr}(\mathbf{AZ/BZC}) = (\text{vec } \mathbf{Z}')(\mathbf{CA} \otimes \mathbf{B}')\text{vec } \mathbf{Z}$ .

**Proof.**

$$\begin{aligned} \text{tr}(\mathbf{AZ/BZC}) &= \text{tr}(\mathbf{Z}'\mathbf{BZCA}) \\ &= (\text{vec } \mathbf{Z}')\text{vec}(\mathbf{BZCA}), \quad \text{from Theorem 2} \\ &= (\text{vec } \mathbf{Z}')(\mathbf{A}'\mathbf{C}' \otimes \mathbf{B}')\text{vec } \mathbf{Z}, \quad \text{from Theorem 1,} \end{aligned}$$

and, being a scalar, this equals its own transpose.

QED

### c. Vec-permutation matrices

The vectors  $\text{vec } A$  and  $\text{vec } A'$  contain the same elements, but in different sequences. We define  $\mathbf{I}_{m,n}$  as that permutation matrix (see Section 6.9) which rearranges  $\text{vec } A'$  to be  $\text{vec } A$ :

$$\text{vec}(A_{m \times n}) = \mathbf{I}_{m,n} \text{vec}(A'). \quad (18)$$

$\mathbf{I}_{m,n}$  is called a *vec-permutation matrix*.

The nature of the relationship between  $\text{vec } A$  and  $\text{vec } A'$  yields a variety of characteristics for the *vec-permutation matrix*, not the least of which is the part played in connecting the Kronecker products (Section 10.7)  $A \otimes B$  and  $B \otimes A$ :

$$B_{p \times q} \otimes A_{m \times n} = \mathbf{I}_{m,p} (A_{m \times n} \otimes B_{p \times q}) \mathbf{I}_{q,n}. \quad (19)$$

This and more, including the many names, notations and definitions given to  $\mathbf{I}_{m,n}$  during the last fifty years, are presented at length in Henderson and Searle (1981b).

### d. Relationships between $\text{vec}$ and $\text{vech}$

When  $X$  is symmetric, the elements of  $\text{vec } X$  are those of  $\text{vech } X$  with some repetitions. Therefore  $\text{vec } X$  and  $\text{vech } X$  for symmetric  $X$  are linear transformations of one another:

$$\text{for } X = X', \quad \text{vech } X = H \text{vec } X \quad \text{and} \quad \text{vec } X = G \text{vech } X, \quad (20)$$

so defining  $H$  and  $G$ . They have the following properties, for  $X$  being  $n \times n$ :

$G$  is unique, of full column rank  $\frac{1}{2}n(n+1)$ , with  $\mathbf{I}_{n,n} G = G$ .

$H$  is not unique, but does have full row rank  $\frac{1}{2}n(n+1)$ .

$H$  is a left inverse of  $G$ , i.e.,  $HG = I$ .

One form of  $H$  is  $H = G^+ = (G'G)^{-1}G'$ , the Moore-Penrose inverse of  $G$ , for which  $H\mathbf{I}_{n,n} = H$  and  $GH = \frac{1}{2}(\mathbf{I}_{n^2} + \mathbf{I}_{n,n})$ .

For any  $H$ ,  $|H(X \otimes X)G| = |X|^{n+1}$ .

These and more are discussed and established by Henderson and Searle (1981a), who also give specific examples of  $H$  and  $G$ .

## 10. OTHER CALCULUS RESULTS

The formulations in Section 8 use matrix notation to represent the differentiating of functions with respect to an array of variables. This

principle can also be applied to a variety of functions of matrices, of which we give but a small selection here. The reader interested in further examples is referred to Deemer and Olkin (1951), Dwyer (1967), Neudecker (1969), Henderson and Searle (1979), and Nel (1980), and to the many references they contain.

### a. Differentiating inverses

Suppose elements of  $A_{r \times c}$  are functions of a scalar  $x$ . The differential of  $A$  with respect to  $x$  is defined as the matrix of each element of  $A$  differentiated with respect to  $x$ :

$$\frac{\partial A}{\partial x} = \left\{ \frac{\partial a_{ij}}{\partial x} \right\} \quad \text{for } i = 1, \dots, r \quad \text{and} \quad j = 1, \dots, c. \quad (21)$$

Applying this to the product  $AA^{-1} = I$  for nonsingular  $A$  gives

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1}, \quad (22)$$

and (21) applied to  $AA^{-1}A = A$ , followed by premultiplication by  $AA^{-1}$  and postmultiplication by  $A^{-1}A$  gives

$$A \frac{\partial A^{-1}}{\partial x} A = -AA^{-1} \frac{\partial A}{\partial x} A^{-1}A, \quad (23)$$

which does, of course, reduce to (22) for nonsingular  $A$ . Also, using  $A'A$  in place of  $A$  in (23), and applying (6) of Section 3.1b, yields

$$A \left[ \frac{\partial (A'A)^{-1}}{\partial x} \right] A' = -A(A'A)^{-1} \frac{\partial (A'A)}{\partial x} (A'A)^{-1} A'. \quad (24)$$

Similarly, if  $A$  is symmetric and if  $T$  has elements that are not functions of  $x$ , then for

$$P = T(T'AT)^{-1}T', \quad \text{we have} \quad \frac{\partial P}{\partial x} = -P \frac{\partial A}{\partial x} P. \quad (25)$$

### b. Differentiating traces

A variation on (13), for differentiating a scalar  $\theta$  that is a function of elements of  $X_{p \times q}$ , is to define

$$\frac{\partial \theta}{\partial X} = \left\{ \frac{\partial \theta}{\partial x_{ij}} \right\} \quad \text{for } i = 1, \dots, p \quad \text{and} \quad j = 1, \dots, q. \quad (26)$$

Trace functions involving  $\mathbf{X}$  are special cases of this. For  $\mathbf{X}$  having functionally independent elements

$$\frac{\partial}{\partial x_{ij}} \operatorname{tr}(\mathbf{X}\mathbf{A}) = \frac{\partial}{\partial x_{ij}} \sum_r \sum_s x_{rs} a_{sr} = a_{ji}$$

and the matrix of all such results is

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{X}\mathbf{A}) = \mathbf{A}'. \quad (27)$$

But (27) does not hold when elements of  $\mathbf{X}$  are functionally related as, for example, when  $\mathbf{X}$  is symmetric. Then, as is shown in Henderson and Searle [1981a, equation (58)],

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{X}\mathbf{A}) = \mathbf{A} + \mathbf{A}' - \operatorname{diag}(\mathbf{A}), \quad \text{for symmetric } \mathbf{X}, \quad (28)$$

where  $\operatorname{diag}(\mathbf{A})$  is the diagonal matrix of diagonal elements of  $\mathbf{A}$ , a variation on the notation for diagonal matrices given in Section 1.6.

### c. Differentiating determinants

For  $\mathbf{X}$  having all elements functionally independent, it is clear that

$$\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{E}_{ij}, \quad (29)$$

a null matrix except for unity as the  $(i, j)$ th element, the matrix  $\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j'$  being as defined in Section 3.1c. Hence

$$\frac{\partial(\mathbf{P}\mathbf{X})}{\partial x_{ij}} = \mathbf{P}\mathbf{E}_{ij} \quad \text{and so} \quad \frac{\partial(\mathbf{P}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{0} \quad \text{implies} \quad \mathbf{P} = \mathbf{0}. \quad (30)$$

Similar results apply for symmetric  $\mathbf{X}$ , for which

$$\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{E}_{ij} + \mathbf{E}_{ji} - \delta_{ij} \mathbf{E}_{ij} \quad (31)$$

for  $\delta_{ij} = 0$  when  $i \neq j$  and  $\delta_{ij} = 1$  when  $i = j$ .

Recalling from Chapter 4 that  $|\mathbf{X}| = \sum_j x_{ij} |\mathbf{X}_{ij}|$  for each  $i$ , where  $\mathbf{X} = \{x_{ij}\}$  and  $|\mathbf{X}_{ij}|$  is the cofactor of  $x_{ij}$ , we have, in contrast to (29),

$$\frac{\partial |\mathbf{X}|}{\partial x_{ij}} = |\mathbf{X}_{ij}| \quad \text{for } \mathbf{X} \text{ having functionally independent elements.} \quad (32)$$

But, similar to (31),

$$\begin{aligned} \frac{\partial |\mathbf{X}|}{\partial x_{ij}} &= |\mathbf{X}_{ij}| + |\mathbf{X}_{ji}| - \delta_{ij} |\mathbf{X}_{ij}| \quad \text{for symmetric } \mathbf{X}, \\ &= (2 - \delta_{ij}) |\mathbf{X}_{ij}| \quad \text{for symmetric } \mathbf{X}. \end{aligned} \quad (33)$$

Therefore, on denoting the adjoint (or adjugate) matrix of  $\mathbf{X}$  by  $\operatorname{adj} \mathbf{X}$  (as at the end of Section 5.4), and on arraying results (32) and (33) as matrices, we have, for  $\mathbf{X}^{-1} = \operatorname{adj}(\mathbf{X})/|\mathbf{X}|$ ,

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = (\operatorname{adj} \mathbf{X})' = |\mathbf{X}| (\mathbf{X}^{-1})', \quad \text{for } \mathbf{X} \text{ having functionally} \\ \text{unrelated elements} \quad (34)$$

and

$$\begin{aligned} \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} &= 2(\operatorname{adj} \mathbf{X}) - \operatorname{diag}(\operatorname{adj} \mathbf{X}) \\ &= |\mathbf{X}| [2\mathbf{X}^{-1} - \operatorname{diag}(\mathbf{X}^{-1})], \quad \text{for symmetric } \mathbf{X}. \end{aligned} \quad (35)$$

Furthermore, for nonsingular  $\mathbf{X}$

$$\frac{\partial}{\partial \mathbf{X}} \log |\mathbf{X}| = \frac{1}{|\mathbf{X}|} \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}}$$

$$= \mathbf{X}^{-1}, \quad \text{from (34), for } \mathbf{X} \text{ with functionally independent elements} \quad (36)$$

$$= 2\mathbf{X}^{-1} - \operatorname{diag}(\mathbf{X}^{-1}), \quad \text{from (35), for symmetric } \mathbf{X}. \quad (37)$$

Results (27) and (36) are well known, but (28) and (37) are not. A recent appearance is Henderson and Searle (1981a).

Another useful derivative is  $\partial \log |\mathbf{X}| / \partial y$ . For  $\mathbf{X}$  having functionally independent elements, it is

$$\frac{\partial}{\partial y} \log |\mathbf{X}| = \frac{1}{|\mathbf{X}|} \frac{\partial |\mathbf{X}|}{\partial y} = \frac{1}{|\mathbf{X}|} \sum_i \sum_j \frac{\partial |\mathbf{X}|}{\partial x_{ij}} \frac{\partial x_{ij}}{\partial y}. \quad (38)$$

Observing that, in general,  $\operatorname{tr}(\mathbf{K}'\mathbf{M}) = \sum_i \sum_j k_{ij} m_{ij}$ , gives

$$\begin{aligned} \frac{\partial}{\partial y} \log |\mathbf{X}| &= \frac{1}{|\mathbf{X}|} \operatorname{tr} \left[ \left( \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} \right)' \frac{\partial \mathbf{X}}{\partial y} \right] = \operatorname{tr} \left[ \left( \frac{1}{|\mathbf{X}|} \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} \right)' \frac{\partial \mathbf{X}}{\partial y} \right] \\ &= \operatorname{tr} \left( \mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial y} \right), \quad \text{from (36)}. \end{aligned} \quad (39)$$

When  $\mathbf{X}$  is symmetric, differentiation is needed only with respect to the distinct elements of  $\mathbf{X}$  so that, in contrast to (38),

$$\frac{\partial}{\partial \mathbf{y}} \log |\mathbf{X}| = \frac{1}{|\mathbf{X}|} \sum_{i < j} \frac{\partial |\mathbf{X}|}{\partial x_{ij}} \frac{\partial x_{ij}}{\partial y} \quad \text{for symmetric } \mathbf{X}.$$

But this, in combination with (33), gives

$$\begin{aligned} \frac{\partial}{\partial \mathbf{y}} \log |\mathbf{X}| &= \frac{1}{|\mathbf{X}|} \sum_{i < j} (2 - \delta_{ij}) |\mathbf{X}_{ij}| \frac{\partial x_{ij}}{\partial y} \quad \text{for } \mathbf{X} = \mathbf{X}' \\ &= \frac{1}{|\mathbf{X}|} \sum_i \sum_j |\mathbf{X}_{ij}| \frac{\partial x_{ij}}{\partial y} = \text{tr} \left( \mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial \mathbf{y}} \right), \end{aligned}$$

which is (39); i.e., (39) applies for symmetric and nonsymmetric  $\mathbf{X}$ .

#### d. Jacobians

When  $\mathbf{y}$  is a vector of  $n$  differentiable functions of the  $n$  elements of  $\mathbf{x}$ , of such a nature that the transformation  $\mathbf{x} \rightarrow \mathbf{y}$  is 1-to-1, then the matrix

$$\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}} = \left( \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right)' = \left\{ \frac{\partial x_i}{\partial y_j} \right\} \quad \text{for } i, j = 1, \dots, n \quad (40)$$

is the *Jacobian matrix* of the transformation from  $\mathbf{x}$  to  $\mathbf{y}$ . For example, for the linear transformation  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , the 1-to-1 property ensures existence of  $\mathbf{A}^{-1}$  and so  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ . Hence, using (11), the Jacobian matrix (40) for

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

is

$$\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}} = \left[ \frac{\partial (\mathbf{A}^{-1}\mathbf{y})}{\partial \mathbf{y}} \right]' = (\mathbf{A}^{-1})' = \mathbf{A}^{-1}. \quad (41)$$

Determinants of Jacobian matrices are required when making a change of variables in integral calculus. Suppose in an integral involving  $x$ 's, a 1-to-1 transformation to  $y$ 's is to be made. The integral is transformed by substituting for the  $x$ 's in terms of the  $y$ 's and by replacing the differentials  $dx_1 dx_2 dx_3 \cdots dx_n$  (denoted by  $d\mathbf{x}$ ) by  $\|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\| dy_1 dy_2 dy_3 \cdots dy_n$  where  $\|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\|$  is the absolute value of the determinant of  $\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}$ . In brief, for the transformation

$$\mathbf{x} \rightarrow \mathbf{y} \text{ make the replacement } d\mathbf{x} \rightarrow \|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\| d\mathbf{y}. \quad (42)$$

Although  $\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}$  is the Jacobian matrix,  $\|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\|$  is known as the *Jacobian* of the transformation.

Because the transformation in (42) is from  $\mathbf{x}$  to  $\mathbf{y}$ , elements of  $\mathbf{y}$  will usually be expressed as functions of those of  $\mathbf{x}$  (rather than vice versa). Hence derivation of  $\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}$  with elements  $\partial y_i / \partial x_j$  will be easier to obtain than  $\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}$  with elements  $\partial x_i / \partial y_j$ . This presents no problem, though, because it can be shown that  $\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}} \mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}} = \mathbf{I}$  and so we always have

$$\|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\| = 1 / \|\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}\|. \quad (43)$$

Hence, whichever of  $\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}$  and  $\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}$  is most easily obtainable from the transformation can be used in (42), either  $\|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\|$  or  $1 / \|\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}\|$  of (43). This being so, relying upon memory for using (42) can raise a doubt: "Is it  $\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}$  or  $\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}$ ?" This is especially important when symbols other than  $\mathbf{x}$  and  $\mathbf{y}$  are being used. Fortunately a useful mnemonic<sup>2</sup> is available for those who like such an aid. Denote the variables by "o" for "old" and "n" for "new", where the transformation is thought of as being from "old" variables to "new" variables. Then in (42) the Jacobian matrix is  $\mathbf{J}_{o \rightarrow n}$  where the subscripts are "on": in contrast, the wrong matrix  $\mathbf{J}_{n \rightarrow o}$  has subscripts "no" — and it is *not* "no"; i.e., it is not  $\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}$ .

**Example.** For the transformation from  $x_1, x_2$  to  $y_1, y_2$  represented by

$$\left. \begin{aligned} y_1 &= e^{x_1} - x_2 \\ y_2 &= x_2 \end{aligned} \right\} \text{ we have } \|\mathbf{J}_{\mathbf{y} \rightarrow \mathbf{x}}\| = \begin{vmatrix} e^{x_1} & -1 \\ 0 & 1 \end{vmatrix} = e^{x_1} = y_1 + y_2.$$

The inverse relationship is, for  $r = 1/(y_1 + y_2)$ ,

$$\left. \begin{aligned} x_1 &= \log(y_1 + y_2) \\ x_2 &= y_2 \end{aligned} \right\} \text{ with } \|\mathbf{J}_{\mathbf{x} \rightarrow \mathbf{y}}\| = \begin{vmatrix} r & r \\ 0 & 1 \end{vmatrix} = 1/(y_1 + y_2).$$

Sometimes variables are arrayed not as vectors but as matrices, in which case, putting the matrices in vector form (by the operations of Section 9) enables (40) and (42) to be used for deriving Jacobians. In general

$$\mathbf{J}_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{\partial \text{vec } \mathbf{X}}{\partial \text{vec } \mathbf{Y}}$$

and

$$\mathbf{J}_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{\partial \text{vech } \mathbf{X}}{\partial \text{vech } \mathbf{Y}} \quad \text{for } \mathbf{X} \text{ and } \mathbf{Y} \text{ symmetric.}$$

<sup>2</sup>I am indebted to Daniel L. Solomon for this.

TABLE 12.1. JACOBIANS FOR CERTAIN MATRIX TRANSFORMATIONS

Transformation: $X \rightarrow Y$	Jacobian: $\ J_{X \rightarrow Y}\ $	
	Matrices Having Functionally Independent Elements	Symmetric Matrices $X = X'$ and $Y = Y'$
$Y_{p \times q} = AX_{p \times q}B$	$ A ^{-q}  B ^{-p}$	Not applicable
$Y_{p \times p} = AX_{p \times p}A'$	$ A ^{-2p}$	$ A ^{-(p+1)}$
$Y_{n \times n} = X^{-1}$	$ X ^{2n}$	$ X ^{n+1}$
$Y_{n \times n} = X^k$	$\left( \prod_{i=1}^n \prod_{j=1}^n \sum_{r=1}^k \lambda_i^{k-r} \lambda_j^{r-1} \right)^{-1}$	$\left( \prod_{i=1}^n \prod_{j \geq i}^n \sum_{r=1}^k \lambda_i^{k-r} \lambda_j^{r-1} \right)^{-1}$

Note: All determinants are taken positively, and in the last line  $\lambda_1, \lambda_2, \dots, \lambda_n$  are eigenvalues of  $X$ ; symmetric  $X$  differs from non-symmetric only in having  $j \geq i$ .

These techniques provide straightforward derivation of results such as those shown in Table 12.1. Henderson and Searle (1981a) give details and discuss earlier derivations.

**e. Aitken's integral**

For positive definite  $A$ , the result

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}x'Ax} dx_1 \dots dx_n = (2\pi)^{\frac{1}{2}n} |A|^{-\frac{1}{2}} \quad (44)$$

is known as Aitken's integral. It is established by noting that because  $A$  is p.d., there exists a nonsingular matrix  $P$  such that  $P'AP = I$ . Letting  $y = P^{-1}x$ , the Jacobian for which is, by (41),  $\|P\| = |A|^{-\frac{1}{2}}$ , gives

$$\begin{aligned} & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}x'Ax} dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y'P'APy} dy_1 \dots dy_n |A|^{-\frac{1}{2}} \\ &= \prod_{i=1}^n \left[ \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i \right] |A|^{-\frac{1}{2}} = (2\pi)^{\frac{1}{2}n} |A|^{-\frac{1}{2}}. \end{aligned}$$

**f. Hessians**

Another matrix of partial differentials is the Hessian. If  $\theta$  is a function of  $n$  variables  $x_1, x_2, \dots, x_n$ , the matrix of second-order partial derivatives of  $\theta$  with respect to the  $x$ 's is called a Hessian: for  $i, j = 1, 2, \dots, n$ ,

$$H = \left\{ \frac{\partial^2 \theta}{\partial x_i \partial x_j} \right\} = \frac{\partial^2 \theta}{\partial x \partial x'}$$

**Illustration.** One of the properties of maximum likelihood in statistics is that, under certain conditions (see, for example, Mood, Graybill and Boes, 1974), the variance-covariance matrix of a set of simultaneously estimated parameters is the inverse of minus the expected value of the Hessian of the likelihood with respect to those parameters.

**11. MATRICES WITH ELEMENTS THAT ARE COMPLEX NUMBERS**

Knowledge of complex numbers is not a prerequisite for reading this book, which deals primarily with real matrices, i.e., having elements that are real numbers. However, as with other topics that have been touched on lightly in order to introduce the reader to concepts and vocabulary that may be encountered elsewhere, so here—we briefly give some definitions of matrices having elements that are complex numbers.

The starting point is the square root of a negative number. For  $\sqrt{-1}$  the symbol  $i$  is used so that, for example,  $\sqrt{-6} = i\sqrt{6}$  and  $\sqrt{-81} = 9i$ . Numbers that are multiples of  $i$  are called imaginary numbers; those not involving imaginary numbers are real. The sum of a real and an imaginary number is a complex number; e.g.,  $3 + 7i$  and  $-2 + i\sqrt{11}$  are complex numbers. In general, for  $a$  and  $b$  being real,  $a + ib$  is complex. The number  $a - ib$  is called the complex conjugate of  $a + ib$  and the pair of numbers  $a + ib$  and  $a - ib$  are called a conjugate pair.

An example of a matrix having complex numbers for elements is

$$M = \begin{bmatrix} 3 + 2i & 4 - 5i \\ -6i & 9 + i\sqrt{2} \\ 7 - i & 6 + 7i \end{bmatrix}; \quad \text{and} \quad \bar{M} = \begin{bmatrix} 3 - 2i & 4 + 5i \\ 6i & 9 - i\sqrt{2} \\ 7 + i & 6 - 7i \end{bmatrix}$$

is the complex conjugate of  $M$ . It is the matrix whose elements are the complex conjugates of the elements of  $M$ . If, for  $A$  and  $B$  being real,  $M = A + iB$ , then  $\bar{M} = A - iB$ , analogous to scalar complex numbers. A Hermitian matrix is one in which the transpose of the complex conjugate

equals the matrix itself:  $\bar{M}' = M$ , in which case the symbol  $M^*$  is often used. For example, if

$$M^* = \begin{bmatrix} 3 & 7+8i \\ 7-8i & 4 \end{bmatrix},$$

$$\bar{M}^{*'} = \begin{bmatrix} 3 & 7-8i \\ 7+8i & 4 \end{bmatrix}' = \begin{bmatrix} 3 & 7+8i \\ 7-8i & 4 \end{bmatrix} = M^*.$$

The diagonal elements of a Hermitian matrix are real. When  $\bar{M}'M = I$ , the matrix  $M$  is said to be a *unitary* matrix.

Clearly, when  $B$  is null, matrices of the form  $A + iB$  become real, if Hermitian they become symmetric, and if unitary they become orthogonal; i.e., being Hermitian is the counterpart of being symmetric, as is unitary of orthogonal. Indeed, symmetric matrices are a subset of Hermitian matrices, as are orthogonal matrices of unitary ones.

## EXERCISES

1. Using Section 1f, calculate orthogonal matrices from

$$\begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ -1 & -2 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 5 \\ 2 & -5 & 0 \end{bmatrix},$$

and use them to verify Sections 1a through 1e.

2. Calculate  $P$  of Section 2k, using

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

Using  $P$ , verify results in Section 2.

3. Using an orthogonal matrix calculated in Exercise 1 as  $Q$ , and  $P$  from Exercise 2, show that  $Q'PQ$  is idempotent.
4. Verify results of Section 2 for  $P$  of Section 2k.
5. Define  $Q = X(YX)^{-1}Y$ . Prove the following:
- $Q$  is idempotent.
  - $Q = I$  when  $X$  and  $Y$  are nonsingular.
  - $\text{tr}(Q) = \text{number of rows in } Y$ .

6. For  $D = X(X'V^{-1}X)^{-1}X'V^{-1}$  with  $X$  rectangular and  $V$  symmetric, prove the following:
- $D$  is idempotent.
  - $D' = V^{-1}DV$ .
  - $(I - D')V^{-1}(I - D) = (I - D')V^{-1} = V^{-1}(I - D)$ .

7. For  $A$  being idempotent of order  $n$ , prove  $r_{I-A} = n - r_A$  in two ways: (a) using the trace-equal-to-rank property of  $A$ , and (b) considering LIN solutions of  $(I - A)x = 0$ .

8. Prove that a symmetric matrix is idempotent if and only if its eigenvalues are all zero or unity.

9. Suppose  $A$  and  $V$  are symmetric, and  $V$  is p.d. If  $AV$  has eigenvalues that are all zero or unity, prove that  $AV$  is idempotent.

10. Prove that

$$\begin{vmatrix} a & c\mathbf{I}' \\ c\mathbf{I} & (a-b)\mathbf{I}_r + b\mathbf{J}_r \end{vmatrix} = (a-b)^{r-1} [a^2 + (r-1)ab - rc^2].$$

11. Prove that

$$(a\mathbf{I}_r + b\mathbf{J}_r)^n = a^n\mathbf{I}_r + \frac{1}{r} [(a+rb)^n - a^n]\mathbf{J}_r,$$

and

$$e^{a\mathbf{I}_r + b\mathbf{J}_r} = e^a\mathbf{I}_r + \frac{1}{r}(e^{a+rb} - e^a)\mathbf{J}_r.$$

12. Show that  $a\mathbf{I} + b\mathbf{J}$  of order  $n$  is diagonalizable, and verify that  $\mathbf{H}'_n$  of Section 3.4b is a matrix of its eigenvectors.

13. Under what condition is  $a\mathbf{I} + b\mathbf{J}$  of order  $n$  (a) idempotent and (b) orthogonal?

14. When  $V$  is p.d., prove that  $(X'V^{-1}X)^{-1}X'V^{-1}$  is a generalized inverse of  $X$ .

15. Prove that  $A(A'A)^{-2}A'$  is the Moore-Penrose inverse of the n.n.d. matrix  $AA'$ .

16. If  $K$  is square of order  $n$ , what values of  $K$  make  $e^K$  equal to the following?

- |                      |                                 |
|----------------------|---------------------------------|
| (a) $e\mathbf{I}$    | (b) $\mathbf{I} - K + eK$       |
| (c) $\mathbf{I}$     | (d) $e^\lambda\mathbf{I}$       |
| (e) $\mathbf{I} + K$ | (f) $\mathbf{I} + K(e^n - 1)/n$ |



17. (a) Show that  $(-a - b)^{n-1} \mathbf{R}$  is the  $n$ th power of

$$\mathbf{R} = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}.$$

- (b) Show that  $e^{\mathbf{R}} = \begin{bmatrix} p_1 & 1-p_1 \\ p_2 & 1-p_2 \end{bmatrix}$  implies

$$a = \frac{-(1-p_1)\log_e(p_1-p_2)}{1-p_1+p_2} \quad \text{and} \quad a+b = -\log_e(p_1-p_2).$$

These are the matrices used by Pielou (1964).

18. In Section 8, derive (17) from maximizing  $r_{IH}$ .
19. With numerical examples of your own creation (without using matrices  $\mathbf{0}$ ,  $\mathbf{I}$  or  $\mathbf{J}$ ), illustrate the three theorems of Section 9.

20. For  $\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$ , derive  $\mathbf{G}$  and  $\mathbf{H}$  of (20) and illustrate their properties stated in Section 9d.

21. Prove:  $\text{vec}(\mathbf{PQ}) = (\mathbf{Q}' \otimes \mathbf{I})\text{vec } \mathbf{P} = (\mathbf{Q}' \otimes \mathbf{P})\text{vec } \mathbf{I} = (\mathbf{I} \otimes \mathbf{P})\text{vec } \mathbf{Q}$ .

22. Prove and illustrate:  $\text{vec}(\mathbf{xy}') = \mathbf{y} \otimes \mathbf{x}$ . Use this result, along with  $\mathbf{B} = \sum_i \sum_j b_{ij} \mathbf{e}_i \mathbf{e}_j'$ , to provide an alternative proof of Theorem 1 in Section 9b.

23. Prove and/or illustrate:

(a)  $\text{vec}(\mathbf{A} \otimes \mathbf{b}) = (\text{vec } \mathbf{A}) \otimes \mathbf{b}$ .

(b)  $\text{vec}(\mathbf{a}_{p \times 1} \otimes \mathbf{B}_{m \times n}) = (\mathbf{I}_{p,n} \otimes \mathbf{I}_m)(\mathbf{a} \otimes \text{vec } \mathbf{B})$ .

(c)  $\text{vec}(\mathbf{A}_{p \times q} \otimes \mathbf{B}_{m \times n}) = (\mathbf{I}_q \otimes \mathbf{I}_{p,n} \otimes \mathbf{I})(\text{vec } \mathbf{A} \otimes \text{vec } \mathbf{B})$ .

24. Show that with respect to  $x_1$  and  $x_2$ , the Jacobian of

$$y_1 = 6x_1^2x_2 + 2x_1x_2 + x_2^2 \quad \text{and} \quad y_2 = 2x_1^3 + x_1^2 + 2x_1x_2$$

is the same as the Hessian of

$$y = 2x_1^3x_2 + x_1^2x_2 + x_1x_2^2.$$

25. Prove:  $\partial(\mathbf{PX} + \mathbf{X}'\mathbf{Q})/\partial \mathbf{X} = \mathbf{0}$  implies  $\mathbf{P} = -\mathbf{Q}'$ .

26. Express the quadratic form

$$7x_1^2 + 4x_1x_2 - 5x_2^2 - 6x_2x_3 + 3x_3^2 + 6x_1x_3$$

in the form  $\mathbf{x}'\mathbf{A}\mathbf{x}$  where  $\mathbf{A} = \mathbf{A}'$ , and verify (13).

27. For  $\mathbf{A}$  and  $\mathbf{B}$  being p.d., prove that solutions to  $|\mathbf{B} - \lambda\mathbf{A}| = 0$  for  $\lambda$  are positive.
28. Find the value of  $\beta$  that maximizes  $\beta'\mathbf{X}\text{tr}'\mathbf{X}'\beta/\beta'\mathbf{X}'\mathbf{X}\beta$ ; and find that maximum value.
29. For  $\mathbf{A}$  being n.n.d., prove that  $(\mathbf{x}'\mathbf{A}\mathbf{y})^2 \leq \mathbf{x}'\mathbf{A}\mathbf{x}(\mathbf{y}'\mathbf{A}\mathbf{y})$ . Applications of this to statistics are given by Wolkowicz and Styan (1979). [Hint: use the Cauchy-Schwarz inequality of Exercise 2.23.]
30. Verify:  $\text{tr}(\mathbf{AX}) = (\text{vec } \mathbf{A})'\text{vec } \mathbf{X} = (\text{vec } \mathbf{X}')\text{vec } \mathbf{A}$ . By differentiating with respect to  $\text{vec } \mathbf{X}$ , establish (27).

APPLICATIONS IN STATISTICS

Mathematical description and development of many statistical methods are greatly aided by matrix algebra. A few examples from a multitude of possibilities are presented in this chapter.

**Notation.** Thus far we have maintained the conventional mathematics notation of using capital letters for matrices and lowercase for vectors. This will be continued, despite its clash with the statistics convention of using capital letters for random variables and lowercase for realized values thereof. For example, the use of  $\mathbf{X}$  for a matrix and  $\mathbf{x}$  for a vector will be maintained, whereas the notation  $\mathbf{X} = [X_1 \ X_2 \ \dots \ X_k]'$  for a vector of random variables  $X_1, X_2, \dots, X_k$  will not be used. The matrix use of  $\mathbf{X}$  is retained by making  $\mathbf{x} = \{x_i\}$  do double duty: on some occasions  $\mathbf{x}$  will represent random variables and on other occasions it will represent realized values. Maintaining this distinction is easier than having  $\mathbf{X}$  sometimes be a vector when it is usually a matrix.

The operation of taking expected values will be represented by  $E$ . Thus  $E(\mathbf{x})$  is the vector of the expected value of each element of  $\mathbf{x}$ :

$$E(\mathbf{x}) = E \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_k) \end{bmatrix}$$

Similarly,  $E(\mathbf{X})$  is the matrix of the expected value of each element of  $\mathbf{X}$ . For convenience,  $E(\mathbf{x})$  and  $E \mathbf{x}$  will mean the same thing, the latter being useful when taking the expectation of a function, e.g.,  $E(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'$ . This avoids excessive parentheses, and the possibility of  $E(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'$  being read as  $E(\mathbf{x} - \boldsymbol{\mu})$  postmultiplied by  $(\mathbf{x} - \boldsymbol{\mu})'$ .

1. VARIANCE-COVARIANCE MATRICES

Suppose  $x_1, x_2, \dots, x_k$  are  $k$  random variables with means  $\mu_1, \mu_2, \dots, \mu_k$ , variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$  and covariances  $\sigma_{12}, \sigma_{13}, \dots, \sigma_{k-1, k}$ . On representing the random variables and their means by the vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \quad \text{we have} \quad E(\mathbf{x}) = \boldsymbol{\mu}; \quad (1)$$

and assembling the variances and covariances in a matrix gives

$$\text{var}(\mathbf{x}) = \mathbf{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \dots & \sigma_k^2 \end{bmatrix}. \quad (2)$$

This is the *variance-covariance matrix*, or *dispersion matrix*, of the random variables  $x_1, x_2, \dots, x_k$ . It is symmetric,  $\mathbf{V}' = \mathbf{V}$ ; its  $i$ th diagonal element is the variance of  $x_i$  and its  $ij$ th off-diagonal element (for  $i \neq j$ ) is the covariance between  $x_i$  and  $x_j$ .

The familiar definitions of variance and covariance are

$$\sigma_i^2 = E(x_i - \mu_i)^2 \quad \text{and} \quad E(x_i - \mu_i)(x_j - \mu_j) = \sigma_{ij} \quad \text{for} \quad i \neq j,$$

respectively. They provide the means for expressing  $\mathbf{V}$  as

$$\mathbf{V} = \text{var}(\mathbf{x}) = E(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'. \quad (3)$$

When the means are zero,  $\boldsymbol{\mu} = \mathbf{0}$ , and (3) becomes

$$\mathbf{V} = \text{var}(\mathbf{x}) = E(\mathbf{x}\mathbf{x}').$$

The dispersion matrix of a linear transformation  $\mathbf{y} = \mathbf{T}\mathbf{x}$  of the  $x_i$ 's is easily obtained: first we have the mean,

$$E(\mathbf{y}) = E(\mathbf{T}\mathbf{x}) = \mathbf{T}E(\mathbf{x}) = \mathbf{T}\boldsymbol{\mu}. \quad (4)$$

Then the dispersion matrix of  $\mathbf{y}$  is

$$\text{var}(\mathbf{y}) = E[\mathbf{y} - E(\mathbf{y})][\mathbf{y} - E(\mathbf{y})]' = E\mathbf{T}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{T}' = \mathbf{T}\mathbf{V}\mathbf{T}'. \quad (5)$$

When  $\mathbf{T}$  is just a single row, i.e., a row vector  $\mathbf{t}'$ , (5) is  $\text{var}(\mathbf{t}'\mathbf{y}) = \mathbf{t}'\mathbf{V}\mathbf{t}$ . Because  $\text{var}(\mathbf{t}'\mathbf{y})$  is a variance, it is never negative and so therefore neither is  $\mathbf{t}'\mathbf{V}\mathbf{t}$ . Thus  $\mathbf{t}'\mathbf{V}\mathbf{t}$  is a n.n.d. quadratic form. Hence  $\mathbf{V}$  is a non-negative definite matrix (see Section 3.6). This is a characteristic of all variance-covariance matrices; they are all non-negative definite. In many cases they are specifically positive definite.

## 2. CORRELATION MATRICES

Suppose  $\rho_{ij}$  is the correlation between the random variables  $x_i$  and  $x_j$ . Then, provided every  $\sigma_i^2$  is nonzero,

$$\rho_{ij} = \rho_{ji} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (6)$$

and arraying the  $\rho_{ij}$ 's in a (symmetric) matrix  $\mathbf{R}$ , with diagonal elements  $\rho_{ii} = \sigma_{ii}/\sigma_i \sigma_i = \sigma_i^2/\sigma_i^2 = 1$ , gives the correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{12} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1k} & \rho_{2k} & \cdots & 1 \end{bmatrix}. \quad (7)$$

Define

$$\mathbf{D} = \text{diag}\{\sigma_i^2\} \quad \text{for } i = 1, \dots, k,$$

a diagonal matrix of the variances in the diagonal of  $\mathbf{V}$  and observe that  $\sqrt{\mathbf{D}} = \mathbf{D}^{\frac{1}{2}} = \text{diag}\{\sigma_i\}$ . Then

$$\mathbf{R} = \mathbf{D}^{-\frac{1}{2}} \mathbf{V} \mathbf{D}^{-\frac{1}{2}}. \quad (8)$$

**Illustration.** Suppose  $\mathbf{x}$  represents  $k$  random variables each having variance  $\sigma^2$ , with the covariance between every pair of  $\mathbf{x}$ 's being  $\rho\sigma^2$ . Then the variance-covariance matrix of  $\mathbf{x}$  is, from (2),

$$\mathbf{V} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{R} \quad \text{for } \mathbf{R} = (1 - \rho)\mathbf{I}_k + \rho\mathbf{J}_k.$$

When the  $\mathbf{x}$ 's follow a normal distribution, the determinant and inverse of the correlation matrix  $\mathbf{R}$  are needed for the density function. They are

$$|\mathbf{R}| = (1 - \rho)^{k-1} (1 - \rho + k\rho)$$

and

$$\mathbf{R}^{-1} = \frac{1}{1 - \rho} \left( \mathbf{I}_k - \frac{\rho}{1 - \rho + k\rho} \mathbf{J}_k \right).$$

Since  $\mathbf{V}$  is p.s.d.,  $|\mathbf{V}|$  and  $|\mathbf{R}|$  are non-negative, and therefore feasible values for  $\rho$  are restricted to  $1 \geq \rho \geq -1/(k-1)$ . If either equality holds,  $\mathbf{R}^{-1}$  does not exist.

## 3. MATRICES OF SUMS OF SQUARES AND CROSS-PRODUCTS

### a. Data matrices

We use the symbol  $x$  in this section to represent an observation. Suppose we have  $N$  samples of data on each of  $k$  variables, with  $x_{ij}$  being the  $i$ th observation on variable  $j$ . Then it is natural to array all  $Nk$  observations in what is called a *data matrix*:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} \quad (9)$$

$$= \{x_{ij}\} \quad \text{for } i = 1, \dots, N \text{ and } j = 1, \dots, k.$$

In this definition of  $\mathbf{X}$  of order  $N \times k$ , each column represents the  $N$  observations on a particular variable and each row represents a single observation on all  $k$  variables.

Some authors define  $\mathbf{X}$  to have order  $k \times N$ , in the form of our  $\mathbf{X}'$ , in which case each column represents one observation on all  $k$  variables and a row represents the  $N$  observations on a particular variable. One therefore needs to take care, when working with data matrices, to be sure of which representation of the data  $\mathbf{X}$  is being used for.

### b. Uncorrected sums of squares and products

The uncorrected sum of squares of the observations on variable  $j$  in  $\mathbf{X}$  is

$$\sum_{i=1}^N x_{ij}^2 = \mathbf{x}'_j \mathbf{x}_j$$

where  $\mathbf{x}_j$  is the  $j$ th column of  $\mathbf{X}$ ; similarly, the corresponding sum of products of the observations on the  $j$ th and  $t$ th variables is

$$\sum_{i=1}^N x_{ij}x_{it} = \mathbf{x}'_j\mathbf{x}_t = \mathbf{x}'_t\mathbf{x}_j.$$

Assembling all of these sums of squares and cross-products into a matrix, which is symmetric, we have

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \sum_{i=1}^N x_{i1}^2 & \sum_{i=1}^N x_{i1}x_{i2} & \cdots & \sum_{i=1}^N x_{i1}x_{ik} \\ \sum_{i=1}^N x_{i1}x_{i2} & \sum_{i=1}^N x_{i2}^2 & \cdots & \sum_{i=1}^N x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N x_{i1}x_{ik} & \sum_{i=1}^N x_{i2}x_{ik} & \cdots & \sum_{i=1}^N x_{ik}^2 \end{bmatrix} \quad (10)$$

as the matrix of uncorrected sums of squares and cross-products of the  $x$ 's. It is sometimes called the *uncorrected sum of squares* (and cross-products) *matrix*.

### c. Means, and the centering matrix

The mean of the  $N$  observations on the  $j$ th variable, i.e., of the elements in the  $j$ th column of  $\mathbf{X}$ , is

$$\bar{x}_{.j} = \sum_{i=1}^N x_{ij} / N = (x_{1j} + \cdots + x_{Nj}) / N = \mathbf{1}'\mathbf{x}_j / N = \mathbf{x}'_j\mathbf{1} / N, \quad (11)$$

where  $\mathbf{1}$  is the summing vector (all elements unity, see Section 3.2) of order  $N$ . This is the observed mean for the  $j$ th variable. And the row vector of all  $k$  such observed means is

$$\bar{\mathbf{x}}' = [\bar{x}_{.1} \quad \bar{x}_{.2} \quad \cdots \quad \bar{x}_{.k}] = \frac{1}{N} \mathbf{1}'\mathbf{X}. \quad (12)$$

Now consider each element  $x_{ij}$  of  $\mathbf{X}$  replaced by its deviation from its column mean: for  $i = 1, \dots, N$  and  $j = 1, \dots, k$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} - \begin{bmatrix} \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.k} \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.k} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.k} \end{bmatrix} \quad (13)$$

$$= \{x_{ij} - \bar{x}_{.j}\} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'. \quad (14)$$

Inspection of (13) reveals that its  $j$ th column is

$$\begin{bmatrix} x_{1j} - \bar{x}_{.j} \\ x_{2j} - \bar{x}_{.j} \\ \vdots \\ x_{Nj} - \bar{x}_{.j} \end{bmatrix} = \mathbf{x}_j - \frac{1}{N} \mathbf{J}_N \mathbf{x}_j = (\mathbf{I} - \bar{\mathbf{J}}_N) \mathbf{x}_j = \mathbf{C}_N \mathbf{x}_j \quad (15)$$

where  $\mathbf{C}_N$  is the centering matrix of Section 3.2:

$$\mathbf{C}_N = \mathbf{I} - \bar{\mathbf{J}}_N = \mathbf{I} - \frac{1}{N} \mathbf{J}_N = \mathbf{I} - \frac{1}{N} \mathbf{1}_N \mathbf{1}'_N. \quad (16)$$

Hence (13) is

$$\{x_{ij} - \bar{x}_{.j}\} = \mathbf{C}_N \mathbf{X}. \quad (17)$$

### d. Corrected sums of squares and products

From (15), using the symmetry and idempotency of  $\mathbf{C}_N$ , the sum of squares of observations in the  $j$ th column of  $\mathbf{X}$ , corrected for the mean, is

$$\sum_{i=1}^N (x_{ij} - \bar{x}_{.j})^2 = \mathbf{x}'_j \mathbf{C}_N \mathbf{x}_j \quad (18)$$

just like (19) of Section 3.2. Similarly, the corrected sum of cross-products of observations on the  $j$ th and  $t$ th variables is

$$\sum_{i=1}^N (x_{ij} - \bar{x}_{.j})(x_{it} - \bar{x}_{.t}) = \mathbf{x}'_j \mathbf{C}_N \mathbf{x}_t. \quad (19)$$

Arraying in a matrix all values of (18) and (19), for  $j, t = 1, \dots, k$ , gives the matrix of corrected sums of squares and cross-products:

$$\left\{ \sum_{i=1}^N (x_{ij} - \bar{x}_{.j})(x_{it} - \bar{x}_{.t}) \right\} = \mathbf{X}' \mathbf{C}_N \mathbf{X}. \quad (20)$$

This is, of course, exactly analogous to (10) with  $\mathbf{C}_N \mathbf{X}$  of (17) in place of  $\mathbf{X}$ , and using the symmetry and idempotency of  $\mathbf{C}_N$ :

$$(\mathbf{C}_N \mathbf{X})' \mathbf{C}_N \mathbf{X} = \mathbf{X}' \mathbf{C}_N^2 \mathbf{X} = \mathbf{X}' \mathbf{C}_N \mathbf{X}.$$

Hence denoting the matrix of corrected sums of squares and products in

where  $\mathbf{x}_j$  is the  $j$ th column of  $\mathbf{X}$ ; similarly, the corresponding sum of products of the observations on the  $j$ th and  $t$ th variables is

$$\sum_{i=1}^N x_{ij}x_{it} = \mathbf{x}'_j \mathbf{x}_t = \mathbf{x}'_t \mathbf{x}_j.$$

Assembling all of these sums of squares and cross-products into a matrix, which is symmetric, we have

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \sum_{i=1}^N x_{i1}^2 & \sum_{i=1}^N x_{i1}x_{i2} & \cdots & \sum_{i=1}^N x_{i1}x_{ik} \\ \sum_{i=1}^N x_{i1}x_{i2} & \sum_{i=1}^N x_{i2}^2 & \cdots & \sum_{i=1}^N x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N x_{i1}x_{ik} & \sum_{i=1}^N x_{i2}x_{ik} & \cdots & \sum_{i=1}^N x_{ik}^2 \end{bmatrix} \quad (10)$$

as the matrix of uncorrected sums of squares and cross-products of the  $x$ 's. It is sometimes called the *uncorrected sum of squares* (and cross-products) *matrix*.

#### c. Means, and the centering matrix

The mean of the  $N$  observations on the  $j$ th variable, i.e., of the elements in the  $j$ th column of  $\mathbf{X}$ , is

$$\bar{x}_{.j} = \sum_{i=1}^N x_{ij} / N = (x_{1j} + \cdots + x_{Nj}) / N = \mathbf{1}'\mathbf{x}_j / N = \mathbf{x}'_j \mathbf{1} / N, \quad (11)$$

where  $\mathbf{1}$  is the summing vector (all elements unity, see Section 3.2) of order  $N$ . This is the observed mean for the  $j$ th variable. And the row vector of all  $k$  such observed means is

$$\bar{\mathbf{x}}' = [\bar{x}_{.1} \quad \bar{x}_{.2} \quad \cdots \quad \bar{x}_{.k}] = \frac{1}{N} \mathbf{1}'\mathbf{X}. \quad (12)$$

Now consider each element  $x_{ij}$  of  $\mathbf{X}$  replaced by its deviation from its column mean: for  $i = 1, \dots, N$  and  $j = 1, \dots, k$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} - \begin{bmatrix} \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.k} \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.k} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.k} \end{bmatrix} \quad (13)$$

$$= \{x_{ij} - \bar{x}_{.j}\} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'. \quad (14)$$

Inspection of (13) reveals that its  $j$ th column is

$$\begin{bmatrix} x_{1j} - \bar{x}_{.j} \\ x_{2j} - \bar{x}_{.j} \\ \vdots \\ x_{Nj} - \bar{x}_{.j} \end{bmatrix} = \mathbf{x}_j - \frac{1}{N} \mathbf{J}_N \mathbf{x}_j = (\mathbf{I} - \bar{\mathbf{J}}_N) \mathbf{x}_j = \mathbf{C}_N \mathbf{x}_j, \quad (15)$$

where  $\mathbf{C}_N$  is the centering matrix of Section 3.2:

$$\mathbf{C}_N = \mathbf{I} - \bar{\mathbf{J}}_N = \mathbf{I} - \frac{1}{N} \mathbf{J}_N = \mathbf{I} - \frac{1}{N} \mathbf{1}_N \mathbf{1}'_N. \quad (16)$$

Hence (13) is

$$\{x_{ij} - \bar{x}_{.j}\} = \mathbf{C}_N \mathbf{X}. \quad (17)$$

#### d. Corrected sums of squares and products

From (15), using the symmetry and idempotency of  $\mathbf{C}_N$ , the sum of squares of observations in the  $j$ th column of  $\mathbf{X}$ , corrected for the mean, is

$$\sum_{i=1}^N (x_{ij} - \bar{x}_{.j})^2 = \mathbf{x}'_j \mathbf{C}_N \mathbf{x}_j, \quad (18)$$

just like (19) of Section 3.2. Similarly, the corrected sum of cross-products of observations on the  $j$ th and  $t$ th variables is

$$\sum_{i=1}^N (x_{ij} - \bar{x}_{.j})(x_{it} - \bar{x}_{.t}) = \mathbf{x}'_j \mathbf{C}_N \mathbf{x}_t, \quad (19)$$

Arrayed in a matrix all values of (18) and (19), for  $j, t = 1, \dots, k$ , gives the matrix of corrected sums of squares and cross-products:

$$\left\{ \sum_{i=1}^N (x_{ij} - \bar{x}_{.j})(x_{it} - \bar{x}_{.t}) \right\} = \mathbf{X}' \mathbf{C}_N \mathbf{X}. \quad (20)$$

This is, of course, exactly analogous to (10) with  $\mathbf{C}_N \mathbf{X}$  of (17) in place of  $\mathbf{X}$ , and using the symmetry and idempotency of  $\mathbf{C}_N$ :

$$(\mathbf{C}_N \mathbf{X})' \mathbf{C}_N \mathbf{X} = \mathbf{X}' \mathbf{C}_N^2 \mathbf{X} = \mathbf{X}' \mathbf{C}_N \mathbf{X}.$$

Hence denoting the matrix of corrected sums of squares and products in

(20) by  $\mathbf{S}$ , we have

$$\mathbf{S} = \mathbf{X}'\mathbf{C}_N\mathbf{X}. \quad (21)$$

The matrix  $\mathbf{S}$  of (21) arises in multivariate statistics, where it is called the *Wishart matrix* (see Section 4); it also occurs in multiple regression (see Section 14.3). An elementary example of the preceding expressions is the following illustration of (18).

**Illustration.** When  $\mathbf{X}$  is a vector,  $\mathbf{x} = [x_1 \ \dots \ x_N]'$ , equation (18) is

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \mathbf{x}'\mathbf{C}_N\mathbf{x}, \quad (22)$$

as in (19) of Section 3.2. Because  $\mathbf{C}_N$  is symmetric, we know from (52) of Section 11.6c that its canonical form  $\mathbf{C}_N = \mathbf{UDU}'$  under orthogonal similarity exists. Hence

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \mathbf{x}'\mathbf{UDU}'\mathbf{x} = \mathbf{y}'\mathbf{D}\mathbf{y}, \quad (23)$$

for  $\mathbf{y} = \mathbf{U}'\mathbf{x}$  and where  $\mathbf{D}$  is the diagonal matrix of eigenvalues of  $\mathbf{C}_N$ . But, from Section 12.3f,  $\mathbf{C}_N$  is idempotent, with  $N - 1$  eigenvalues of 1 and one of zero. Hence

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_{N-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}$$

and so from (23)

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^{N-1} y_i^2 \quad \text{for } \mathbf{y} = \mathbf{U}'\mathbf{x} \text{ with } \mathbf{U} \text{ orthogonal.} \quad (24)$$

The importance of (24) in relation to sums of squares having  $\chi^2$ -distributions is seen in the next section.

#### 4. THE MULTIVARIATE NORMAL DISTRIBUTION

The probability density and moment generating functions for a normally distributed random variable  $x$  having mean  $\mu$  and variance  $\sigma^2$  are, respectively,

$$f(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \quad \text{and} \quad m_x(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}. \quad (25)$$

We say that  $\mathbf{x}$  is normally distributed and write this as  $\mathbf{x}$  is  $N(\boldsymbol{\mu}, \sigma^2)$  or  $\mathbf{x} \sim N(\boldsymbol{\mu}, \sigma^2)$ .

The multivariate analogue of (25) is for a vector  $\mathbf{x}$ , of  $k$  random variables, with  $E(\mathbf{x}) = \boldsymbol{\mu}$  and  $\text{var}(\mathbf{x}) = \mathbf{V}$  as in (1) and (3). When all  $k$  variables are normally distributed, the joint density and moment generating functions, respectively, are

$$f(\mathbf{x}) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}}{\sqrt{|2\pi\mathbf{V}|}} \quad \text{and} \quad m_{\mathbf{x}}(\mathbf{t}) = e^{t'\boldsymbol{\mu} + \frac{1}{2}t'\mathbf{V}\mathbf{t}}. \quad (26)$$

The analogy between (25) and (26) is clear; and, of course, when  $k = 1$ , the functions in (26) reduce to those in (25). Notice in (26) that  $f(\mathbf{x})$  exists only if  $\mathbf{V}^{-1}$  does, that the denominator of  $f(\mathbf{x})$  has  $|\mathbf{V}|$  as the analogue of  $\sigma^2$  in (25) and that  $\sqrt{|2\pi\mathbf{V}|}$  is usually written as  $(2\pi)^{k/2}|\mathbf{V}|^{1/2}$ . A summary statement for (26) is  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$ , just as  $x \sim N(\mu, \sigma^2)$  is for (25).

Suppose  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$ . We know from (4) and (5) that  $\mathbf{K}\mathbf{x}$  has mean  $\mathbf{K}\boldsymbol{\mu}$  and dispersion matrix  $\mathbf{K}\mathbf{V}\mathbf{K}'$ . Using the result that a moment generating function does, under wide regularity conditions, uniquely define a corresponding density function, we use the moment generating function of (26) to show that

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V}) \quad \text{implies} \quad \mathbf{K}\mathbf{x} \sim N(\mathbf{K}\boldsymbol{\mu}, \mathbf{K}\mathbf{V}\mathbf{K}'). \quad (27)$$

The definition of  $m_{\mathbf{x}}(\mathbf{t})$  shown in (26) is equivalent to

$$m_{\mathbf{x}}(\mathbf{t}) = E e^{t'\mathbf{x}}. \quad (28)$$

Therefore the moment generating function of  $\mathbf{K}\mathbf{x}$  is

$$m_{\mathbf{K}\mathbf{x}}(\mathbf{t}) = E e^{t'\mathbf{K}\mathbf{x}} = E e^{(\mathbf{K}'\mathbf{t})'\mathbf{x}}. \quad (29)$$

The last expression in (29) is, by analogy with (28),  $m_{\mathbf{x}}(\mathbf{K}'\mathbf{t})$ . Thus

$$m_{\mathbf{K}\mathbf{x}}(\mathbf{t}) = m_{\mathbf{x}}(\mathbf{K}'\mathbf{t}). \quad (30)$$

Hence using  $\mathbf{K}'\mathbf{t}$  in place of  $\mathbf{t}$  in (26) for the right-hand side of (30) gives

$$m_{\mathbf{K}\mathbf{x}}(\mathbf{t}) = e^{t'\mathbf{K}\boldsymbol{\mu} + \frac{1}{2}t'\mathbf{K}\mathbf{V}\mathbf{K}'\mathbf{t}}. \quad (31)$$

Comparing (31) with (26) leads to conclusion (27).

**Illustration** (continued). Suppose that  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$ . Then for  $\mathbf{y} = \mathbf{U}'\mathbf{x}$  of (24), with  $\mathbf{U}$  orthogonal, (27) gives  $\mathbf{y} \sim N(\mathbf{U}'\mathbf{0}, \mathbf{U}'\mathbf{I}\mathbf{U})$ , i.e.,  $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I})$ . This is always a consequence of an orthogonal transformation, that the distribution of an orthogonal transform of  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$  has the same distribution as  $\mathbf{x}$ .

In the case of (24) there is a further consequence. The  $y_i$ 's are normally distributed with zero mean and unit variance, so each  $y_i^2$  is distributed as a  $\chi^2$  on 1 degree of freedom. The  $y_i$ 's are also independent, so

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^{N-1} y_i^2$$

of (24) is distributed as the sum of  $N - 1$  independent 1-degree-of-freedom  $\chi^2$ 's, i.e., as  $\chi^2$  on  $N - 1$  degrees of freedom.

The brevity of matrix notation is well illustrated in the preceding expressions and is further evident in considering marginal and conditional distributions of the multivariate normal distribution. Suppose  $\mathbf{x}$ ,  $\boldsymbol{\mu}$  and  $\mathbf{V}$  are partitioned conformably as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \quad (32)$$

with  $\mathbf{V}_{21} = (\mathbf{V}_{12})'$ . Define  $\mathbf{W}_{11} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ , the Schur complement (see Section 10.4) of  $\mathbf{V}_{22}$  in  $\mathbf{V}$ . Then it can be shown (e.g., Searle, 1971, Section 2.4f) that the marginal distribution of  $\mathbf{x}_1$  is  $\mathbf{x}_1 \sim N(\boldsymbol{\mu}_1, \mathbf{V}_{11})$ ; and the conditional distribution of  $\mathbf{x}_1$  given  $\mathbf{x}_2$  is  $N[\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \mathbf{W}_{11}]$ .

These and numerous other uses of matrix algebra are widely found wherever multivariate statistical methods are discussed. For example, when the  $k$  random variables underlying the data matrix  $\mathbf{X}_{N \times k}$  of (9) are normally distributed,  $N(\boldsymbol{\mu}, \mathbf{V})$ , then the  $\frac{1}{2}k(k+1)$  distinct elements  $s_{ij}$  for  $i \leq j = 1, \dots, k$  of the Wishart matrix  $\mathbf{S}$  of (21) are jointly distributed with probability density function

$$f(s_{11}, s_{12}, \dots, s_{kk}) = \frac{|\mathbf{S}|^{\frac{1}{2}(N-k-1)} e^{-\frac{1}{2}\text{tr}(\mathbf{V}^{-1}\mathbf{S})}}{2^{\frac{1}{2}Nk} \pi^{\frac{1}{2}k(k-1)} \prod_{i=1}^k \Gamma_{\frac{1}{2}}(N+1-i)}$$

This is known as the *Wishart density function*.

In being elements of a matrix, the random variables of the Wishart density are not conveniently arrayed for considering their dispersion matrix. But this inconvenience can be overcome by using the *vec* and *vech* operators of Section 12.9, because *vech*( $\mathbf{S}$ ) is the vector of  $\frac{1}{2}k(k+1)$  distinct elements

of  $\mathbf{S}$ . Then it can be shown (see Henderson and Searle, 1979) that the dispersion matrix of elements of  $\mathbf{S}$  is

$$\text{var}(\text{vec } \mathbf{S}) = k(\mathbf{V} \otimes \mathbf{V})(\mathbf{I}_{k^2} + \mathbf{I}_{k,k})$$

where  $\mathbf{I}_{k^2}$  is an identity matrix of order  $k^2$ , and  $\mathbf{I}_{k,k}$  is a *vec*-permutation matrix (Henderson and Searle, 1981b) of Section 12.9.

### 5. QUADRATIC FORMS AND $\chi^2$ -DISTRIBUTIONS

Any sum of squares in an analysis of variance is a quadratic function of observations. Representing the observations by  $\mathbf{x}$ , a sum of squares is therefore the quadratic form  $\mathbf{x}'\mathbf{A}\mathbf{x}$  for some symmetric matrix  $\mathbf{A}$ .

Hypothesis testing and variance component estimation often utilize expected values of sums of squares, i.e., of  $E(\mathbf{x}'\mathbf{A}\mathbf{x})$ . When  $\mathbf{x}$  has mean  $\boldsymbol{\mu}$  and dispersion matrix  $\mathbf{V}$  as in (1) and (3), we obtain  $E(\mathbf{x}'\mathbf{A}\mathbf{x})$  by first writing  $\mathbf{V} = E(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = E(\mathbf{x}\mathbf{x}') - \boldsymbol{\mu}\boldsymbol{\mu}'$ , which is the multivariate analogue of the well-known univariate expression  $\sigma^2 = E(x^2) - \mu^2$ . Hence  $E(\mathbf{x}\mathbf{x}') = \mathbf{V} + \boldsymbol{\mu}\boldsymbol{\mu}'$ . Then, because  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is a scalar, it equals its trace,  $\mathbf{x}'\mathbf{A}\mathbf{x} = \text{tr}(\mathbf{x}'\mathbf{A}\mathbf{x})$ , so that on using  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$  from Section 2.7,

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = E \text{tr}(\mathbf{x}'\mathbf{A}\mathbf{x}) = E \text{tr}(\mathbf{A}\mathbf{x}\mathbf{x}') = \text{tr}[\mathbf{A}E(\mathbf{x}\mathbf{x}')],$$

thus giving

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = \text{tr}[\mathbf{A}(\mathbf{V} + \boldsymbol{\mu}\boldsymbol{\mu}')] = \text{tr}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}. \quad (33)$$

This result holds whether or not  $\mathbf{x}$  is normally distributed. When it is, there is a further result for the variance of  $\mathbf{x}'\mathbf{A}\mathbf{x}$ :

$$\text{var}(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2 \text{tr}(\mathbf{A}\mathbf{V})^2 + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}. \quad (34)$$

Results (33) and (34), the latter under normality, apply quite generally for any quadratic form  $\mathbf{x}'\mathbf{A}\mathbf{x}$ . They are therefore available for providing the mean and variance of any sum of squares (expressed as a quadratic form) that may be of interest. Furthermore, generalizations to bilinear forms (end of Section 2.5) are also available, by writing a bilinear form as a quadratic form,

$$\mathbf{x}'\mathbf{A}\mathbf{y} = [\mathbf{x}' \quad \mathbf{y}] \begin{bmatrix} \frac{1}{2}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (35)$$

and applying results like (33) and (34) to the right-hand side of (35). A variety of such results is given in Searle (1971, Sections 2.5 and 2.6).

$F$ -statistics are ratios of mean squares or, more generally, of quadratic forms. And in order for such ratios to have an  $F$ -distribution, the numerator and denominator quadratic forms must be distributed independently of each other and each must have a  $\chi^2$ -distribution. Theory dictates that for  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$  these requirements generalize as follows.

**Theorem.** (i)  $\mathbf{x}'\mathbf{A}\mathbf{x}$  and  $\mathbf{x}'\mathbf{B}\mathbf{x}$  are distributed independently if  $\mathbf{A}\mathbf{V}\mathbf{B} = \mathbf{0}$ ; and (ii)  $\mathbf{x}'\mathbf{A}\mathbf{x}$  has a  $\chi^2$ -distribution with  $r(\mathbf{A}\mathbf{V})$  degrees of freedom if  $\mathbf{A}\mathbf{V}$  is idempotent.

**Illustration** (continued). The distribution of  $\sum_{i=1}^N (x_i - \bar{x})^2$  has already been considered for  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$ . Here we consider it for  $\mathbf{x} \sim N(\boldsymbol{\mu}\mathbf{1}, \sigma^2\mathbf{I})$ , i.e., each  $x_i \sim N(\mu, \sigma^2)$ , with the  $x_i$ 's being independent. Then, using

$$\sum_{i=1}^N (x_i - \bar{x})^2 / \sigma^2 = \mathbf{x}'(\mathbf{C}_N / \sigma^2)\mathbf{x}, \quad (36)$$

we can invoke (ii) of the theorem just stated to show that (36) has a  $\chi^2$ -distribution. With  $\mathbf{A}$  and  $\mathbf{V}$  of the theorem being  $\mathbf{C}_N / \sigma^2$  and  $\sigma^2\mathbf{I}$ , respectively,  $\mathbf{A}\mathbf{V} = \mathbf{C}_N$ , which (by Section 12.3f) is idempotent. And with  $\mathbf{C}_N\mathbf{x}$  having expected value  $\mathbf{C}_N\boldsymbol{\mu}\mathbf{1} = \boldsymbol{\mu}\mathbf{C}_N\mathbf{1} = \mathbf{0}$ , and  $r_{\mathbf{C}_N}$  being  $r_{\mathbf{C}_N} = N - 1$ , part (ii) of the theorem gives

$$\sum_{i=1}^N (x_i - \bar{x})^2 / \sigma^2 \sim \chi_{N-1}^2.$$

Extension to data from a completely randomized design experiment is given as Exercise 4. General results of this nature are to be found in Searle (1971, Chapters 3 and 5-8).

**Illustration.** The  $r$ th cumulant  $k_r$  of  $\mathbf{x}'\mathbf{A}\mathbf{x}$  for  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{V})$  is given by the identity in  $t$  (e.g., Searle, 1971, p. 55):

$$\sum_{r=1}^{\infty} k_r t^r / r! \equiv -\frac{1}{2} \log |\mathbf{I} - 2t\mathbf{A}\mathbf{V}|.$$

Deriving  $k_r$  explicitly, through expanding the right-hand side of this expression, provides an interesting illustration of using properties of eigenvalues.

Let  $\lambda_i(\mathbf{M})$  be the  $i$ th eigenvalue of the matrix  $\mathbf{M}$ . Then

$$\begin{aligned} \sum_{r=1}^{\infty} k_r t^r / r! &= -\frac{1}{2} \log \left\{ \prod_{i=1}^N [\lambda_i(\mathbf{I} - 2t\mathbf{A}\mathbf{V})] \right\} = -\frac{1}{2} \sum_{i=1}^N \log [\lambda_i(\mathbf{I} - 2t\mathbf{A}\mathbf{V})] \\ &= -\frac{1}{2} \sum_{i=1}^N \log [1 - 2t\lambda_i(\mathbf{A}\mathbf{V})] = -\frac{1}{2} \sum_{i=1}^N \sum_{r=1}^{\infty} -[2t\lambda_i(\mathbf{A}\mathbf{V})]^r / r \\ &= \frac{1}{2} \sum_{r=1}^{\infty} (2t)^r r^{-1} \sum_{i=1}^N [\lambda_i(\mathbf{A}\mathbf{V})]^r. \end{aligned}$$

Therefore, on equating coefficients of  $t^r$  from both sides,

$$k_r = 2^{r-1} r! r^{-1} \sum_{i=1}^N [\lambda_i(\mathbf{A}\mathbf{V})]^r = 2^{r-1} (r-1)! \text{tr}(\mathbf{A}\mathbf{V})^r.$$

Special cases of this result are  $r = 1$  and  $r = 2$ , which (for  $\boldsymbol{\mu} = \mathbf{0}$ ) give (33) and (34), respectively.

## 6. LEAST SQUARES EQUATIONS

The least squares equations

$$\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y} \quad (37)$$

have already been introduced in Sections 9.9 and 12.7. A more general form is

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad (38)$$

known as the *generalized least squares (GLS) equations*, or *Aitken equations*. In contrast, (37) represents the *ordinary least squares (OLS) equations*.

In the GLS equations,  $\mathbf{V}$  is a dispersion matrix and can, in many situations, be taken as positive definite (Section 1) and therefore  $\mathbf{V}^{-1}$  can be written as  $\mathbf{V}^{-1} = \mathbf{L}'\mathbf{L}$  for some  $\mathbf{L}$  (Section 7.7c). On defining  $\mathbf{W} = \mathbf{L}\mathbf{X}$  and  $\mathbf{z} = \mathbf{L}'\mathbf{y}$ , the GLS equations (38) then reduce to  $\mathbf{W}'\mathbf{W}\hat{\mathbf{b}} = \mathbf{W}'\mathbf{z}$ , which are the same form as the OLS equations (37), and can be handled similarly. They are considered at length in Chapters 14 and 15.



The matrix  $\mathbf{X}$  in this context is often partitioned as  $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$ , in which case with the conformable partitioning of  $\mathbf{b}$  and  $\mathbf{y}$  the OLS equations become

$$\begin{bmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1 \mathbf{y} \\ \mathbf{X}'_2 \mathbf{y} \end{bmatrix}. \quad (39)$$

When the matrix on the left of (39) is nonsingular, two expressions for its inverse are available in Exercise 10.7. Otherwise, a generalized inverse can be used from  $Q^*$  of Section 10.5. Thus can (39) be solved.

An important calculation based on (39) is  $R(\mathbf{b}_1, \mathbf{b}_2) = \hat{\mathbf{b}}'(\mathbf{X}'\mathbf{y})$ , the reduction in sum of squares. In the case of  $\mathbf{X}'\mathbf{X}$  being nonsingular and using the inverse of Exercise 10.7, this becomes

$$R(\mathbf{b}_1, \mathbf{b}_2) = \mathbf{y}'\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y} + \mathbf{y}'\mathbf{M}_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{M}_1\mathbf{y}. \quad (40)$$

for

$$\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1, \quad (41)$$

analogous to  $\mathbf{M}$  of Exercise 8.27. Defining  $R(\mathbf{b}_1)$  as

$$R(\mathbf{b}_1) = \mathbf{y}'\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y} \quad (42)$$

gives the difference

$$R(\mathbf{b}_1, \mathbf{b}_2) - R(\mathbf{b}_1) = \mathbf{y}'\mathbf{M}_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{M}_1\mathbf{y}, \quad (43)$$

which has important uses as a sum of squares in analysis of variance for regression and linear models (see Sections 14.15 and 15.13). The reader should verify (43) and develop a similar expression using  $R(\mathbf{b}_2)$  in place of  $R(\mathbf{b}_1)$ , where  $R(\mathbf{b}_2)$  is defined exactly like (42) with the subscript 2 in place of 1. Similar results can also be obtained when  $\mathbf{X}'\mathbf{X}$  is singular and the generalized inverses  $Q^*$  and  $Q^{**}$  of Section 10.5 have to be used.

## 7. CONTRASTS AMONG MEANS

In multi-factorial experiments the calculation of estimated contrasts can be specified using direct products of the matrices and vectors involved in one-factor experiments. The procedure is illustrated with a simple example.

First, some notation must be established. Suppose a horticulturalist carries out an experiment involving two fertilizer treatments on three

TABLE 13.1. TABLE OF MEANS

Treatment (A)	Variety (B)			Treatment Mean
	1	2	3	
1	$ab_{11}$	$ab_{12}$	$ab_{13}$	$a_1$
2	$ab_{21}$	$ab_{22}$	$ab_{23}$	$a_2$
Variety mean	$b_1$	$b_2$	$b_3$	Grand mean

varieties of wheat. Denote treatments by the letter A and varieties by B, the observed treatment means being  $a_1$  and  $a_2$ , and the variety means being  $b_1$ ,  $b_2$  and  $b_3$ . With notation  $ab_{ij}$  for treatment  $\times$  variety means, Table 13.1 is the table of means. Furthermore, let  $I$ , a scalar, represent the overall population mean, let  $A_1$  be the linear effect among the treatments,  $B_1$  the linear effect among the varieties and  $B_2$  the quadratic effect among the varieties. These are some of the contrasts to be estimated.

For the moment, assume first that there had been only one variety of wheat, and second that there had been no fertilizer treatments. Then the estimates of  $I$  and  $A_1$  in the first case, and of  $I$ ,  $B_1$  and  $B_2$  in the second would, respectively, be

$$\begin{bmatrix} \hat{I} \\ \hat{A}_1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \hat{I} \\ \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix},$$

where  $\hat{I}$ , for example, indicates an estimate of  $I$ . But, on taking treatments and varieties into account simultaneously, estimates of  $I$ ,  $A_1$ ,  $B_1$ ,  $B_2$  and the interactions  $A_1B_1$  and  $A_1B_2$  are given by the direct product of the preceding equations which, by property (v) of Section 10.7, is

$$\begin{bmatrix} \hat{I} \\ \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} \otimes \begin{bmatrix} \hat{I} \\ \hat{A}_1 \end{bmatrix} = \frac{1}{6} \left\{ \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{bmatrix} \otimes \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right\} \left\{ \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \otimes \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right\}.$$

In expanding the direct products of vectors in this equation the following convention is adopted: although  $a_1b_1$  as a product is the product of two means, it is interpreted as the mean  $ab_{11}$ . Similarly,  $\hat{A}_1\hat{B}_1$  is interpreted as

$\widehat{AB}_{12}$ . With this convention, the preceding equation expands to

$$\begin{bmatrix} \hat{I} \\ \hat{A}_1 \\ \hat{B}_1 \\ \widehat{AB}_{11} \\ \hat{B}_2 \\ \widehat{AB}_{12} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ -1 & -1 & 2 & 2 & -1 & -1 \\ 1 & -1 & -2 & 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} ab_{11} \\ ab_{21} \\ ab_{12} \\ ab_{22} \\ ab_{13} \\ ab_{23} \end{bmatrix}$$

This specifies the calculation of all the estimated contrasts.

The general result is clear: if  $\hat{\mu}_\beta$  is the vector of estimated effects for the  $\beta$ -factor, e.g.,  $\hat{\mu}'_\beta = [I \ \hat{B}_1 \ \hat{B}_2]$ ; if  $\mathbf{m}_\beta$  is the corresponding vector of observed means, e.g.,  $\mathbf{m}'_\beta = [b_1 \ b_2 \ b_3]$ ; and if  $\mathbf{A}_\beta$  is the matrix such that  $\hat{\mu}_\beta = \mathbf{A}_\beta \mathbf{m}_\beta$  when all other factors are ignored, then estimated contrasts for a series of factors  $\alpha, \beta, \gamma, \delta$  say, are given by

$$\hat{\mu}_\delta \otimes \hat{\mu}_\gamma \otimes \hat{\mu}_\beta \otimes \hat{\mu}_\alpha = (\mathbf{A}_\delta \otimes \mathbf{A}_\gamma \otimes \mathbf{A}_\beta \otimes \mathbf{A}_\alpha)(\mathbf{m}_\delta \otimes \mathbf{m}_\gamma \otimes \mathbf{m}_\beta \otimes \mathbf{m}_\alpha).$$

The convention of replacing  $a_1 b_1$ , for example, by  $ab_{11}$ , applies to all elements of the direct products of vectors in this expression.

Rules for establishing equations such as these have been suggested by several authors, for example, Yates (1937) and Fisher (1951), and are summarized by Federer (1955). The generality of the expression is appealing. It is easily remembered, is universally true no matter how many factors are involved, it eliminates the need for remembering lengthy detailed results and it is easily derived from general, direct-product results. Extensions of these ideas are to be found in many places, for example, Robson (1959), Kurkjian and Zelen (1962) and Zelen and Federer (1965).

Kronecker products are also helpful in variance components models for experimental data (e.g., Searle and Henderson, 1979) and are made use of in a wide variety of other statistical analyses (see, for example, Henderson and Searle, 1981b).

EXERCISES

1. If the dispersion matrix of  $\mathbf{x}$  is  $\mathbf{I}$ , and if the rows of  $\mathbf{T}$  are orthonormal, prove that  $\mathbf{y} = \mathbf{T}\mathbf{x}$  also has an identity matrix as its dispersion matrix.
2. Suppose  $\theta$  is subtracted from every element of a data matrix. Show that the Wishart matrix is unchanged.

3. Consider  $\mathbf{Y} = [\mathbf{I} \ \mathbf{X}]$ . Show that

$$(\mathbf{Y}'\mathbf{Y})^{-1} = \begin{bmatrix} \frac{1}{N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\bar{\mathbf{x}}' \\ \mathbf{I} \end{bmatrix} \mathbf{S}^{-1} [-\bar{\mathbf{x}} \ \mathbf{I}],$$

for  $\bar{\mathbf{x}}'$  of (12) and  $\mathbf{S}$  of (21).

4. In a completely randomized experiment of  $n$  observations in each of  $a$  classes, let  $x_{ij}$  be the  $j$ th observation in the  $i$ th class for  $i = 1, \dots, a$  and  $j = 1, \dots, n$ . Define  $\mathbf{x}$  as the vector of observations in lexicon order

$$\mathbf{x}' = [x_{11} \ \dots \ x_{1n} \ x_{21} \ \dots \ x_{2n} \ \dots \ \dots \ x_{i1} \ \dots \ x_{in} \ \dots \ \dots \ x_{a1} \ \dots \ x_{an}]$$

and define the means

$$\bar{x}_{i.} = \sum_{j=1}^n \frac{x_{ij}}{n} \quad \text{and} \quad \bar{x}_{.j} = \sum_{i=1}^a \sum_{j=1}^n \frac{x_{ij}}{an}.$$

Then in the analysis of variance of such data the two sums of squares are

$$\text{SSA} = n \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})^2 \quad \text{and} \quad \text{SSE} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2.$$

Show that

- (a)  $\text{SSA} = \mathbf{x}'(\mathbf{I}_a \otimes \bar{\mathbf{J}}_n - \bar{\mathbf{J}}_{an})\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x}$ , so defining  $\mathbf{A}$ .
- (b)  $\text{SSE} = \mathbf{x}'(\mathbf{I}_{an} - \mathbf{I}_a \otimes \bar{\mathbf{J}}_n)\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$ , so defining  $\mathbf{B}$ .
- (c)  $\mathbf{I}_{an} = \mathbf{I}_a \otimes \mathbf{I}_n$  and  $\bar{\mathbf{J}}_{an} = \bar{\mathbf{J}}_a \otimes \bar{\mathbf{J}}_n$ .
- (d)  $\mathbf{A}\mathbf{I}_{an}$  and  $\mathbf{B}(\mathbf{I}_a \otimes \mathbf{I}_n)$  are both null.

Assume normality of the  $x_{ij}$ 's with dispersion matrix  $\sigma^2 \mathbf{I}_{an}$ . Prove

- (e)  $\text{SSA}/\sigma^2$  and  $\text{SSB}/\sigma^2$  are distributed independently.
  - (f)  $\text{SSA}/\sigma^2$  and  $\text{SSB}/\sigma^2$  are each distributed as a  $\chi^2$ -distribution, with degrees of freedom  $a - 1$  and  $a(n - 1)$ , respectively.
5. For  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  of order  $n$  show that for  $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I})$ 
    - (a)  $E(\mathbf{y}'\mathbf{M}\mathbf{y}) = (n - r_{\mathbf{X}})\sigma^2$ .
    - (b)  $\mathbf{y}'\mathbf{M}\mathbf{y}$  and  $\mathbf{y}'(\mathbf{I} - \mathbf{M})\mathbf{y}$  are independent.
    - (c)  $\mathbf{y}'\mathbf{M}\mathbf{y}/\sigma^2 \sim \chi^2_{n-r_{\mathbf{X}}}$ .
    - (d)  $\mathbf{y}'(\mathbf{I} - \mathbf{M})\mathbf{y}/\sigma^2 \sim \chi^2_{r_{\mathbf{X}}}$  when  $\mathbf{X}\mathbf{b} = \mathbf{0}$ .

6. When  $\mathbf{X}$  is a matrix of random variables there is no such thing as the dispersion matrix of  $\mathbf{X}$ . In contrast, the dispersion matrix of  $\text{vec } \mathbf{X}$  (see Section 12.8) contains variances and covariances of all elements of  $\mathbf{X}$ . Suppose  $\mathbf{x}_j$  and  $\mathbf{x}_k$  are columns of  $\mathbf{X}$  having a matrix of covariances  $\mathbf{M}_{jk}$ :

$$\text{cov}(\mathbf{x}_j, \mathbf{x}_k) = \mathbf{M}_{jk}.$$

Then the variance-covariance matrix of  $\text{vec } \mathbf{X}$ , to be denoted by  $\text{var}(\text{vec } \mathbf{X})$ , is the partitioned matrix having submatrices  $\mathbf{M}_{jk}$ :

$$\text{var}(\text{vec } \mathbf{X}) = \{\mathbf{M}_{jk}\}.$$

For the special case  $\mathbf{M}_{jk} = m_{jk}\mathbf{V}$  with  $\mathbf{M} = \{m_{ij}\}$ , prove

- (a)  $\text{var}(\text{vec } \mathbf{X}) = \mathbf{M} \otimes \mathbf{V}$ ,  
 (b)  $\text{var}(\text{vec } \mathbf{TX}) = \mathbf{M} \otimes \mathbf{TVT}'$ ,  
 (c)  $\text{var}(\text{vec } \mathbf{X}') = \mathbf{V} \otimes \mathbf{M}$ .
7. (a) Explain why  $\mathbf{X}$  having full column rank leads to the GLS equations (38) having solution  $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ .  
 (b) Suppose  $E(\mathbf{Y}_{N \times p}) = \mathbf{X}_{N \times q}\mathbf{B}_{q \times p}$  with  $\text{cov}(\mathbf{y}_j, \mathbf{y}_k) = \sigma_{jk}\mathbf{I}_N$  for  $\mathbf{y}_j$  and  $\mathbf{y}_k$  being two columns of  $\mathbf{Y}$ . Prove that the generalized least squares estimator of  $\mathbf{B}$  is  $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ .  
 [Hint: use the vec operator.]
8. Suppose

$$\text{var}\begin{bmatrix} X & Y & Z \end{bmatrix}' = \begin{bmatrix} 8 & -7 & 2 \\ -7 & 10 & -3 \\ 2 & -3 & 1 \end{bmatrix},$$

where  $X$ ,  $Y$  and  $Z$  are three random variables.

- (a) Prove that the  $3 \times 3$  matrix is legitimate for this purpose.  
 (b) Show that  $X + 2Y + 3Z$  and  $Y + 4Z$  have variances 5 and 2, respectively. Are they independent?
9. Prove that the determinant of  $\mathbf{R}$  in (8) does not exceed unity.  
 [Hint: Use Section 5.2.]
10. Why is  $\mathbf{S}$  of (21) non-negative definite? Assuming it is positive definite, define an estimated correlation matrix  $\hat{\mathbf{R}}$ , analogous to (8), and show that its determinant does not exceed unity.
11. Prove (28).

## CHAPTER 14

### THE MATRIX ALGEBRA OF REGRESSION ANALYSIS

We conclude this book with chapters devoted to regression and linear models. Both topics are ideally suited to being described in matrix terminology and gain considerable clarity therefrom. Only a modest acquaintance with statistics is assumed, for it is hoped that the chapters will spark in the reader an interest both in the statistical methods themselves and in the merits of using matrix algebra for describing them. Such references as Snedecor and Cochran (1967), Kendall and Stuart (1974), Mood, Graybill and Boes (1974) and Steel and Torrie (1980) give readers the relevant details on statistics, while texts like Kempthorne (1952), Searle (1971), Rao (1973), Graybill (1976), Seber (1977) and Arnold (1981) provide more extensive matrix descriptions than are presented here.

#### 1. GENERAL DESCRIPTION

**Illustration.** The feed consumed by a beef animal is used for maintaining body weight and for making gains in body weight. In studying the feed intake of beef cows Lemenager et al. (1980) therefore considered daily intake as a linear combination of the weight of a cow and a body conformation score ranging from 1 for very thin to 5 for very fat. From measurements on these three variables on a sample of cows they estimate the linear combination as

daily nutrient requirement in kilograms

$$= 4.08 + 0.0148(\text{weight in kilograms}) - 2.3012(\text{conformation score}).$$

### a. Linear models

The technique of regression analysis is just that of the illustration: of studying situations in which a variable is thought to be some linear combination of one or more other variables, usually measured on the same observational unit; and from a series of such measurements estimating the coefficients in that linear combination. To be specific, suppose that on each of  $N$  units we measure three variables  $y$ ,  $x_1$  and  $x_2$ . Suppose further, that for given values of  $x_1$  and  $x_2$ , we assume the average value of  $y$  in the population of units (of which we are assuming we have just a random sample) is of the form

$$E(y) = a + b_1x_1 + b_2x_2. \quad (1)$$

In this equation  $E(y)$  represents the average or expected value of  $y$  corresponding to values of  $x_1$  and  $x_2$ ; and  $a$ ,  $b_1$  and  $b_2$  are some unknown constants. Regression analysis simply uses the observations on  $y$  and the  $x$ 's to obtain estimates of the  $b$ 's in equation (1). The variable denoted by  $y$  is usually called the *dependent variable* (because it depends on the  $x$ 's) and the  $x$ -variables are correspondingly referred to as *independent variables* or nowadays often as *regressor variables*.

Equation (1) is an attempt at describing how we think one variable is related to others, and as such is an example of *model building*. The model here is said to be a *linear model*, linear because we envisage  $E(y)$  as linear in the unknowns  $a$ ,  $b_1$  and  $b_2$ , which are called *parameters*. As such, the analysis we are to describe, referred to so far as regression analysis, is more correctly called *linear regression*, or (for more than one  $x$ -variable) *multiple linear regression analysis*.

### b. Observations

In the illustration of the preceding section, the feed intake of a cow with weight  $x_1$  and score  $x_2$  will not be exactly  $a + b_1x_1 + b_2x_2$ . Indeed, this is already recognized in the writing of the model equation (1) as  $E(y) = a + b_1x_1 + b_2x_2$  rather than as  $y = a + b_1x_1 + b_2x_2$ . More particularly, if  $y_i$ ,  $x_{i1}$  and  $x_{i2}$  are the values of  $y$ ,  $x_1$  and  $x_2$  for cow  $i$ , we write

$$E(y_i) = a + b_1x_{i1} + b_2x_{i2}, \quad (2)$$

where  $E(y_i)$  is not the same as  $y_i$ . The difference,  $y_i - E(y_i)$ , represents the deviation of the observed  $y_i$  from its expected value  $E(y_i)$  and is written as

$$e_i = y_i - E(y_i) = y_i - a - b_1x_{i1} - b_2x_{i2}. \quad (3)$$

Hence

$$y_i = a + b_1x_{i1} + b_2x_{i2} + e_i. \quad (4)$$

The deviation  $e_i$  defined in (3) represents the extent to which an observed  $y_i$  differs from its expected value  $E(y_i)$ . Thus the  $e$ 's include all manner of discrepancies between observed  $y$ 's and their expected values and, as such, are considered as random variables, usually called *random errors* or *random residuals*.

A change in notation is now made: in place of  $a$  write  $b_0$ , and then for  $b_0$  write  $b_0x_{i0}$  with all values of  $x_{i0}$  being unity. This gives

$$y_i = b_0x_{i0} + b_1x_{i1} + b_2x_{i2} + e_i, \quad (5)$$

and this equation occurs for every set of observations  $y_i$ ,  $x_{i0} = 1$ ,  $x_{i1}$  and  $x_{i2}$  that constitutes the data; i.e., (5) exists for  $i = 1, 2, \dots, N$ . Now define the following matrix and vectors:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ x_{N0} & x_{N1} & x_{N2} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}. \quad (6)$$

Then the complete set of equations represented by (5) is

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}. \quad (7)$$

Equation (7) is called the *equation of the model*, or *model equation*. It is the basis for estimating  $\mathbf{b}$  from the observations arrayed in  $\mathbf{X}$  and  $\mathbf{y}$ .

### c. Nonlinear models

Numerous other models could be postulated, many of which are nonlinear in their parameters. For example, instead of (4) one could, if appropriate, postulate

$$y_i = \alpha x_1^\beta x_2^\gamma \epsilon_i. \quad (8)$$

This is a nonlinear model because the parameters  $\alpha$ ,  $\beta$  and  $\gamma$  occur nonlinearly; but  $E(y) = a + b_1x_1 + b_2x_1^2$  is a linear model because it is linear in  $a$  and the  $b$ 's.

Although the variety of nonlinear models is extensive, linear models are important, both in theory and in practice, because they are highly tractable mathematically and because, in a wide range of applications, they have been found of great value. Furthermore, many nonlinear models can often be rearranged to be linear; for example, (8) can be rearranged as the linear

model

$$\log y = \alpha + \beta \log x_1 + \gamma \log x_2 + \log \epsilon. \quad (9)$$

By these means, nonlinear relationships between variables can often be reduced to linear ones and studied analytically using linear regression. It is therefore a technique that has wide application.

2. ESTIMATION

The customary estimation procedure is to apply the method of "Least Squares" to (7). It results in choosing as the estimator of  $\mathbf{b}$  those values of  $b_0, b_1$  and  $b_2$  which minimize the expression  $\sum_{i=1}^N [y_i - E(y_i)]^2$ . From (2) and (3) this sum of squares is

$$\begin{aligned} \sum_{i=1}^N [y_i - E(y_i)]^2 &= \sum_{i=1}^N [y_i - (b_0 x_{i0} + b_1 x_{i1} + b_2 x_{i2})]^2 = \sum_{i=1}^N e_i^2 \\ &= (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) = \mathbf{e}'\mathbf{e}. \end{aligned}$$

This is the expression considered in the first illustration of Section 12.7; equation (14) there is identical to (7) of this chapter. Consequently the estimator of  $\mathbf{b}$ , which will be denoted by  $\hat{\mathbf{b}}$ , is as derived earlier, namely from

$$\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}. \quad (10)$$

comes

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (11)$$

Since every  $x_{i0}$  in (5) is unity, and every  $x_{i1}$  and  $x_{i2}$  is an observation,  $\mathbf{X}$  is a real matrix. Therefore  $\mathbf{X}'\mathbf{X}$  of (10) is n.n.d. (Section 12.4). And in practically all instances of regression analysis  $\mathbf{X}$  has full column rank, making  $\mathbf{X}'\mathbf{X}$  p.d. Hence  $(\mathbf{X}'\mathbf{X})^{-1}$  of (11) exists. In Chapter 15 we consider situations when  $(\mathbf{X}'\mathbf{X})^{-1}$  does not exist.

**Notation.** A common notation is  $\beta$  for the vector of parameters and  $\mathbf{b}$  for its estimator. The convention adopted here of  $\mathbf{b}$  for parameters and  $\hat{\mathbf{b}}$  for estimators of  $\mathbf{b}$  is used in order to preserve  $\beta$  for later use as a subvector of  $\mathbf{b}$ .

The nature of  $\mathbf{X}$  in (6) decrees that  $\mathbf{X}'\mathbf{X}$  is the square matrix of sums of squares and cross-products of the  $x$ 's: in this instance the  $3 \times 3$  matrix

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \sum x_{i0}^2 & \sum x_{i0}x_{i1} & \sum x_{i0}x_{i2} \\ \sum x_{i0}x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i0}x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix} = \begin{bmatrix} N & x_{\cdot 1} & x_{\cdot 2} \\ x_{\cdot 1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ x_{\cdot 2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix}. \quad (12)$$

All summations here are for  $i = 1, 2, \dots, N$ . Thus  $\mathbf{X}'\mathbf{X}$  is just like (10) of Section 13.3b, except for utilizing  $x_{i0} = 1$  for  $i = 1, \dots, N$ , which gives

$$\sum_{i=1}^N x_{i0}^2 = \sum_{i=1}^N 1^2 = N \quad \text{and} \quad \sum_{i=1}^N x_{i0}x_{i1} = \sum_{i=1}^N x_{i1} = x_{\cdot 1}, \quad (13)$$

$x_{\cdot 1}$  being the total of the  $Nx_{i1}$ 's; similarly for  $x_{\cdot 2}$ .

$\mathbf{X}'\mathbf{y}$  is also needed for  $\hat{\mathbf{b}}$  in (11). Similar to (12), it is the vector of sums of cross-products of the  $x$ 's with the  $y$ 's, in this case the  $3 \times 1$  vector

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum x_{i0}y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix} = \begin{bmatrix} N\bar{y} \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}. \quad (14)$$

Again the summations are for  $i = 1, 2, \dots, N$ ; and the occurrence of  $N\bar{y}$  in (14) arises in the same way as does  $N\bar{x}_{\cdot 1} = x_{\cdot 1}$  of (13) in (12).

Visualizing (12) and (14) substituted in (11) we see that  $\hat{\mathbf{b}}$  is the vector of sums of products of the  $x$ 's and  $y$ 's pre-multiplied by the inverse of the matrix of sums of squares and products of the  $x$ 's.

**Example.** Suppose we have the following five sets of observations (including  $x_{i0} = 1$  for  $i = 1, \dots, 5$ ):

$i$	$y_i$	$x_{i0}$	$x_{i1}$	$x_{i2}$
1	62	1	2	6
2	60	1	9	10
3	57	1	6	4
4	48	1	3	13
5	23	1	5	2
$N = 5$	$y = 250$	$x_{\cdot 0} = 5$	$x_{\cdot 1} = 25$	$x_{\cdot 2} = 35$
	$\bar{y} = 50$		$\bar{x}_{\cdot 1} = 5$	$\bar{x}_{\cdot 2} = 7$

Then the estimated  $\mathbf{b}$  is calculated as  $\hat{\mathbf{b}}$  from (11) using

$$\mathbf{y} = \begin{bmatrix} 62 \\ 60 \\ 57 \\ 48 \\ 23 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 2 & 6 \\ 1 & 9 & 10 \\ 1 & 6 & 4 \\ 1 & 3 & 13 \\ 1 & 5 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 250 \\ 1265 \\ 1870 \end{bmatrix}, \quad (15)$$

with

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 5 & 25 & 35 \\ 25 & 155 & 175 \\ 35 & 175 & 325 \end{bmatrix}^{-1} = \frac{1}{480} \begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix}. \quad (16)$$

Hence from (11)

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \frac{1}{480} \begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix} \begin{bmatrix} 250 \\ 1265 \\ 1870 \end{bmatrix} = \begin{bmatrix} 37 \\ \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix}, \quad (17)$$

and so the estimated regression line is

$$\widehat{E}(y) = 37 + \frac{1}{2}x_1 + 1\frac{1}{2}x_2. \quad (18)$$

### 3. SEVERAL REGRESSOR VARIABLES

The procedure for fitting more than two  $x$ -variables,  $k$  of them say, is a direct extension of the preceding section. Analogous to (5) the model equation is

$$y_i = b_0x_{i0} + b_1x_{i1} + b_2x_{i2} + \cdots + b_kx_{ik} + e_i. \quad (19)$$

The vector equation for the  $N$  sets of observations is still as in (7),

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e},$$

where  $\mathbf{y}$  and  $\mathbf{e}$  are exactly as in (6); and  $\mathbf{X}$ , similar to that in (6), is now of order  $N \times (k+1)$ :

$$\mathbf{X} = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1k} \\ x_{20} & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N0} & x_{N1} & \cdots & x_{Nk} \end{bmatrix}, \quad \text{and } \mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix}. \quad (20)$$

The equations for estimating  $\mathbf{b}$  are exactly as in (10),

$$\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}, \quad \text{with solution } \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (21)$$

as in (11). Likewise,  $\mathbf{X}'\mathbf{X}$  and  $\mathbf{X}'\mathbf{y}$  have precisely the same form as in (12) and (14) except they are now of order  $k+1$  instead of 3:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & x_{\cdot 1} & x_{\cdot 2} & \cdots & x_{\cdot k} \\ x_{\cdot 1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ik} \\ x_{\cdot 2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 & \cdots & \sum x_{i2}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{\cdot k} & \sum x_{i1}x_{ik} & \sum x_{i2}x_{ik} & \cdots & \sum x_{ik}^2 \end{bmatrix}, \quad \text{and } \mathbf{X}'\mathbf{y} = \begin{bmatrix} y_{\cdot} \\ \sum x_{i1}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}. \quad (22)$$

### 4. DEVIATIONS FROM MEANS

The matrix  $\mathbf{X}'\mathbf{X}$  and vector  $\mathbf{X}'\mathbf{y}$  shown in (22) have as elements the sums of squares and products of the observations. But it is well known that the regression coefficients  $b_1, \dots, b_k$  can be estimated using a matrix and vector that are just like  $\mathbf{X}'\mathbf{X}$  and  $\mathbf{X}'\mathbf{y}$  only involving sums of squares and products corrected for their means. Indeed, this is the customary manner in which estimates are calculated. We now establish this formulation.

First, in using the summing vector  $\mathbf{1}_N$  of order  $N$ , defining

$$\mathbf{X}_1 = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}, \quad \text{gives } \mathbf{X} = [\mathbf{1} \quad \mathbf{X}_1]. \quad (23)$$

Then using  $\bar{y}$  for the mean of the  $y$ 's, and  $\bar{\mathbf{x}}' = [\bar{x}_{\cdot 1} \quad \bar{x}_{\cdot 2} \quad \cdots \quad \bar{x}_{\cdot k}] = \mathbf{1}'\mathbf{X}_1/N$  for the vector of means as in (12) of Section 13.3c,

$$\begin{aligned} \hat{\mathbf{b}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{X}_1 \\ \mathbf{X}_1'\mathbf{1} & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{X}_1'\mathbf{y} \end{bmatrix} = \begin{bmatrix} N & N\bar{\mathbf{x}}' \\ N\bar{\mathbf{x}} & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1} \begin{bmatrix} N\bar{y} \\ \mathbf{X}_1'\mathbf{y} \end{bmatrix}. \end{aligned} \quad (24)$$

Now define  $\mathbf{b}' = [b_0 \quad \delta']$ , where  $\delta$  ("curly" bee) is  $\delta' = [b_1 \quad b_2 \quad \cdots \quad b_k]$ , the vector of regression coefficients excluding the "constant" term  $b_0$ . We will apply Exercise 13.3 to the inverse matrix in (24). This requires the Schur

complement

$$\begin{aligned} \mathbf{S} &= \mathbf{X}'_1 \mathbf{X}_1 - N\bar{x}(N^{-1})N\bar{x}' \\ &= \mathbf{X}'_1 \mathbf{C}_N \mathbf{X}_1, \quad \text{with } \mathbf{C}_N = \mathbf{I} - \bar{\mathbf{J}}_N, \text{ the centering matrix} \\ &= \mathcal{X}'\mathcal{X} \end{aligned}$$

for

$$\mathcal{X} = \mathbf{C}_N \mathbf{X}_1 = \{x_{ij} - \bar{x}_{.j}\} \quad \text{for } i = 1, \dots, N \text{ and } j = 1, \dots, k$$

with  $\mathbf{C}_N$  being idempotent. Thus  $\mathcal{X}$  ("curly"  $x$ ) is the matrix of the  $x$ -variable observations, each expressed as a deviation from its mean, and  $\mathbf{S}$  is the sum of squares of such deviations; i.e.,  $\mathbf{S}$  is the matrix of corrected sums of squares and products of the observed  $x$ 's. Then using Exercise 13.3 on (24) yields

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{b}_0 \\ \hat{\lambda} \end{bmatrix} = \left\{ \begin{bmatrix} 1/N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\bar{x}' \\ \mathbf{I} \end{bmatrix} \mathbf{S}^{-1} \begin{bmatrix} -\bar{x} & \mathbf{1} \end{bmatrix} \right\} \begin{bmatrix} N\bar{y} \\ \mathbf{X}'_1 \mathbf{y} \end{bmatrix} \quad (25)$$

$$= \begin{bmatrix} \bar{y} - \bar{x}' \mathbf{S}^{-1} (\mathbf{X}'_1 \mathbf{y} - N\bar{y}\bar{x}) \\ \mathbf{S}^{-1} (\mathbf{X}'_1 \mathbf{y} - N\bar{y}\bar{x}) \end{bmatrix}. \quad (26)$$

Hence

$$\hat{b}_0 = \bar{y} - \bar{x}' \hat{\lambda} \quad (27)$$

and

$$\hat{\lambda} = \mathbf{S}^{-1} (\mathbf{X}'_1 \mathbf{y} - N\bar{y}\bar{x}). \quad (28)$$

Furthermore, for  $\hat{\lambda}$ ,

$$\begin{aligned} \mathbf{X}'_1 \mathbf{y} - N\bar{y}\bar{x} &= \mathbf{X}'_1 \mathbf{y} - \mathbf{X}'_1 \mathbf{1} \bar{y} / N = \mathbf{X}'_1 (\mathbf{I} - \bar{\mathbf{J}}_N) \mathbf{y} \\ &= \mathbf{X}'_1 \mathbf{C}_N \mathbf{y} = \mathcal{X}' \mathbf{y} = \mathcal{X}' \mathbf{y}, \end{aligned} \quad (29)$$

where "curly"  $y$  is

$$\mathbf{y} = \mathbf{C}_N \mathbf{y} = \{y_i - \bar{y}\} \quad \text{for } i = 1, \dots, N. \quad (30)$$

Hence in (28)

$$\hat{\lambda} = (\mathcal{X}'\mathcal{X})^{-1} \mathcal{X}' \mathbf{y} = (\mathcal{X}'\mathcal{X})^{-1} \mathcal{X}' \mathbf{y}. \quad (31)$$

Thus  $\hat{\lambda}$  is precisely  $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$  with  $\mathbf{X}'\mathbf{X}$  and  $\mathbf{X}'\mathbf{y}$  replaced by  $\mathcal{X}'\mathcal{X}$  and  $\mathcal{X}'\mathbf{y}$  having elements that are corrected sums of squares and products of the observations.

**Special case.** The special case of just a single  $x$ -variable,  $k=1$ , is of particular interest. This is the case of simple regression, whereupon the model equation (1) is  $E(y_i) = a + bx_i$ , with

$$\mathcal{X}'\mathcal{X} = \sum_{i=1}^N (x_i - \bar{x})^2 \quad \text{and} \quad \mathcal{X}'\mathbf{y} = \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}).$$

Thus (27) and (31) give the familiar results

$$\hat{a} = \hat{b}_0 = \bar{y} - \hat{b}\bar{x} \quad \text{and} \quad \hat{b} = \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^N (x_i - \bar{x})^2. \quad (32)$$

**Example (continued).** From (15),

$$\mathcal{X} = \begin{bmatrix} -3 & -1 \\ 4 & 3 \\ 1 & -3 \\ -2 & 6 \\ 0 & -5 \end{bmatrix}, \quad \text{and so} \quad \mathcal{X}'\mathcal{X} = \begin{bmatrix} 30 & 0 \\ 0 & 80 \end{bmatrix}. \quad (33)$$

Hence with

$$(\mathcal{X}'\mathcal{X})^{-1} = \begin{bmatrix} \frac{1}{30} & 0 \\ 0 & \frac{1}{80} \end{bmatrix} \quad \text{and} \quad \mathcal{X}'\mathbf{y} = \begin{bmatrix} 15 \\ 120 \end{bmatrix} \quad (34)$$

(31) gives

$$\hat{\lambda} = \begin{bmatrix} \frac{1}{30} & 0 \\ 0 & \frac{1}{80} \end{bmatrix} \begin{bmatrix} 15 \\ 120 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix}, \quad (35)$$

as in (17). And from (27) comes the other result in (17):

$$\hat{b}_0 = 50 - \begin{bmatrix} \frac{1}{2} & 1\frac{1}{2} \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} = 37. \quad (36)$$

## 5. THE STATISTICAL MODEL

Statistical properties of  $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$  arise from properties attributed to elements  $e_i$  of the residual vector  $\mathbf{e}$  in the model equation  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ . As

in (3), and this is true generally,  $e_i$  is defined as  $e_i = y_i - E(y_i)$ , so that

$$\mathbf{e} = \mathbf{y} - E(\mathbf{y}) \quad \text{and hence} \quad E(\mathbf{e}) = E(\mathbf{y}) - E(\mathbf{y}) = \mathbf{0}. \quad (37)$$

Thus, by definition, the error terms  $e_i$  have zero means. And, of course,

$$E(\mathbf{y}) = \mathbf{y} - \mathbf{e} = \mathbf{X}\mathbf{b}. \quad (38)$$

The further assumption usually made about the  $e_i$ 's is that they all have the same variance  $\sigma^2$ , and that each of them is uncorrelated (has zero covariance) with all others. Hence the variance-covariance matrix of the  $e_i$ 's is taken as  $\sigma^2\mathbf{I}_N$ ; i.e., in terms of (2) and (3) of Section 13.1,

$$\text{var}(\mathbf{e}) = E[\mathbf{e} - E(\mathbf{e})][\mathbf{e} - E(\mathbf{e})]' = E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{I}_N. \quad (39)$$

#### 6. UNBIASEDNESS AND VARIANCES

The expected value of  $\hat{\mathbf{b}}$  is  $\mathbf{b}$ . This is so because, with  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$  from (38),

$$E(\hat{\mathbf{b}}) = E(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{b};$$

i.e.,  $\hat{\mathbf{b}}$  is an unbiased estimator of  $\mathbf{b}$ .

The dispersion matrix of  $\mathbf{y}$  is  $\sigma^2\mathbf{I}_N$  and that of  $\hat{\mathbf{b}}$  is  $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ . First, using  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$  and  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ ,

$$\text{var}(\mathbf{y}) = E[\mathbf{y} - E(\mathbf{y})][\mathbf{y} - E(\mathbf{y})]' = E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{I}_N;$$

and second

$$\begin{aligned} \text{var}(\hat{\mathbf{b}}) &= \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \text{by (5) of Section 13.1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_N\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2. \end{aligned} \quad (40)$$

A similar result holds for  $\hat{\beta}$ : from (31)

$$\text{var}(\hat{\beta}) = \text{var}[(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'\mathbf{y}] = (\mathcal{X}'\mathcal{X})^{-1}\sigma^2, \quad (41)$$

as can also be verified directly from (40), using  $(\mathbf{X}'\mathbf{X})^{-1}$  implicit in (25).

derived from Exercise 13.3. That same inverse also yields

$$v(\hat{b}_0) = \sigma^2/N + \bar{x}'\text{var}(\hat{\beta})\bar{x} = [1/N + \bar{x}'(\mathcal{X}'\mathcal{X})^{-1}\bar{x}]\sigma^2 \quad (42)$$

and

$$\text{cov}(\hat{b}_0, \hat{\beta}') = -\bar{x}'\text{var}(\hat{\beta}). \quad (43)$$

#### 7. PREDICTED $y$ -VALUES

The estimator  $\hat{\mathbf{b}}$  can be used for estimating the expected value  $E(y) = a + b_1x_1 + \cdots + b_kx_k$  in the form

$$\widehat{E(y)} = \hat{b}_0x_0 + \hat{b}_1x_1 + \hat{b}_2x_2 + \cdots + \hat{b}_kx_k$$

with  $x_0 = 1$ . This can be done for any set of  $x$ -values

$$\mathbf{x}'_0 = [1 \quad x_{01} \quad \cdots \quad x_{0k}] \quad (44)$$

for which we wish to estimate the corresponding value of  $E(y)$ , in the form

$$\widehat{E(y_0)} = \hat{b}_0 + \hat{b}_1x_{01} + \cdots + \hat{b}_kx_{0k} = \mathbf{x}'_0\hat{\mathbf{b}}. \quad (45)$$

This is called the *estimated expected value of  $y$*  corresponding to the values  $1, x_{01}, \dots, x_{0k}$ . When this set of  $x$ 's is one of those in the data,  $\mathbf{x}'_0$  is a row of  $\mathbf{X}$ , in which case (45) is an element of  $\mathbf{X}\hat{\mathbf{b}}$ . Corresponding to  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$  we therefore have, as  $N$  special cases of (45),

$$\hat{\mathbf{y}} = \widehat{E(\mathbf{y})} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \quad \text{for} \quad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (46)$$

where  $\mathbf{H}$  is symmetric and idempotent.  $\mathbf{H}$  is sometimes called the "hat" matrix because it puts a "hat" over  $\mathbf{y}$ . The values in  $\hat{\mathbf{y}}$  correspond to those in the data. They are sometimes called *fitted  $y$ -values*, or *estimated  $y$ -values*, names which can be misleading because (45) and (46) are both estimates of expected values of  $y$ . They correspond, in (45), to any set of predetermined  $x$ 's in  $\mathbf{x}'_0$  of (44), and in (46) to the observed  $x$ 's in  $\mathbf{X}$ .

Variances of the estimators (45) and (46) are readily obtained using  $\text{var}(\hat{\mathbf{b}})$  of (40). Thus

$$v[\widehat{E(y_0)}] = \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\sigma^2$$

and

$$\text{var}[\widehat{E(\mathbf{y})}] = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2 = [\mathbf{J}_N + \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}']\sigma^2. \quad (47)$$



Equation (45) can also be used to predict a future observation. Implications of this are discussed at the end of Section 3.4c of Searle (1971).

### 8. ESTIMATING THE ERROR VARIANCE

The sum of squares of the deviations of the observed  $y$ 's from their predicted values is known as the error sum of squares:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (48)$$

It can be expressed as

$$SSE = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}),$$

and on substituting  $\hat{\mathbf{y}}$  from (46) it becomes

$$SSE = \mathbf{y}'(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \quad (49)$$

$$= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{H}\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$= \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} \quad (50)$$

because  $\hat{\mathbf{b}}' = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ . This is sometimes convenient for computing SSE:  $\mathbf{y}'\mathbf{y}$  in (50) is the total sum of squares of the observations, and  $\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}$  is the sum of products of the elements of the solution  $\hat{\mathbf{b}}$  with their corresponding elements of the right-hand side,  $\mathbf{X}'\mathbf{y}$ , of the equations from which  $\hat{\mathbf{b}}$  is derived, namely  $\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}$ .

An expression for SSE based on corrected sums of squares and products can also be developed. From (50)

$$\begin{aligned} SSE &= \mathbf{y}'\mathbf{y} - \begin{bmatrix} \bar{y} & \hat{\ell}'\bar{\mathbf{x}} & \hat{\ell}' \end{bmatrix} \begin{bmatrix} N\bar{y} \\ \mathbf{X}'_1\mathbf{y} \end{bmatrix} \\ &= \mathbf{y}'\mathbf{y} - N\bar{y}^2 - \hat{\ell}'(\mathbf{X}'_1\mathbf{y} - N\bar{y}\bar{\mathbf{x}}) = \mathbf{y}'\mathbf{y} - \hat{\ell}'\mathcal{X}'\mathbf{y}. \end{aligned} \quad (51)$$

Clearly, (51) is analogous to (50):  $\mathbf{y}'\mathbf{y}$  is the corrected sum of squares of the  $y$ 's, and  $\hat{\ell}'\mathcal{X}'\mathbf{y}$  is the sum of products of elements of the solution with the corresponding elements of the right-hand side,  $\mathcal{X}'\mathbf{y}$ , of the equations from which  $\hat{\ell}$  is derived, namely  $\mathcal{X}'\mathcal{X}\hat{\ell} = \mathcal{X}'\mathbf{y} = \mathcal{X}'\mathbf{y}$  of (31).

The expected value of SSE comes from applying (28) of Section 13.5 to (49) with  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$  and  $\mathbf{V} = \text{var}(\mathbf{y}) = \text{var}(\mathbf{e}) = \sigma^2\mathbf{I}$  of (39). Thus

$$\begin{aligned} E(SSE) &= \text{tr}(\mathbf{I} - \mathbf{H})\mathbf{I}\sigma^2 + \mathbf{b}'\mathbf{X}'(\mathbf{I} - \mathbf{H})\mathbf{X}\mathbf{b} \\ &= \sigma^2 \text{tr}(\mathbf{I} - \mathbf{H}), \quad \text{because } \mathbf{H}\mathbf{X} = \mathbf{X}, \\ &= (N - r_{\mathbf{X}})\sigma^2, \end{aligned} \quad (52)$$

on using the idempotency of  $\mathbf{H}$  which therefore has trace equal to rank (see Section 12.2c), which in this case is  $r_{\mathbf{X}} = k + 1$ .

An unbiased estimator of  $\sigma^2$  is, from (52),

$$\hat{\sigma}^2 = \frac{SSE}{N - r_{\mathbf{X}}} = \frac{SSE}{N - (k + 1)}. \quad (53)$$

In this regression situation of  $k$   $x$ -variables we know that  $r_{\mathbf{X}} = k + 1$ , but retaining  $r_{\mathbf{X}}$  in (53) is useful because it is a quite general result and makes for easier transition to the more general linear models of Chapter 15.

**Example (continued).** Using (15) and (17) in (46) gives

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} = \begin{bmatrix} 1 & 2 & 6 \\ 1 & 9 & 10 \\ 1 & 6 & 4 \\ 1 & 3 & 13 \\ 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} 37 \\ \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 47 \\ 56\frac{1}{2} \\ 46 \\ 58 \\ 42\frac{1}{2} \end{bmatrix}.$$

Therefore, with  $\mathbf{y}$  from (15)

$$(\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 62 \\ 60 \\ 57 \\ 48 \\ 23 \end{bmatrix} - \begin{bmatrix} 47 \\ 56\frac{1}{2} \\ 46 \\ 58 \\ 42\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 15 \\ 3\frac{1}{2} \\ 11 \\ -10 \\ -19\frac{1}{2} \end{bmatrix}$$

and so from (48)

$$SSE = 15^2 + (3\frac{1}{2})^2 + 11^2 + 10^2 + (19\frac{1}{2})^2 = 838\frac{1}{2}. \quad (54)$$

An alternative expression for SSE is (50). It requires  $y'y = \sum y_i^2 = 13526$  obtainable from (15). Then with  $\hat{\mathbf{b}}'$  from (17) and  $\mathbf{X}'\mathbf{y}$  from (16), equation (50) is

$$\text{SSE} = 13526 - \begin{bmatrix} 37 & \frac{1}{2} & 1\frac{1}{2} \end{bmatrix} \begin{bmatrix} 250 \\ 1265 \\ 1870 \end{bmatrix} = 13526 - 12687\frac{1}{2} = 838\frac{1}{2}, \text{ as before.}$$

Likewise for (51), using (35) for  $\mathcal{X}'\mathbf{y}$

$$\text{SSE} = 13526 - 5(50^2) - \begin{bmatrix} \frac{1}{2} & 1\frac{1}{2} \end{bmatrix} \begin{bmatrix} 15 \\ 120 \end{bmatrix} = 1026 - 7\frac{1}{2} - 180 = 838\frac{1}{2}, \text{ again.}$$

From all of these (53) gives

$$\hat{\sigma}^2 = \frac{838\frac{1}{2}}{5-3} = 419\frac{1}{4}. \quad (55)$$

#### 9. PARTITIONING THE TOTAL SUM OF SQUARES

The total sum of squares, which we call SST, is

$$\text{SST} = \mathbf{y}'\mathbf{y} = \sum_{i=1}^N y_i^2.$$

And the sum of squares of deviations of the observed  $y_i$ 's from their predicted values is, from (50) and (51)

$$\text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - N\bar{y}^2 - \hat{\ell}'\mathcal{X}'\mathbf{y}.$$

The difference is

$$\text{SSR} = \text{SST} - \text{SSE} = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = N\bar{y}^2 + \hat{\ell}'\mathcal{X}'\mathbf{y}.$$

It represents that portion of SST attributable to having fitted the regression and is called the *sum of squares due to regression*, SSR. It is also called the *reduction in sum of squares*. Clearly,  $\text{SST} = \text{SSR} + \text{SSE}$ , as is now summarized for developing the traditional analysis of variance table:

$$\begin{array}{l} \text{SSR} = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} (= \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}) = N\bar{y}^2 + \hat{\ell}'\mathcal{X}'\mathbf{y} \\ \text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - N\bar{y}^2 - \hat{\ell}'\mathcal{X}'\mathbf{y} \\ \hline \text{SST} = \mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{y} \end{array} \quad (56)$$

$N\bar{y}^2$  occurring in SSR of (56) is recognizable as the usual correction for the mean, denoted as SSM. Subtracting it from SSR gives what we call

$$\text{SSR}_m = \text{SSR} - \text{SSM} = \hat{\ell}'\mathcal{X}'\mathbf{y} = \hat{\ell}'\mathcal{X}'\mathcal{X}\hat{\ell},$$

the regression sum of squares corrected for the mean. In this way (56) becomes

$$\begin{array}{l} \text{SSM} = N\bar{y}^2 \\ \text{SSR}_m = \hat{\ell}'\mathcal{X}'\mathbf{y} = \hat{\ell}'\mathcal{X}'\mathcal{X}\hat{\ell} \\ \text{SSE} = \mathbf{y}'\mathbf{y} - N\bar{y}^2 - \hat{\ell}'\mathcal{X}'\mathbf{y} \\ \hline \text{SST} = \mathbf{y}'\mathbf{y} \end{array} \quad (57)$$

Similar to  $\text{SSR}_m$  we also have

$$\text{SST}_m = \text{SST} - \text{SSM} = \mathbf{y}'\mathbf{y} - N\bar{y}^2 = \mathbf{y}'\mathbf{y} \quad (58)$$

as the corrected sum of squares of the  $y$ 's. Then, with  $\text{SST}_m = \text{SSR}_m + \text{SSE}$ , we have the summary

$$\begin{array}{l} \text{SSR}_m = \hat{\ell}'\mathcal{X}'\mathbf{y} = \hat{\ell}'\mathcal{X}'\mathbf{y} \\ \text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\ell}'\mathcal{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - N\bar{y}^2 - \hat{\ell}'\mathcal{X}'\mathbf{y} \\ \hline \text{SST}_m = \mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{y} - N\bar{y}^2 \end{array} \quad (59)$$

This format is identical to that of (56); in the one case, (56), uncorrected sums of squares are used with total SST, and in the other, (59), corrected sums of squares are used with total  $\text{SST}_m$ . The error terms are the same in the two cases, namely SSE.

The summary shown in (59) is the basis of the traditional analysis of variance table for fitting linear regression, considered in Section 12.

**Example** (continued). From (54),  $\text{SSE} = 838\frac{1}{2}$ , and from the data themselves in (15),  $\text{SST} = \sum y_i^2 = 13526$  and  $\text{SSM} = N\bar{y}^2 = 5(50^2) = 12500$ . Therefore the summaries of (56), (57) and (59) are as shown in Table 14.1.

TABLE 14.1. PARTITIONING OF SUM OF SQUARES

Equations (56)	Equations (57)	Equations (59)
	SSM = 12500	
SSR = 12687 $\frac{1}{2}$	SSR <sub>m</sub> = 187 $\frac{1}{2}$	SSR <sub>m</sub> = 187 $\frac{1}{2}$
SSE = 838 $\frac{1}{2}$	SSE = 838 $\frac{1}{2}$	SSE = 838 $\frac{1}{2}$
SST = 13526	SST	

## 10. MULTIPLE CORRELATION

A measure of the goodness of fit of the regression is the multiple correlation coefficient, estimated as the product moment correlation between the observed  $y_i$ 's and the predicted  $\hat{y}_i$ 's. Denoted by  $R$ , it can be calculated as

$$R^2 = \frac{[\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}})^2}$$

Fortunately there is an easier formula, derived as follows.

First, in the denominator of  $R^2$

$$\sum(\hat{y}_i - \bar{\hat{y}})^2 = \mathbf{y}'\mathbf{C}\mathbf{y} = \mathbf{y}'(\mathbf{I} - \bar{\mathbf{J}})\mathbf{y} = \text{SST}_m,$$

where  $\mathbf{C}$  is a centering matrix. Second, the numerator of  $R^2$  is the square of

$$\begin{aligned} (\mathbf{C}\mathbf{y})'\mathbf{C}\hat{\mathbf{y}} &= \mathbf{y}'\mathbf{C}\mathbf{H}\mathbf{y} = \mathbf{y}'(\mathbf{I} - \bar{\mathbf{J}})\mathbf{H}\mathbf{y} \\ &= \mathbf{y}'(\mathbf{H} - \bar{\mathbf{J}})\mathbf{y} \\ &= \mathbf{y}'[(\mathbf{I} - \bar{\mathbf{J}}) - (\mathbf{I} - \mathbf{H})]\mathbf{y} \\ &= \text{SST}_m - \text{SSE}, \quad \text{using (49)} \\ &= \text{SSR}_m. \end{aligned} \quad (60)$$

(60) is true because, from (23),  $\mathbf{1}$  is a column of  $\mathbf{X}$ , and so part (viii) of Exercise 8.37 gives  $\mathbf{1}'\mathbf{H} = \mathbf{1}'$  and hence  $\bar{\mathbf{J}}\mathbf{H} = \bar{\mathbf{J}} = \mathbf{H}\bar{\mathbf{J}}$ . Finally, in the denominator of  $R^2$ ,

$$\sum(\hat{y}_i - \bar{\hat{y}})^2 = \hat{\mathbf{y}}'\mathbf{C}\hat{\mathbf{y}} = \hat{\mathbf{y}}'\mathbf{H}\mathbf{C}\mathbf{H}\mathbf{y} = \hat{\mathbf{y}}'(\mathbf{H} - \bar{\mathbf{J}})\mathbf{y} = \text{SSR}_m,$$

and so

$$R^2 = (\text{SSR}_m)^2 / \text{SST}_m(\text{SSR}_m) = \text{SSR}_m / \text{SST}_m. \quad (61)$$

In the example,

$$R^2 = 187\frac{1}{2} / 1026 = .183.$$

Intuitively the ratio  $\text{SSR}_m / \text{SST}_m$  of (61) has appeal, since it represents that fraction of the total (corrected) sum of squares which is accounted for

by fitting the model—in this case fitting the regression. Thus although  $R$  has traditionally been thought of and used as a multiple correlation coefficient in some sense, its more frequent use nowadays is in the form (61), where it represents the fraction of the total sum of squares accounted for by fitting the model.

Care must be taken in using (61) for, although  $\text{SSR}_m = \text{SSR} - N\bar{y}^2$  and  $\text{SST}_m = \text{SST} - N\bar{y}^2$  in the intercept model, the ratio is  $\text{SSR}/\text{SST}$  in the no-intercept model, as we now see.

## 11. THE NO-INTERCEPT MODEL

When all  $x$ 's are zero in the above models,  $\widehat{E}(y) = \hat{b}_0$ . Thus for  $x_1 = 0 = x_2$  in the preceding example the estimated value  $\widehat{E}(y)$  is  $\hat{b}_0 = 37$ . Models of this nature are called *intercept models*; the intercept is  $b_0$ , the value of  $E(y)$  when all  $x$ 's are zero.

Sometimes it is appropriate to have no term  $b_0$  in the model, in which case the model is called a *no-intercept model*. The matrix  $\mathbf{X}$  then has no vector  $\mathbf{1}$  in it, [it is  $\mathbf{X}_1$  of (23)], and  $\mathbf{X}'\mathbf{X}$  is then the matrix of sums of squares and products of the observations, without the first row and column of totals seen in (22).

**Example** (continued). For the no-intercept model, from (15) and (16)

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 155 & 175 \\ 175 & 325 \end{bmatrix} \quad \text{and} \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 1265 \\ 1870 \end{bmatrix}$$

for which solution to (20) is

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{790} \begin{bmatrix} 3355 \\ 2739 \end{bmatrix} = \begin{bmatrix} 4.25 \\ 3.47 \end{bmatrix}. \quad (62)$$

The no-intercept model thus has  $E(y)$  being estimated from these data as

$$\widehat{E}(y) = 4.25x_1 + 3.47x_2.$$

The partitioning of  $\mathbf{b}'$  as  $[b_0 \ \ell']$  does not exist in the no-intercept model.  $\mathbf{b}'$  is itself the vector of the  $b$ 's corresponding to the  $x$ -variables and  $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  is based on uncorrected sums of squares and products as exemplified in (62). Using corrected sums of squares and products  $\mathcal{X}'\mathcal{X}$  and  $\mathcal{X}'\mathbf{y}$  is therefore not appropriate for the no-intercept model. Similarly, the only appropriate partitioning of  $\text{SST}$  is that of (56).

For the example, (62) yields

$$SSE = \hat{\mathbf{b}}' \mathbf{X}' \mathbf{y} = \frac{1}{790} [3355 \quad 2739] \begin{bmatrix} 1265 \\ 1870 \end{bmatrix} = 11855.7025$$

so that the partitioning of SST is

$$\begin{array}{r} SSR = 11855.7025 \\ SSE = 1670.2975 \\ \hline SST = 13526 \end{array}$$

Analogous to (60), the multiple correlation  $R$  in the no-intercept model is defined without reference to  $\bar{y}$  and so is

$$R^2 = (\sum y_i \hat{y}_i)^2 / [(\sum y_i^2)(\sum \hat{y}_i^2)] = SSR/SST.$$

Its value for the example is therefore  $11855.7025/13526 = .877$ .

## 12. ANALYSIS OF VARIANCE

Partitioning the total sum of squares as shown in (56), (57) and (59) is the basis for analysis of variance. But using the customary  $F$ -statistics emanating from analysis of variance tables demands normality assumptions, and up to this point no need has arisen for, nor has any assumption been made about, probability distributional properties of the observations  $y_i$  or more particularly of the residuals  $e_i$ . We now assume that they are normally distributed. This leads to sums of squares in the partitioning of SST having  $\chi^2$ -distributions, and thence to ratios of mean squares having  $F$ -distributions from which come familiar tests of hypotheses based on  $F$ -statistics. Derivation of these results is now sketched. The reader wanting details is referred to the texts mentioned at the beginning of this chapter. The outline given here is based on Section 3.5 of Searle (1971).

The basic distributional assumption to be made is that the error terms in  $\mathbf{e}$  of the model  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$  are normally distributed with zero means and variance-covariance matrix  $\sigma^2 \mathbf{I}$ ; i.e., using the notation following (25) of Section 13.4, we write

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Distribution properties of  $\mathbf{y}$  and of functions of  $\mathbf{y}$  follow at once:

- (i)  $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I})$ .
- (ii)  $\hat{\mathbf{b}} \sim N[\mathbf{b}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2]$ .

- (iii)  $\hat{\boldsymbol{\theta}} \sim N[\boldsymbol{\theta}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2]$ .
- (iv)  $\hat{\mathbf{b}}$  and  $\hat{\sigma}^2$  are independent.
- (v)  $SSE/\sigma^2$  has a  $\chi^2$ -distribution on  $N - r_x$  degrees of freedom.
- (vi)  $SSR/\sigma^2$ ,  $SSM/\sigma^2$  and  $SSR_m/\sigma^2$  each has a (noncentral)  $\chi^2$ -distribution on  $k + 1$ , 1, and  $k$  degrees of freedom, respectively, and each is independent of SSE.

Application of (v) and (vi) to the sums of squares in (57) yields the analysis of variance of Table 14.2.

Hypotheses tested by the  $F$ -statistics in Table 14.2 are as follows:

$$F(M) \text{ tests } H: \mathbf{1}'\mathbf{X}\mathbf{b} = 0, \text{ equivalent to } H: E(\bar{y}) = 0,$$

$$F(R_m) \text{ tests } H: \boldsymbol{\theta} = \mathbf{0}$$

and

(63)

$$F(R) = \frac{SSR}{(k+1)MSE} \text{ tests } H: \mathbf{b} = \mathbf{0},$$

with  $SSR = SSR_m + SSM$ . Note that  $F(M)$  is the square of a  $t$ -statistic:

$$F(M) = \frac{MSM}{MSE} = \frac{N\bar{y}^2}{\hat{\sigma}^2} = \left( \frac{\bar{y}}{\hat{\sigma}/\sqrt{N}} \right)^2$$

TABLE 14.2. ANALYSIS OF VARIANCE FOR REGRESSION ON  $k$   $x$ -VARIABLES

Source of Variation <sup>a</sup>	d.f. <sup>b</sup>	Sum of Squares	Mean Square	$F$ -statistic
Mean	1	$SSM = N\bar{y}^2$	$MSM = \frac{SSM}{1}$	$F(M) = \frac{MSM}{MSE}$
Regression (c.f.m.)	$k$	$SSR_m = \hat{\boldsymbol{\theta}}'\mathbf{X}'\mathbf{y}$	$MSR_m = \frac{SSR_m}{k}$	$F(R_m) = \frac{MSR_m}{MSE}$
Residual error	$N - r_x$	$SSE = \mathbf{y}'\mathbf{y} - N\bar{y}^2 - \hat{\boldsymbol{\theta}}'\mathbf{X}'\mathbf{y}$	$MSE = \frac{SSE}{N - r_x} = \hat{\sigma}^2$	
Total	$N$	$SST = \mathbf{y}'\mathbf{y}$		

<sup>a</sup>c.f.m. = corrected for the mean.

<sup>b</sup>d.f. = degrees of freedom:  $k + 1 = r_x$  for  $k$   $x$ -variables.

TABLE 14.3. ANALYSIS OF VARIANCE FOR EXAMPLE

Source of Variation	d.f.	Sum of Squares	Mean Square	F-statistic
Mean	1	SSM = 12500	MSM = 12500	$F(M) = 12500/419\frac{1}{2} = 29.8$
Regression (c.f.m.)	2	SSR <sub>m</sub> = 187 $\frac{1}{2}$	MSR <sub>m</sub> = 93 $\frac{1}{4}$	$F(R_m) = 93\frac{1}{4}/419\frac{1}{2} = 0.2$
Residual error	2	SSE = 838 $\frac{1}{2}$	MSE = 419 $\frac{1}{2}$	
Total	5	SST = 13526		

where  $\bar{y}/(\hat{\sigma}/\sqrt{N})$  is a  $t$ -statistic with  $N - k - 1$  degrees of freedom. Also,  $F(R_m)$  can be calculated using the multiple correlation squared. Using (61)

$$F(R_m) = \frac{MSR_m}{MSE} = \frac{SSR_m}{k} \frac{(N - k - 1)}{SSE} = \frac{N - k - 1}{k} \frac{SSR_m}{SST_m - SSR_m}$$

$$= \frac{N - k - 1}{k} \frac{R^2}{1 - R^2} = \frac{(N - 1)/k - 1}{1/R^2 - 1}. \quad (64)$$

**Example (continued).** Using the middle column of Table 14.1, the analysis of variance Table 14.2 is shown for the example in Table 14.3. Then  $F(M)$  tested against the tabulated  $F$ -distribution on 1 and 2 degrees of freedom tests  $H: \mathbf{1}'\mathbf{X}\mathbf{b} = 0$ , or equivalently  $H: E(\bar{y}) = 0$ , in this case  $H: a + 5b_1 + 7b_2 = 0$ . And  $F(R_m) = 0.2$  tested against the tabulated  $F$ -distribution on 2 and 2 degrees of freedom tests  $H: b_1 = 0 = b_2$ .

Demonstration of (64) is also readily available. From the end of Section 10,  $R^2 = 187\frac{1}{2}/1026$ , and so (64) gives the same value as Table 14.3:

$$F(R_m) = (4/2 - 1)/(1026/187\frac{1}{2} - 1) = 187\frac{1}{2}/838\frac{1}{2} = 0.2.$$

### 13. TESTING LINEAR HYPOTHESES

#### a. Stating a hypothesis

Suppose in the example that we wanted to test the joint hypothesis

$$H: \begin{cases} b_1 = \frac{1}{2} \\ b_1 + 3b_2 = 3 \end{cases}, \text{ equivalent to } H: \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix}. \quad (65)$$

Its second form is an example of

$$H: \mathbf{K}'\mathbf{b} = \mathbf{m} \quad (66)$$

where  $\mathbf{b}$  is the vector of parameters in the model  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ , and where  $\mathbf{K}'$  and  $\mathbf{m}$  are a matrix and vector, respectively, of known values that define the hypothesis of interest. In (65)

$$\mathbf{K}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \text{ and } \mathbf{m} = \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix}.$$

In this way (66) is a quite general formulation of any linear hypothesis: every linear hypothesis can be expressed this way. For example,  $H: \mathbf{b} = \mathbf{0}$  and  $H: \mathbf{t} = \mathbf{0}$  can be expressed, respectively, as

$$H: \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } H: \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (67)$$

Only one limitation is placed on the form  $H: \mathbf{K}'\mathbf{b} = \mathbf{m}$ , and it is a very practical one.  $\mathbf{K}'\mathbf{b}$  must not contain any redundancies. For example, in

$$H: \begin{cases} b_1 = \frac{1}{2} \\ b_1 + 3b_2 = 3 \\ 2b_1 + 3b_2 = 3\frac{1}{2} \end{cases}, \text{ i.e., } H: \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \\ 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 3 \\ 3\frac{1}{2} \end{bmatrix}, \quad (68)$$

the last equation is clearly the sum of the other two. If the statements  $b_1 = \frac{1}{2}$  and  $b_1 + 3b_2 = 3$  are part of the hypothesis there is no need to also have  $2b_1 + 3b_2 = 3\frac{1}{2}$ , because it is the sum of  $b_1 = \frac{1}{2}$  and  $b_1 + 3b_2 = 3$ . In this sense  $2b_1 + 3b_2 = 3\frac{1}{2}$  is redundant and its presence along with  $b_1 = \frac{1}{2}$  and  $b_1 + 3b_2 = 3$  is not permitted in the hypothesis. In general, functions of  $\mathbf{b}$  are not permitted in  $\mathbf{K}'\mathbf{b}$  if they can be expressed as linear combinations of other such functions in  $\mathbf{K}'\mathbf{b}$ . This is tantamount to not permitting any rows of  $\mathbf{K}'$  to be linearly dependent on other rows, as occurs in (68). Hence in stating any linear hypothesis as  $H: \mathbf{K}'\mathbf{b} = \mathbf{m}$ , the matrix  $\mathbf{K}'$  must always be of full row rank. Not only is this limitation practical, as just explained, but it also materially affects development of the general expression for the  $F$ -statistic for testing  $H: \mathbf{K}'\mathbf{b} = \mathbf{m}$  quoted in the next subsection.

#### b. The $F$ -statistic

In the model  $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2\mathbf{I}_N)$  that is being considered in this chapter, with  $\mathbf{X}$  of full column rank, it can be shown (e.g., Searle, 1971, Section 3.6a)

that the  $F$ -statistic for testing  $H: \mathbf{K}\mathbf{b} = \mathbf{m}$  is

$$F(H) = Q/\hat{\sigma}^2 r_{\mathbf{K}}, \quad \text{with } r_{\mathbf{K}} \text{ and } N - k - 1 \text{ degrees of freedom, (69)}$$

$r_{\mathbf{K}}$  being the rank of  $\mathbf{K}$ , equal to the number of rows in  $\mathbf{K}'$ . The numerator sum of squares in  $F(H)$  of (69) is

$$Q = (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})' [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m}); \quad (70)$$

and

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{and} \quad \hat{\sigma}^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}) / (N - k - 1) \quad (71)$$

as have already been developed in (11), and (50) and (53).

Hypotheses concerned with the regression slopes of  $\boldsymbol{\beta}' = [b_1 \dots b_k]$  and not with the intercept  $b_0$  can be stated and tested in similar fashion. The  $F$ -statistic for testing  $H^*: \mathbf{K}^*\boldsymbol{\beta} = \mathbf{m}^*$  is

$$F(H^*) = Q^*/\hat{\sigma}^2 r_{\mathbf{K}^*} \quad \text{with } r_{\mathbf{K}^*} \text{ and } N - k - 1 \text{ degrees of freedom, (72)}$$

where

$$Q^* = (\mathbf{K}^{*\prime}\hat{\boldsymbol{\beta}} - \mathbf{m}^*)' [\mathbf{K}^{*\prime}(\mathcal{X}'\mathcal{X})^{-1}\mathbf{K}^*]^{-1} (\mathbf{K}^{*\prime}\hat{\boldsymbol{\beta}} - \mathbf{m}^*), \quad (73)$$

and with

$$\hat{\boldsymbol{\beta}} = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'\mathbf{y} \quad \text{and} \quad \hat{\sigma}^2 = (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathcal{X}'\mathbf{y}) / (N - k - 1), \quad (74)$$

from (31), and (51) and (53).

**Example (continued).** For the hypothesis of (65) we illustrate the calculation of  $Q$  and  $Q^*$  of (70) and (73). First, for  $Q$ , with  $\hat{\mathbf{b}}$  from (17)

$$\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 37 \\ \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

and with  $(\mathbf{X}'\mathbf{X})^{-1}$  from (16),

$$\begin{aligned} [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} &= \left\{ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \frac{1}{480} \begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 3 \end{bmatrix} \right\}^{-1} \\ &= 480 \begin{bmatrix} 16 & 16 \\ 16 & 70 \end{bmatrix}^{-1} = \frac{5}{9} \begin{bmatrix} 70 & -16 \\ -16 & 16 \end{bmatrix}. \end{aligned}$$

Therefore  $Q$  of (70) is

$$Q = [0 \quad 2] \left( \frac{5}{9} \right) \begin{bmatrix} 70 & -16 \\ -16 & 16 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = 35 \frac{5}{9}.$$

$$Q^* \text{ uses } \mathbf{K}^{*\prime}\hat{\mathbf{b}} - \mathbf{m}^* = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

and, with  $(\mathbf{X}'\mathbf{X})^{-1}$  from (34),

$$\begin{aligned} [\mathbf{K}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}^*]^{-1} &= \left\{ \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1/30 & 0 \\ 0 & 1/80 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \right\}^{-1} \\ &= 30 \begin{bmatrix} 1 & 1 \\ 1 & 4\frac{3}{8} \end{bmatrix}^{-1} = \frac{5}{9} \begin{bmatrix} 70 & -16 \\ -16 & 16 \end{bmatrix} \end{aligned}$$

so that (73) yields  $Q^* = Q = 35 \frac{5}{9}$ .

### c. Equivalent statements of a hypothesis

A feature of testing hypotheses is, of course, that even when a hypothesis can be expressed in different equivalent ways, for each of those ways the  $F$ -statistic given by (69) will be the same. For example, the hypothesis (65)

$$H: \begin{cases} b_1 = \frac{1}{2} \\ b_1 + 3b_2 = 3 \end{cases}, \quad \text{is equivalent to} \quad H: \begin{cases} b_1 = \frac{1}{2} \\ b_2 = \frac{5}{6} \end{cases}.$$

For the latter,

$$\mathbf{K}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix}, \quad \mathbf{K}'\hat{\mathbf{b}} - \mathbf{m} = \begin{bmatrix} \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{3} \end{bmatrix},$$

$$\begin{aligned} [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} &= \left\{ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \frac{1}{480} \begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}^{-1} \\ &= \begin{bmatrix} 30 & 0 \\ 0 & 80 \end{bmatrix} \end{aligned}$$

and (69) gives

$$Q = [0 \quad \frac{2}{3}] \begin{bmatrix} 30 & 0 \\ 0 & 80 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{2}{3} \end{bmatrix} = 35 \frac{5}{9}, \quad \text{as before.}$$

## d. Special cases

At least three special cases of the general linear hypothesis  $H: \mathbf{K}'\mathbf{b} = \mathbf{m}$  are of interest: (i) that all elements of  $\mathbf{b}$  are zero, i.e.,  $H: \mathbf{b} = \mathbf{0}$ ; (ii) that each element of  $\mathbf{b}$  equals some specific value, i.e.,  $H: \mathbf{b} = \mathbf{b}_0$ ; and (iii) that some linear combination of the elements of  $\mathbf{b}$  equals a specific value, e.g., for given  $\lambda$  and  $m$ ,  $H: \lambda'\mathbf{b} = m$ . For each of these the  $F$ -statistic can be derived using (69), as is now shown. A fourth case, of wide applicability, is dealt with in Section 14.

(i)  $H: \mathbf{b} = \mathbf{0}$ . The  $F$ -statistic for testing this hypothesis is shown in (63). It can also be derived from (69), for which the equations  $\mathbf{b} = \mathbf{0}$  have to be expressed as  $\mathbf{K}'\mathbf{b} = \mathbf{m}$ . This means  $\mathbf{K}' = \mathbf{I}_{k+1}$ ,  $r_{\mathbf{K}'} = k + 1$ ,  $\mathbf{m} = \mathbf{0}$ ,  $\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m} = \hat{\mathbf{b}}$ , and  $[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} = \mathbf{X}'\mathbf{X}$  and so

$$F(H) = \frac{\hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}}{\hat{\sigma}^2(k+1)} = \frac{\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}}{\text{MSE}(k+1)} = \frac{\text{SSR}}{(k+1)\text{MSE}}$$

as in (63). Degrees of freedom for  $F(H)$  are  $k + 1$  and  $N - k - 1$ .

(ii)  $H: \mathbf{b} = \mathbf{b}_0$ . This differs from (i) only in that  $\mathbf{m}$  of  $H: \mathbf{K}'\mathbf{b} = \mathbf{m}$  is  $\mathbf{b}_0$ . Thus

$$F(H) = (\hat{\mathbf{b}} - \mathbf{b}_0)'\mathbf{X}'\mathbf{X}(\hat{\mathbf{b}} - \mathbf{b}_0) / [(k+1)\hat{\sigma}^2],$$

with  $k + 1$  and  $N - k - 1$  degrees of freedom.

(iii)  $H: \lambda'\mathbf{b} = m$ . Here,  $\mathbf{K}' = \lambda'$  with  $r_{\mathbf{K}'} = 1$ ,  $\mathbf{m} = m$  and

$$F(H) = \frac{(\lambda'\hat{\mathbf{b}} - m)'\lambda'(\mathbf{X}'\mathbf{X})^{-1}\lambda}{\hat{\sigma}^2} (\lambda'\hat{\mathbf{b}} - m)$$

Because  $\lambda'$  is a vector, this can be rewritten as

$$F(H) = \frac{(\lambda'\hat{\mathbf{b}} - m)^2}{\lambda'(\mathbf{X}'\mathbf{X})^{-1}\lambda\hat{\sigma}^2};$$

it has 1 and  $N - k - 1$  degrees of freedom and, just like  $F(M)$  of Section 12, it is the square of a  $t$ -statistic having  $N - k - 1$  degrees of freedom.

## 14. CONFIDENCE INTERVALS

On the basis of assuming  $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2\mathbf{I})$ , we have from Section 6,  $\hat{b}_i \sim N[\hat{b}_i, (\mathcal{X}'\mathcal{X})^{-1}\sigma^2]$  and so  $(\hat{b}_i - b_i) / \sqrt{a^{ii}\sigma^2} \sim N(0, 1)$ , where  $a^{ii}$  is the

$i$ th diagonal element of  $(\mathcal{X}'\mathcal{X})^{-1}$ . Therefore

$$(\hat{b}_i - b_i) / \sqrt{a^{ii}\hat{\sigma}^2} \sim t_{N-r_x}$$

where  $t_{N-r}$  represents the  $t$ -distribution on  $N - r$  degrees of freedom, with  $r = r_{\mathbf{X}}$ . On defining  $t_\alpha$  and  $z_\alpha$  by

$$P\{t_{N-r} > t_\alpha\} = \frac{1}{2}\alpha \quad \text{and} \quad (2\pi)^{-\frac{1}{2}} \int_{z_\alpha}^{\infty} e^{-\frac{1}{2}u^2} du = \frac{1}{2}\alpha,$$

the  $100(1 - \alpha)\%$  symmetric confidence interval on  $b_i$  is

$$\hat{b}_i \pm \hat{\sigma} t_\alpha \sqrt{a^{ii}} \quad \text{or} \quad \hat{b}_i \pm \hat{\sigma} z_\alpha \sqrt{a^{ii}},$$

the former when degrees of freedom are small ( $N - r < 100$ , say) and the latter otherwise. Derivation of these intervals and of nonsymmetric intervals is given in Searle (1971, p. 107). In the case of  $b_0$ , these intervals are used with

$$a^{00} = 1/N + \bar{\mathbf{x}}'(\mathcal{X}'\mathcal{X})^{-1}\bar{\mathbf{x}},$$

in accord with (43).

15. FITTING SUBSETS OF THE  $x$ -VARIABLES

Consider a regression analysis of plant growth on five  $x$ -variables:  $x_1$  and  $x_2$  are climatic variables (e.g., mean daily temperature, and rainfall) and  $x_3$ ,  $x_4$  and  $x_5$  are seedling measurements at time of transplanting (e.g., age, leaf development and height of transplant). We might wish to ascertain whether or not the two climatic variables make a significant contribution to the regression. If they do not, the regression would be based solely on the seedling measurements  $x_3$ ,  $x_4$  and  $x_5$ , and for just these three variables a regression could be estimated. In other words, we are interested in testing the hypothesis  $H: b_1 = b_2 = 0$ .

Suppose that instead of writing the regression as

$$E(y) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$$

we change the order of the terms to have

$$E(y) = b_1x_1 + b_2x_2 + b_0 + b_3x_3 + b_4x_4 + b_5x_5$$

and with this order partition  $\mathbf{Xb}$  of  $E(\mathbf{y}) = \mathbf{Xb}$  to be

$$E(\mathbf{y}) = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 = [\mathbf{X}_1 \quad \mathbf{X}_2] \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \quad (75)$$

with

$$\mathbf{b}'_1 = [b_1 \quad b_2] \quad \text{and} \quad \mathbf{b}'_2 = [b_3 \quad b_4 \quad b_5].$$

Then the hypothesis  $H: b_1 = b_2 = 0$  is

$$H: \mathbf{b}_1 = \mathbf{0}. \quad (76)$$

Hypotheses of this nature are frequently of interest: testing whether some subset of the  $k$   $x$ -variables is contributing to the regression.

We consider the partitioned model (75) and hypothesis (76) in the general case where  $\mathbf{b}_1$  is taken as having  $k_1$  elements and  $\mathbf{b}_2$  having  $k_2$ , with  $k_1 + k_2 = k + 1$ . Then for testing  $H: \mathbf{b}_1 = \mathbf{0}$  with (69) we have  $\mathbf{K}' = [\mathbf{I} \quad \mathbf{0}]$ ,  $r_{\mathbf{K}'} = k_1$  and  $\mathbf{m} = \mathbf{0}$ , so that  $\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m} = \hat{\mathbf{b}}_1$  and, on defining

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{X}'_1\mathbf{X}_2 \\ \mathbf{X}'_2\mathbf{X}_1 & \mathbf{X}'_2\mathbf{X}_2 \end{bmatrix}^{-1} \equiv \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \quad \text{with} \quad \mathbf{T}_{21} = (\mathbf{T}_{12})', \quad (77)$$

$$[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} = \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right\}^{-1} = \mathbf{T}_{11}^{-1}. \quad (78)$$

Hence the  $F$ -statistic for testing  $H: \mathbf{b}_1 = \mathbf{0}$  is

$$F(\mathbf{b}_1) = Q_1 / (k_1\hat{\sigma}^2) \quad \text{with} \quad Q_1 = \hat{\mathbf{b}}_1'\mathbf{T}_{11}^{-1}\hat{\mathbf{b}}_1. \quad (79)$$

Since  $Q_1$  involves  $\mathbf{T}_{11}^{-1}$  where, by (77),  $\mathbf{T}_{11}$  is part of the inverse  $(\mathbf{X}'\mathbf{X})^{-1}$ , the expression for  $Q_1$  in (79) has been referred to earlier by Searle (1966, 1971) as the "invert part of the inverse" procedure.  $Q_1$  is, in fact, a quadratic form in the solution vector  $\hat{\mathbf{b}}_1$  (corresponding to the vector of the hypothesis  $H: \mathbf{b}_1 = \mathbf{0}$ ) using the inverse of the submatrix in  $(\mathbf{X}'\mathbf{X})^{-1}$  corresponding to that same vector.  $Q_1$  of (79) also has been called an "indirect" method of calculating the numerator sum of squares for testing  $H: b_1 = 0$ . (See Searle, Speed and Henderson, 1981, which also contains an extension to the case when  $\mathbf{X}'\mathbf{X}$  is singular.)

A special case of (79) of particular interest is when  $\mathbf{b}_1$  is just a single element of  $\mathbf{b}$ ,  $b_i$ , say. Then (79) gives  $Q_i = \hat{b}_i\mathbf{T}_{ii}^{-1}\hat{b}_i = \hat{b}_i^2/\mathbf{T}_{ii}$  for  $\mathbf{T}_{ii}$  being the  $(i, i)$ th element of  $(\mathbf{X}'\mathbf{X})^{-1}$ . More particularly, for  $i = 1, 2, \dots, k$  (but not for

$i = 0$ ),  $\mathbf{T}_{ii}$  is the  $i$ th diagonal element of  $(\mathcal{X}'\mathcal{X})^{-1}$ ,  $a^{ii}$ , and so in (79)

$$F(b_i) = \frac{\hat{b}_i^2}{a^{ii}\hat{\sigma}^2}. \quad (80)$$

This has 1 and  $N - k - 1$  degrees of freedom; and so, similar both to  $F(H)$  in (iii) of the preceding section and to  $F(M)$  of Section 12,  $F(b_i)$  is the square of a  $t$ -statistic:

$$F(b_i) = \left[ \frac{\hat{b}_i}{\hat{\sigma}\sqrt{a^{ii}}} \right]^2 = \frac{\hat{b}_i^2}{\text{estimate of } v(\hat{b}_i)}. \quad (81)$$

The second equality in (81) is based on  $\text{var}(\hat{b}) = (\mathcal{X}'\mathcal{X})^{-1}\sigma^2$  of (41).

**Example** (continued). The  $F$ -statistic for testing  $H: b_1 = 0$  uses  $\hat{b}_1 = \frac{1}{2}$  from (35), it uses  $a^{11} = 1/30$  from (34) and  $\hat{\sigma}^2 = 419\frac{1}{4}$  from (55). Substituted in (80) these give

$$F(b_1) = (\frac{1}{2})^2 / [(1/30)419\frac{1}{4}] = 7\frac{1}{2} / 419\frac{1}{4}. \quad (82)$$

## 16. REDUCTIONS IN SUMS OF SQUARES: THE $R(\cdot | \cdot)$ NOTATION

The numerator sum of squares  $Q_1$  of (79) for testing the hypothesis  $H: \mathbf{b}_1 = \mathbf{0}$  can also be viewed as the difference between the reductions in sums of squares for fitting two different models. This we now show. First, from (56), the reduction in sum of squares for fitting the model

$$E(\mathbf{y}) = \mathbf{Xb} \quad \text{is} \quad \text{SSR} = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (83)$$

In using this for a variety of different models, as will be done, i.e., for models with different  $\mathbf{b}$ 's and  $\mathbf{X}$ 's, it is convenient to introduce the notation

$$R(\mathbf{b}) = \text{reduction in sum of squares due to fitting } E(\mathbf{y}) = \mathbf{Xb} \quad (84)$$

$$= \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (85)$$

Thus the reduction in sum of squares for fitting  $E(\mathbf{y}) = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2$  is

$$R(\mathbf{b}_1, \mathbf{b}_2) = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{with} \quad \mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]. \quad (86)$$



And the reduction for fitting just the  $\mathbf{X}_2\mathbf{b}_2$  part, i.e., for fitting

$$E(y) = \mathbf{X}_2\mathbf{b}_2 \quad \text{is} \quad R(\mathbf{b}_2) = \mathbf{y}'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}. \quad (87)$$

Now consider the difference between the reductions in sums of squares for these two models. It is  $R(\mathbf{b}_1, \mathbf{b}_2) - R(\mathbf{b}_2)$ . It represents the greater reduction in sum of squares due to fitting  $E(y) = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2$  over and above fitting  $E(y) = \mathbf{X}_2\mathbf{b}_2$ . It therefore indicates the extent to which having  $\mathbf{X}_1\mathbf{b}_1$  in the model, over and above having  $\mathbf{X}_2\mathbf{b}_2$ , contributes to the reduction in sum of squares; i.e., in terms of sums of squares it represents the contribution of  $\mathbf{X}_1\mathbf{b}_1$  to the model in the presence of  $\mathbf{X}_2\mathbf{b}_2$ . We denote this by  $R(\mathbf{b}_1 | \mathbf{b}_2)$  and indicate why it equals  $Q_1$  of (79).

First, by definition

$$R(\mathbf{b}_1 | \mathbf{b}_2) = R(\mathbf{b}_1, \mathbf{b}_2) - R(\mathbf{b}_2) \quad (88)$$

and from (86) and (87) this is

$$R(\mathbf{b}_1 | \mathbf{b}_2) = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}. \quad (89)$$

On using  $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$  of (86), and the second expression in Exercise 10.7 for the resulting  $(\mathbf{X}'\mathbf{X})^{-1}$ , we find that (89) reduces to

$$R(\mathbf{b}_1 | \mathbf{b}_2) = \mathbf{y}'\mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \quad (90)$$

with

$$\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2'(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2 = \mathbf{M}_2' = \mathbf{M}_2^2. \quad (91)$$

Furthermore, using that same  $(\mathbf{X}'\mathbf{X})^{-1}$  we find

$$(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1} = \mathbf{T}_{11} \text{ of (77)}. \quad (92)$$

Moreover, the normal equations for the model  $E(y) = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2$ ,

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{bmatrix}, \quad \text{yield} \quad \hat{\mathbf{b}}_1 = (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}. \quad (93)$$

Hence in (90)

$$R(\mathbf{b}_1 | \mathbf{b}_2) = \hat{\mathbf{b}}_1'\mathbf{T}_{11}^{-1}\hat{\mathbf{b}}_1 = Q_1 \text{ of (79)}. \quad (94)$$

**Example** (continued). In Table 14.3

$$R(b_0, b_1, b_2) = \text{SSR} = \text{SSM} + \text{SSR}_m = 12500 + 187\frac{1}{2} = 12687\frac{1}{2}.$$

And for  $E(y) = \mathbf{X}_2\mathbf{b}_2 = b_0 + b_2x_2$ , the reduction in sum of squares is, on using elements of (15) and (16) in (87),

$$R(b_0, b_2) = [250 \quad 1870] \begin{bmatrix} 5 & 35 \\ 35 & 325 \end{bmatrix}^{-1} \begin{bmatrix} 250 \\ 1870 \end{bmatrix} = 12680.$$

Therefore

$$R(b_1 | b_0, b_2) = R(b_0, b_1, b_2) - R(b_0, b_2) = 12687\frac{1}{2} - 12680 = 7\frac{1}{2},$$

which is the numerator of  $F(b_1)$  in (82); i.e.,

$$R(b_1 | b_0, b_2) = Q_1, \quad \text{for testing } H: b_1 = 0,$$

thus illustrating (94).

As final comment, it can be stated that most of the procedures described in terms of  $\mathbf{b}$  and  $\mathbf{X}'\mathbf{X}$  in this and the preceding two sections can also be developed in terms of  $\ell$  and  $\mathcal{X}'\mathcal{X}$ .

## CHAPTER 15

# AN INTRODUCTION TO LINEAR STATISTICAL MODELS

The general subject of linear statistical models is a large one and is presented extensively in many texts such as Kempthorne (1952), Scheffé (1959), Graybill (1961), Searle (1971), Winer (1971), Rao (1973), Graybill (1976) and Seber (1977), to name but a few. It is a subject of great importance in data analysis, because of its close association with analysis of variance. It is also very amenable to presentation in matrix terminology which, in turn, brings considerable clarity to the subject. Nevertheless, because the topic is so vast just its fringe is touched here, in a manner even less complete than that on regression in the preceding chapter. Great reliance is placed on results given there, and considerable use is made of the generalized inverse of a matrix discussed in Chapter 8 and its uses in solving linear equations given in Chapter 9.

To those familiar with linear models it can be said that the main objective of this chapter is to present a simple and general method of deriving estimable functions and their unique, least squares (minimum variance, linear, unbiased) estimators. The method is based entirely on the properties of generalized inverses and is applicable to any situation whatever, whether the data are balanced or unbalanced.

### 1. GENERAL DESCRIPTION

A simple illustration serves to introduce the general idea of linear models and the matrix notation applicable to them.

**Example.** Federer (1955) reports an analysis of rubber-producing plants called guayule, for which the plant weights were available for 54 plants of

three different kinds, 27 of them normal, 15 off-types and 12 aberrants. We will consider just 6 plants for purposes of illustration, 3 normals, 2 off-types and 1 aberrant, as shown in Table 15.1.

TABLE 15.1. WEIGHTS OF SIX PLANTS

	Normal	Off-Type	Aberrant	
	101	84	32	
	105	88		
	94			
Totals	$\overline{300}$	$\overline{172}$	$\overline{32}$	Grand total = 504

For the entries in this table let  $y_{ij}$  denote the weight of the  $j$ th plant of the  $i$ th type,  $i$  taking values 1, 2, 3 for normal, off-type and aberrant, respectively, and  $j = 1, 2, \dots, n_i$ , where  $n_i$  is the number of observations in the  $i$ th type. The problem is to estimate the effect of type of plant on weight of plant. To do this we assume the observation  $y_{ij}$  is the sum of three parts

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad (1)$$

where  $\mu$  represents the population mean of the weight of plant,  $\alpha_i$  is the effect of type  $i$  on weight, and  $e_{ij}$  is a random residual term peculiar to the observation  $y_{ij}$ .

As in regression it is assumed that the  $e_{ij}$ 's are independently distributed with zero mean, i.e.,  $E(e_{ij}) = 0$ . Then  $E(y_{ij}) = \mu + \alpha_i$ . It is also assumed that each  $e_{ij}$  has the same variance,  $\sigma^2$ , so that the variance-covariance matrix of the vector of  $e$ -terms is  $\sigma^2 \mathbf{I}$ . Thus is the model specified, in this case the model of the general *1-way classification*; it is clearly a linear model since it is based on the assumption that  $y_{ij}$  is the simple sum of its three parts  $\mu$ ,  $\alpha_i$  and  $e_{ij}$ .

The problem is to estimate  $\mu$  and the  $\alpha_i$ 's, and also  $\sigma^2$ , the variance of the error terms. We will find that not all of the terms  $\mu$  and  $\alpha_i$  can be estimated satisfactorily, only certain linear functions of them can. At first thought this may seem to be a matter for concern; sometimes it is, but on many occasions it is not, because the number of linear functions that can be estimated is large and frequently includes those in which we are interested; e.g., differences between the effects, such as  $\alpha_1 - \alpha_2$ . Sometimes, however, functions of interest cannot be estimated, because of a paucity of data. On all occasions, though, a method is needed for ascertaining which functions can be estimated and which cannot. This is provided in what follows.

To develop the method of estimation we write down the six observations in terms of equation (1) of the model:

$$\begin{aligned} 101 &= y_{11} = \mu + \alpha_1 && + e_{11} \\ 105 &= y_{12} = \mu + \alpha_1 && + e_{12} \\ 94 &= y_{13} = \mu + \alpha_1 && + e_{13} \\ 84 &= y_{21} = \mu &+ \alpha_2 & + e_{21} \\ 88 &= y_{22} = \mu &+ \alpha_2 & + e_{22} \\ 32 &= y_{31} = \mu &+ \alpha_3 & + e_{31}. \end{aligned}$$

These equations are easily written in matrix form as

$$\begin{bmatrix} 101 \\ 105 \\ 94 \\ 84 \\ 88 \\ 32 \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \end{bmatrix}, \quad (2)$$

which is precisely the form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (3)$$

where  $\mathbf{y}$  is the vector of observations,  $\mathbf{X}$  is the matrix of 0's and 1's,  $\mathbf{b}$  is the vector of parameters to be considered,

$$\mathbf{b}' = [\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3],$$

and  $\mathbf{e}$  is the vector of error terms.

Note that (3) is exactly the same form as (7) of Chapter 14. In both cases  $\mathbf{y}$  and  $\mathbf{e}$  are vectors of observations and errors, respectively, and in both cases  $\mathbf{b}$  is a vector of parameters whose estimates we seek. Also, the properties of  $\mathbf{e}$  are the same:  $E(\mathbf{e}) = \mathbf{0}$ , so that  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ , and  $E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{I}$ . The only difference is in the form of  $\mathbf{X}$ : in regression it is a matrix of observations on  $x$ -variables, whereas in (3) its elements are either 0 or 1, depending on the absence or presence of particular terms of the model in each  $y_{ij}$  observation. But, with respect to applying the principle of least squares for estimating the elements of  $\mathbf{b}$ , there is no difference at all between (3) of this chapter and (7) of Chapter 14. Hence we can go directly to (10) of Chapter 14 to obtain an equation for  $\hat{\mathbf{b}}$ , the least squares estimator of  $\mathbf{b}$ :

$$\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}. \quad (4)$$

In regression analysis the solution  $\hat{\mathbf{b}}$  was obtained from this as  $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , where  $\mathbf{X}'\mathbf{X}$  was nonsingular and therefore had an inverse. Now, however,  $\mathbf{X}'\mathbf{X}$  has no inverse and we have to use the methods of generalized inverses to obtain a solution.

## 2. THE NORMAL EQUATIONS

### a. A general form

Equations (4) are called the *normal equations*. Before discussing their solution let us look briefly at their form. First, the vector of parameters,  $\mathbf{b}$ : it is the vector of all the elements of the model, in this case the elements  $\mu$ ,  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$ . And this is so in general; for example, if data can be arranged in rows and columns according to two different classifications, elements of the vector  $\mathbf{b}$  can be  $\mu$  and terms representing, say,  $a$  row effects and  $b$  column effects and up to  $ab$  interaction effects. Therefore  $\mathbf{b}$  can have as many as  $1 + a + b + ab$  elements in that case.

Second, the matrix  $\mathbf{X}$ : it is called the *incidence matrix* (or, for data from designed experiments, the *design matrix*), because the location of the 0's and 1's throughout its elements represents the incidence of terms of the model among the observations—and hence of the classifications in which the observations lie. This is particularly evident if one writes  $\mathbf{X}$  as a 2-way table with the parameters as headings to the columns and the observations as labels for the rows, as illustrated in Table 15.2.

TABLE 15.2.  $\mathbf{X}$  AS A 2-WAY TABLE (DATA OF TABLE 15.1)

Observations	Parameters of Model			
	$\mu$	$\alpha_1$	$\alpha_2$	$\alpha_3$
$y_{11}$	1	1	0	0
$y_{12}$	1	1	0	0
$y_{13}$	1	1	0	0
$y_{21}$	1	0	1	0
$y_{22}$	1	0	1	0
$y_{31}$	1	0	0	1

In Table 15.2, as in equations (2), it is clear that the sum of the last three columns equals the first column. (This is so because every  $y_{ij}$  observation contains  $\mu$  and so the first column of  $\mathbf{X}$  is all 1's; and every  $y_{ij}$  also contains just one  $\alpha$  and so the sum of the last three columns is also all 1's.) Thus  $\mathbf{X}$  is not of full column rank.

Now consider the normal equations (4). They involve  $X'X$ , which is obviously square and symmetric. Its elements are the inner products of the columns of  $X$  with each other, as in (4) of Section 3.1b. Hence

$$X'X = \begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

Furthermore, because  $X$  does not have full column rank,  $X'X$  is not of full rank.

The normal equations (4) also involve the vector  $X'y$ ; its elements are the inner products of the columns of  $X$  with the vector  $y$ , and since the only nonzero elements of  $X$  are ones, the elements of  $X'y$  are certain sums of elements of  $y$ ; e.g., from (2) and (3)

$$X'y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} y_{1..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \end{bmatrix} = \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix}. \quad (6)$$

This is often the nature of  $X'y$  in linear models—a vector of various subtotals of the observations in  $y$ .

#### b. Many solutions

Whenever  $X'X$  is not of full rank, as in (2), the normal equations (4) cannot be solved with one solitary solution  $\hat{\mathbf{b}} = (X'X)^{-1}X'y$  as in Chapter 14. Many solutions are available. To emphasize this we write the normal equations as

$$X'X\mathbf{b}^0 = X'y, \quad (7)$$

using the symbol  $\mathbf{b}^0$  to distinguish those many solutions (7) from the solitary solution that exists when  $X'X$  has full rank.

**Example** (continued). The normal equations (7) for the data of Table 15.1 are, from (5) and (6)

$$\begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \alpha_3^0 \end{bmatrix} = \begin{bmatrix} y_{1..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \end{bmatrix} = \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix}. \quad (8)$$

Many comments can be made about the detailed form of these equations, for which the reader is referred to Searle (1971, pp. 168 and 232).

### 3. SOLVING THE NORMAL EQUATIONS

We seek a solution for  $\mathbf{b}^0$  to the normal equations (7):  $X'X\mathbf{b}^0 = X'y$ . Since  $X'X$  is singular, it has no inverse, and so there are many solutions for  $\mathbf{b}^0$ , and not just one as there is in regression, in (11) of Section 14.2. This is usually the case with linear models of general nature being discussed here:  $X'X$  is almost invariably singular. We therefore use the results of Chapter 9 to obtain a solution for  $\mathbf{b}^0$  using a generalized inverse of  $X'X$ .

#### a. Generalized inverses of $X'X$

Let  $G$  be a generalized inverse of  $X'X$  satisfying

$$X'XGX'X = X'X \quad (9)$$

in accord with (5) of Section 8.2. Then, on defining

$$X \text{ as having order } N \times p \text{ with } p < N, \text{ and rank } r_X = r \leq p, \quad (10)$$

the theorem in Section 8.6c gives:

$$\begin{array}{ll} \text{(i) } X'XGX'X = X'X, & \text{(ii) } XGX'X = X, \\ \text{(iii) } XGX' \text{ is invariant to } G & \text{(iv) } XGX' \text{ is symmetric.} \end{array} \quad (11)$$

Furthermore, on also defining

$$H = GX'X, \quad (12)$$

we have, following (18) of Section 9.4b,

$$\text{(i) } H \text{ is idempotent, (ii) } r_H = r_X = r \text{ and (iii) } r(H - I) = p - r. \quad (13)$$

#### b. Solutions

Equation (18) of Section 9.4b indicates that

$$\mathbf{b}^0 = GX'y + (H - I)z \quad (14)$$

for arbitrary  $z$  is a solution for  $\mathbf{b}^0$  of the normal equations (7). And Theorems 5 and 6 of Section 9.4c tell us that all possible solutions can be obtained this way, or that they can be obtained using  $GX'y$  for all possible

generalized inverses  $\mathbf{G}$ . For our purposes, only one solution is needed, and we take it as  $\mathbf{GX}'\mathbf{y}$  for some generalized inverse  $\mathbf{G}$ . Thus we use

$$\mathbf{b}^0 = \mathbf{GX}'\mathbf{y} \quad (15)$$

as a solution for the normal equations (7).

The lack of uniqueness to solutions of the normal equations is discussed in many places, usually in terms of including with equations like (8) an equation that is often called a "convenient restraint" or "obvious restriction" (such as  $\alpha_1 + \alpha_2 + \alpha_3 = 0$ ). Taken in combination, the equations then have a single solution. But this procedure is convenient only for very particular cases of (7), and for the general case there *are* no "convenient" or "obvious" restrictions leading to unique solutions. Although a great deal has been written about obtaining solutions through adding extra equations ("conditions", "restraints" or "constraints") of this nature, and many texts on experimental design refer to them, they can create many difficulties when data have unequal numbers of observations in the subclasses (see, for example, Searle, Speed and Henderson, 1981). This whole procedure is avoided here by considering the solution (15), obtained from using a generalized inverse.

The notation  $\mathbf{b}^0$  in (15) for a solution to the normal equations (7) emphasizes that what is derived by solving (7) is *only* a solution to the normal equations and *not* an estimator of  $\mathbf{b}$ . This fact cannot be over-emphasized. In a general discussion of linear models that are not of full rank, it is essential to realize that what is obtained as a solution of the normal equations is just that, a solution and *nothing more*. It is misleading and in most cases wrong for  $\mathbf{b}^0$  to be termed an estimator of  $\mathbf{b}$ . It is true that  $\mathbf{b}^0$  is, as will be shown, an estimator of something, but not of  $\mathbf{b}$ , and indeed the expression it estimates depends entirely upon which generalized inverse of  $\mathbf{X}'\mathbf{X}$  is used in obtaining  $\mathbf{b}^0$ . For this reason  $\mathbf{b}^0$  is always referred to as a solution and not an estimator.

Any particular  $\mathbf{b}^0$  depends on which particular generalized inverse of  $\mathbf{X}'\mathbf{X}$  is used as  $\mathbf{G}$  in calculating  $\mathbf{b}^0 = \mathbf{GX}'\mathbf{y}$ . There is therefore an infinite number of solutions  $\mathbf{b}^0$ —except when  $\mathbf{X}'\mathbf{X}$  has full rank, whereupon  $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1}$  and  $\mathbf{b}^0 = \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  is then the only solution. Nevertheless, we find that the properties of  $\mathbf{G}$  and  $\mathbf{H}$  given in (11) and (13) ensure that this lack of uniqueness in  $\mathbf{b}^0$  is less traumatic than might be anticipated.

**Example** (continued). From (8) a generalized inverse of

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \text{ is } \mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (16)$$

for which the solution (15) is

$$\mathbf{b}^0 = \mathbf{GX}'\mathbf{y} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix} = \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix} \quad (17)$$

Another generalized inverse and corresponding value of  $\mathbf{b}^0$  is

$$\hat{\mathbf{G}} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{5}{6} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & 1\frac{1}{2} \end{bmatrix} \text{ and } \hat{\mathbf{b}}^0 = \hat{\mathbf{G}}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 86 \\ 14 \\ 0 \\ -54 \end{bmatrix} \quad (18)$$

#### 4. EXPECTED VALUES AND VARIANCES

A general statement of the expected values given following (1) is

$$E(\mathbf{e}) = \mathbf{0} \quad \text{and} \quad E(\mathbf{y}) = \mathbf{Xb}. \quad (19)$$

Hence, with  $\mathbf{b}^0 = \mathbf{GX}'\mathbf{y}$ , and using (4) of Section 13.1,

$$E(\mathbf{b}^0) = \mathbf{GX}'E(\mathbf{y}) = \mathbf{GX}'\mathbf{Xb} = \mathbf{Hb}, \quad (20)$$

for  $\mathbf{H} = \mathbf{GX}'\mathbf{X}$  of (12). Thus  $\mathbf{b}^0$  has expected value  $\mathbf{Hb}$ , and so is an unbiased estimator of  $\mathbf{Hb}$  but *not* of  $\mathbf{b}$ .

With  $E(\mathbf{y}) = \mathbf{Xb}$  of (19), the variance-covariance matrix of  $\mathbf{y}$  is, in keeping with (3) of Section 13.1,

$$\text{var}(\mathbf{y}) = E(\mathbf{y} - \mathbf{Xb})(\mathbf{y} - \mathbf{Xb})' = E(\mathbf{ee}')$$

and, similarly,

$$\text{var}(\mathbf{e}) = E(\mathbf{e} - \mathbf{0})(\mathbf{e} - \mathbf{0})' = E(\mathbf{ee}').$$

Furthermore, as indicated following (1), we assume the error terms to be uncorrelated with the variance of each being  $\sigma^2$ , so that by (2) of Section

13.1, we are assuming  $\text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}$ . Therefore

$$\text{var}(\mathbf{y}) = \text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}. \quad (21)$$

And then, using (5) of Section 13.1,

$$\text{var}(\mathbf{b}^0) = \text{var}(\mathbf{GX}'\mathbf{y}) = \mathbf{GX}'\sigma^2\mathbf{IXG}' = \mathbf{GX}'\mathbf{XG}'\sigma^2. \quad (22)$$

Although this is not the analogue of its counterpart  $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$  in regression [see (40) of Section 14.6] as would be  $\mathbf{G}\sigma^2$ , we find that (22) does, in fact, create no difficulties in applications. And in using a symmetric, reflexive, generalized inverse  $\mathbf{G}^* = \mathbf{GX}'\mathbf{XG}'$  (see Section 8.3) in place of  $\mathbf{G}$  we get (22) as  $\mathbf{G}^*\sigma^2$ , should that direct analogue with  $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$  be required.

## 5. PREDICTED $y$ -VALUES

Just as in Section 14.7 dealing with predicted  $y$ -values in regression, so here. We denote by  $\hat{\mathbf{y}}$  the vector  $\mathbf{Xb}^0$  of predicted  $y$ -values corresponding to the vector of observations  $\mathbf{y}$ :

$$\hat{\mathbf{y}} = \mathbf{Xb}^0 = \mathbf{XGX}'\mathbf{y}. \quad (23)$$

Because  $\mathbf{G}$  occurs in  $\hat{\mathbf{y}}$  only in the form  $\mathbf{XGX}'$ , we use (iii) of (11) to observe that  $\hat{\mathbf{y}}$  is invariant to  $\mathbf{G}$ . This means that it does not matter what generalized inverse of  $\mathbf{X}'\mathbf{X}$  is used for  $\mathbf{G}$ : for each and every  $\mathbf{G}$ , the value  $\hat{\mathbf{y}} = \mathbf{XGX}'\mathbf{y}$  of (23) is the same. This is further emphasized by noting that for any  $\mathbf{b}^0$  of the general form (14), i.e., for any  $\mathbf{z}$  in (14),

$$\hat{\mathbf{y}} = \mathbf{Xb}^0 = \mathbf{X}[\mathbf{GX}'\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}] = \mathbf{XGX}'\mathbf{y} + (\mathbf{XH} - \mathbf{X})\mathbf{z} = \mathbf{XGX}'\mathbf{y}$$

because, by (12) and (ii) of (11),  $\mathbf{XH} = \mathbf{X}$ .

**Example** (continued).  $\mathbf{X}$  is the matrix of 0's and 1's in (2). Using  $\mathbf{b}^0$  of (17) in  $\mathbf{Xb}^0$  of (23) therefore gives

$$\hat{\mathbf{y}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 100 \\ 86 \\ 86 \\ 32 \end{bmatrix} \quad (24)$$

and using  $\mathbf{b}^0$  of (18) for  $\mathbf{b}^0$  in (23) gives the same value for  $\hat{\mathbf{y}}$ :

$$\hat{\mathbf{y}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 86 \\ 14 \\ 0 \\ -54 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 100 \\ 86 \\ 86 \\ 32 \end{bmatrix}.$$

## 6. ESTIMATING THE ERROR VARIANCE

### a. Error sum of squares

Having obtained  $\hat{\mathbf{y}} = \mathbf{Xb}^0$  as the vector of predicted values corresponding to the vector of observations  $\mathbf{y}$ , we can now calculate the sum of squares of the deviations of the observed  $y$ 's from their predicted values. This is the residual (or error) sum of squares, SSE, and in regression analysis (Section 14.8) we found it to be expressible as  $\text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}$ . Essentially the same result holds for the general linear model of this chapter. As before,

$$\text{SSE} = \sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2 = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}),$$

so that now

$$\begin{aligned} \text{SSE} &= (\mathbf{y} - \mathbf{Xb}^0)'(\mathbf{y} - \mathbf{Xb}^0) = \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')(\mathbf{I} - \mathbf{XGX}')\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y}, \end{aligned} \quad (25)$$

because from (11),  $\mathbf{XGX}'$  is both idempotent and symmetric. The occurrence of  $\mathbf{G}$  in (25) only in the form  $\mathbf{XGX}'$  has the same consequence as it does in  $\hat{\mathbf{y}}$  of (23): the residual sum of squares SSE is invariant to whatever generalized inverse of  $\mathbf{X}'\mathbf{X}$  is used for  $\mathbf{G}$ , and, equivalently, is invariant to whatever solution of the normal equations  $\mathbf{X}'\mathbf{Xb}^0 = \mathbf{X}'\mathbf{y}$  is used for  $\mathbf{b}^0$ . This is especially useful because from (25) a computing formula is obtainable that is based on  $\mathbf{b}^0$ :

$$\begin{aligned} \text{SSE} &= \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{XGX}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}^0'\mathbf{X}'\mathbf{y}. \end{aligned} \quad (26)$$

This is exactly the same as in regression, at (50) of Section 14.8:  $\mathbf{y}'\mathbf{y}$  is the total sum of squares of the observed  $y$ 's; and  $\mathbf{b}^0'\mathbf{X}'\mathbf{y}$  is the sum of products of the solutions in  $\mathbf{b}^0$ ' multiplied by the corresponding elements of the right-hand sides of the equations  $\mathbf{X}'\mathbf{Xb}^0 = \mathbf{X}'\mathbf{y}$  from which  $\mathbf{b}^0$  is derived.

**Example** (continued). From (2) the total sum of squares, SST, is

$$\text{SST} = y'y = 101^2 + 105^2 + 94^2 + 84^2 + 88^2 + 32^2 = 45,886; \quad (27)$$

and using (17) in  $\mathbf{b}^0'X'y$ , with  $X'y$  being the right-hand side of (8), gives

$$\mathbf{b}^0'X'y = 0(504) + 100(300) + 86(172) + 32(32) = 45,816. \quad (28)$$

Similarly, using  $\hat{\mathbf{b}}^0$  of (18) gives the same value:

$$\hat{\mathbf{b}}^0'X'y = 86(504) + 14(300) + 0(172) - 54(32) = 45,816. \quad (29)$$

Hence, on using either solution (17) or (18), calculation of SSE from (26) is

$$\text{SSE} = y'y - \mathbf{b}^0'X'y = 45,886 - 45,816 = 70. \quad (30)$$

#### b. Expected value

From (19) and (21), the mean and variance of  $y$  are, respectively,  $E(y) = X\mathbf{b}$  and  $\text{var}(y) = \sigma^2\mathbf{I}$ . And in (25), SSE is a quadratic form in  $y$ . Therefore (28) of Section 13.5 can be used to derive the expected value of SSE as

$$E(\text{SSE}) = E[y'(I - XGX')y] = \text{tr}[(I - XGX')I\sigma^2] + \mathbf{b}'X'(I - XGX')X\mathbf{b}.$$

On using the results given in (11) to (13) this simplifies to

$$\begin{aligned} E(\text{SSE}) &= \sigma^2[\text{tr}(\mathbf{I}) - \text{tr}(XGX')] + \mathbf{b}'X'(X - XGX'X)\mathbf{b} \\ &= \sigma^2[N - \text{tr}(GX'X)] + 0 = \sigma^2(N - r_H) \\ &= (N - r_X)\sigma^2. \end{aligned} \quad (31)$$

#### c. Estimation

Equation (31) immediately gives an unbiased estimator of  $\sigma^2$  as

$$\hat{\sigma}^2 = \frac{\text{SSE}}{N - r_X}. \quad (32)$$

Again we see a similarity with regression, since (31) is the same as the first equality in (53) of Section 14.8. Only now, the importance of using  $r_X$  in the expectation is clear, because  $X$  is not of full column rank and its rank is therefore not equal to its number of columns. In fact, the rank of  $X$  depends on the nature of the data available.

**Example.**

$$\hat{\sigma}^2 = \frac{70}{6 - 3} = 23\frac{1}{3}. \quad (33)$$

### 7. PARTITIONING THE TOTAL SUM OF SQUARES

Partitioning the total sum of squares as shown in Section 14.9 for regression takes place in the same way for the general linear model being dealt with in this chapter. One difference is that corrected sums of squares and products of  $x$ -variables are of neither use nor interest (because every  $x$ -variable is just 0 or 1, representing the incidence of a term of the model in the data), and so matrices such as  $X'X$  do not arise. Nevertheless, use is still made of  $\text{SST}_m = y'y - N\bar{y}^2$ , the corrected sum of squares of the  $y$ -observations. The three forms of partitioning the sums of squares are shown in Table 15.3.

TABLE 15.3. PARTITIONING SUMS OF SQUARES

$\text{SSR} = y'XGX'y$	$\text{SSM} = N\bar{y}^2$	$\text{SSR}_m = y'XGX'y - N\bar{y}^2$	$\text{SSR}_m = y'XGX'y - N\bar{y}^2$
$\text{SSE} = y'(I - XGX')y$	$\text{SSE} = y'(I - XGX')y$	$\text{SSE} = y'(I - XGX')y$	$\text{SSE} = y'(I - XGX')y$
$\text{SST} = y'y$	$\text{SST} = y'y$	$\text{SST}_m = y'y - N\bar{y}^2$	

The three columns in Table 15.3 correspond to the three partitionings shown in (56), (57) and (59) of Chapter 14. The first column shows

$$\text{SSR} = \text{SST} - \text{SSE} = y'XGX'y = \mathbf{b}^0'X'y, \quad (34)$$

the sum of squares attributable to fitting the model  $y = X\mathbf{b} + \mathbf{e}$ , similar to the sum of squares for regression in Section 14.9. In the second column,

$$\text{SSM} = N\bar{y}^2 \quad (35)$$

is the sum of squares due to fitting a general mean, and

$$\text{SSR}_m = \text{SSR} - \text{SSM} = \text{SSR} - N\bar{y}^2 \quad (36)$$

is the sum of squares for fitting the model, corrected for the mean. The third column is identical to the second except that SSM has been deleted from the

body of the table and subtracted from SST to give

$$SST_m = SST - SSM = \sum_{i,j} y_{ij}^2 - N\bar{y}^2 \quad (37)$$

as the total sum of squares corrected for the mean. In all three columns the error sum of squares is the same, SSE of (26).

Table 15.3 forms the basis of traditional analysis of variance tables, as is shown in Section 9.

**Example** (continued). With  $\bar{y} = 504/6$  from Table 15.1, we have  $N\bar{y}^2 = 42,336$ ; and with  $SST = 45,886$  and  $SSE = 70$  from (27) and (30), respectively, Table 15.3 for the example is as shown in Table 15.4.

TABLE 15.4. PARTITIONING SUMS OF SQUARES AS IN TABLE 15.3  
(DATA OF TABLE 15.1)

	SSM = 42,336	
SSR = 45,816	SSR <sub>m</sub> = 3,480	SSR <sub>m</sub> = 3,480
SSE = 70	SSE = 70	SSE = 70
SST = 45,886	SST = 45,886	SST <sub>m</sub> = 3,550

## 8. COEFFICIENT OF DETERMINATION

The predicted values of  $y$  corresponding to the observations  $y$  are the elements of  $\hat{y}$  given in (23). The product-moment correlation between the observed  $y$ 's and the corresponding elements of  $\hat{y}$  is, when squared, commonly referred to as the *coefficient of determination*. Since the usual linear model contains  $\mu$ , which corresponds to the intercept in regression, we define

$$R^2 = \text{coefficient of determination} \\ = \frac{[\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2} \quad (38)$$

similar to (60) of Section 14.10. In simplifying (38), use is made of  $\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}' = \mathbf{X}'$  obtainable from (11), which, because the first row of  $\mathbf{X}'$  is  $\mathbf{1}'$ , also means that

$$\mathbf{1}'\mathbf{X}\mathbf{G}\mathbf{X}' = \mathbf{1}' \quad (39)$$

This leads to the equality  $\bar{\hat{y}} = \bar{y}$ ; hence, as in equation (61) of Section 14.10,

$$R^2 = \frac{(\text{SSR}_m)^2}{\text{SST}_m(\text{SSR}_m)} = \frac{\text{SSR}_m}{\text{SST}_m} \quad (40)$$

**Example** (continued). From Table 15.4,  $R^2 = 3480/3550 = .98$ .

## 9. ANALYSIS OF VARIANCE

Analysis of variance is based on partitioning sums of squares in the manner of Tables 15.3 and 15.4. The right-most column of those tables is the partitioning seen in most analysis of variance tables, with the total sum of squares being that corrected for the mean,  $\text{SST}_m = \mathbf{y}'\mathbf{y} - N\bar{y}^2$ . In contrast, we will concentrate attention on the middle column of Tables 15.3 and 15.4, which has the term  $\text{SSM} = N\bar{y}^2$  in it, as well as  $\text{SSR}_m$  and  $\text{SSE}$ . The analysis of variance based on this column is shown in Table 15.5, first in terms of the sums of squares in Table 15.3, and then for the data, using Table 15.4.

Table 15.5 is just a summary of calculations. To utilize the  $F$ -statistics therein, in the usual manner, we assume that the observations in  $\mathbf{y}$  are normally distributed with mean  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$  and variance-covariance matrix  $\text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$ , already specified in (19) and (21), respectively, and write

$$\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2\mathbf{I}), \quad (41)$$

in the notation introduced in Section 13.4. Applying the theorem of that section to the sums of squares  $\text{SSM}$ ,  $\text{SSR}_m$  and  $\text{SSE}$  then establishes that the  $F$ -statistics in Table 15.5 have  $F$ -distributions with degrees of freedom corresponding to those of the mean squares in their numerator and denominator. Establishing these results is beyond the scope of this book. Details are available in many of the texts mentioned at the beginning of this chapter (e.g., Searle, 1971, Section 5.3).

The hypothesis tested by  $F(M)$  in Table 15.5a is  $H: E(\bar{y}) = 0$ ; and  $F(M)$  is the square of the  $t$ -statistic  $[\bar{y}/(\hat{\sigma}/\sqrt{N})]^2$  with  $N - r_{\mathbf{X}}$  degrees of freedom. Both these results are identical to their counterparts in regression (Section 14.12), where  $r_{\mathbf{X}} = k + 1$ .

The hypothesis tested by  $F(R_m)$  in Table 15.5a is described by first considering an  $F$ -statistic not shown in Table 15.5, namely  $F(R)$  based on recombining  $\text{SSM}$  and  $\text{SSR}_m$  into  $\text{SSR}$  and calculating

$$\text{MSR} = \text{SSR}/r_{\mathbf{X}} \quad \text{and} \quad F(R) = \text{MSR}/\text{MSE}.$$



TABLE 15.5. ANALYSIS OF VARIANCE

Source of Variation	d.f. <sup>a</sup>	Sum of Squares	Mean Square	F-Statistic
(a) Based on Table 15.3				
Mean	1	SSM	$MSM = \frac{SSM}{1}$	$F(M) = \frac{MSM}{MSE}$
Model, a.f.m. <sup>b</sup>	$r_X - 1$	$SSR_m$	$MSR_m = \frac{SSR_m}{(r_X - 1)}$	$F(R_m) = \frac{MSR_m}{MSE}$
Error	$N - r_X$	SSE	$MSE = \frac{SSE}{(N - r_X)}$	
Total	$N$	SST		
(b) For Data of Table 15.1, Based on Table 15.4				
Mean	1	42,336	$\frac{42,336}{1} = 42,336$	$F(M) = \frac{42,336}{23\frac{1}{3}} = 1,814\frac{2}{3}$
Model, a.f.m. <sup>b</sup>	2	3,480	$\frac{3,480}{2} = 1,740$	$F(R_m) = \frac{1,740}{23\frac{1}{3}} = 74\frac{2}{3}$
Error	3	70	$\frac{70}{3} = 23\frac{1}{3}$	
Total	6	45,886		

<sup>a</sup>d.f. = degrees of freedom.

<sup>b</sup>a.f.m. = adjusted for mean.

The theorem of Section 13.4 leads to also establishing  $F(R)$  as having an  $F$ -distribution. This is just like the  $F(R)$  of regression which, as in (63) of Section 14.12, is seen to provide a test of  $H: \mathbf{b} = \mathbf{0}$ . In contrast, for the linear model  $E(y) = \mathbf{Xb}$  where  $\mathbf{X}$  does not have full column rank (as is being dealt with in this chapter),  $F(R)$  tests  $H: \mathbf{Xb} = \mathbf{0}$ . This is *not* equivalent to  $H: \mathbf{b} = \mathbf{0}$  because, with  $\mathbf{X}$  having less than full column rank,  $\mathbf{Xb} = \mathbf{0}$  does not imply  $(\neq) \mathbf{b} = \mathbf{0}$ . [At all times,  $\mathbf{Xb} = \mathbf{0} \Rightarrow \mathbf{X}'\mathbf{Xb} = \mathbf{0}$ , and when  $\mathbf{X}$  does have full column rank,  $\mathbf{X}'\mathbf{X}$  is nonsingular, as in regression (see Chapter 14) and so then  $\mathbf{Xb} = \mathbf{0} \Rightarrow \mathbf{X}'\mathbf{Xb} = \mathbf{0} \Rightarrow \mathbf{b} = \mathbf{0}$ .]

When  $F(R)$  is significant we say that there is concordance of the data with the model  $E(y) = \mathbf{Xb}$ ; i.e., the model accounts for a significant portion of the variation in the  $y$ -variable. This does not mean that the model used is necessarily the most suitable model; there may be a subset of its elements that is as significant as the whole set; or there may be other elements

(factors) which, when used alone or in combination with some or all of those already used, are significantly better than those used; or there may be nonlinear models that are at least as suitable as the model used. None of these contingencies is inconsistent with  $F(R)$  being significant and the ensuing conclusion that the data are in concordance with the model  $E(y) = \mathbf{Xb}$ .

Just as  $F(R)$  provides a test of the model  $E(y) = \mathbf{Xb}$ , so does  $F(R_m)$  of Table 15.5a provide a test of the model  $E(y) = \mathbf{Xb}$  over and above the model  $E(y) = \mu$ . Since the latter can be considered as fitting a general mean, we look upon  $F(R_m)$  as providing a test of the model  $E(y) = \mathbf{Xb}$  over and above the mean. When  $F(R_m)$  is significant we conclude that the model satisfactorily accounts for variation in the  $y$ -variable. This is not to be taken as evidence that all elements of  $\mathbf{b}$  other than  $\mu$  are nonzero but only that at least one of them, or one linear combination of them, may be. If  $F(M)$  has first been found significant, then  $F(R_m)$  being significant indicates that a model with terms in it additional to a mean explains significantly more of the variation in the  $y$ -variable than does the model  $E(y) = \mu$ .

**Example** (continued). The particular hypotheses tested by  $F(M)$  and  $F(R_m)$  in Table 15.5b are as follows:

$$F(M) \text{ tests } H: \mu + \frac{1}{6}(3\alpha_1 + 2\alpha_2 + \alpha_3) = 0 \quad (42)$$

and

$$F(R_m) \text{ tests } H: \alpha_1 = \alpha_2 = \alpha_3. \quad (43)$$

The latter is the familiar hypothesis of equality of the effects of type of plant on yield.

## 10. THE $R(\cdot | \cdot)$ NOTATION

In Section 14.15 the notation  $R(\mathbf{b}_1 | \mathbf{b}_2)$  was introduced in the context of regression as the sum of squares due to fitting  $E(y) = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2$  after fitting  $E(y) = \mathbf{X}_2\mathbf{b}_2$ . This is now applied to the model  $E(y_{ij}) = \mu + \alpha_i$  of (1), denoting the sum of squares for this as  $R(\mu, \alpha)$ . It is, of course, SSR in the analysis of variance:

$$R(\mu, \alpha) = SSR. \quad (44)$$

Similarly,  $R(\mu)$  denotes the sum of squares due to fitting  $E(y_{ij}) = \mu$  and so

$$R(\mu) = N\bar{y}^2 = SSM. \quad (45)$$

Hence, in combination with (36) and analogous to (88) of Section 14.16,

$$SSR_m = SSR - SSM = R(\mu, \alpha) - R(\mu) = R(\alpha | \mu). \quad (46)$$

In this way  $R(\alpha | \mu)$  indicates by its very symbolism that it is the sum of squares due to having  $\alpha$ 's in the model, over and above having just  $\mu$ ; i.e., that it is the sum of squares due to fitting  $E(y_{ij}) = \mu + \alpha_i$ , over and above that due to fitting  $E(y_{ij}) = \mu$ . This proves to be a useful notation in models of this nature that have more than one factor, as indicated in Section 14.

## 11. ESTIMABLE FUNCTIONS

Equation (20) shows that  $\mathbf{b}^0$  is not an unbiased estimator of  $\mathbf{b}$ . Nevertheless, unbiased estimators of certain linear combinations of elements of  $\mathbf{b}$  can be obtained from  $\mathbf{b}^0$ , and they are also estimators that are invariant to  $\mathbf{b}^0$ ; i.e., invariant to whatever generalized inverse of  $\mathbf{X}'\mathbf{X}$  is used for  $\mathbf{G}$ . The situation is as follows.

Applying Theorem 9 of Section 9.6 to the normal equations  $\mathbf{X}'\mathbf{X}\mathbf{b}^0 = \mathbf{X}'\mathbf{y}$  and their solution  $\mathbf{b}^0 = \mathbf{G}\mathbf{X}'\mathbf{y}$  indicates that  $\mathbf{q}'\mathbf{b}^0$  is invariant to  $\mathbf{b}^0$  whenever  $\mathbf{q}'$  is such that  $\mathbf{q}'\mathbf{H} = \mathbf{q}'$ , for  $\mathbf{H} = \mathbf{G}\mathbf{X}'\mathbf{X}$  of (12). Furthermore, any  $\mathbf{q}'$  of the form  $\mathbf{q}' = \mathbf{w}'\mathbf{H}$ , no matter what the vector  $\mathbf{w}'$  is, satisfies the condition  $\mathbf{q}'\mathbf{H} = \mathbf{q}'$ , because then  $\mathbf{q}'\mathbf{H} = \mathbf{w}'\mathbf{H}^2 = \mathbf{w}'\mathbf{H} = \mathbf{q}'$ , since  $\mathbf{H}$  is idempotent as in (13). Thus for an arbitrary vector  $\mathbf{w}'$  and  $\mathbf{q}' = \mathbf{w}'\mathbf{H}$ , the linear combination  $\mathbf{q}'\mathbf{b}^0$  of elements of  $\mathbf{b}^0$  is invariant to  $\mathbf{b}^0$ ; and, using (11) and (20), its expected value is

$$E(\mathbf{q}'\mathbf{b}^0) = \mathbf{w}'\mathbf{H}E(\mathbf{b}^0) = \mathbf{w}'\mathbf{H}\mathbf{H}\mathbf{b} = \mathbf{w}'\mathbf{H}^2\mathbf{b} = \mathbf{w}'\mathbf{H}\mathbf{b} = \mathbf{q}'\mathbf{b}, \quad (47)$$

i.e.,  $\mathbf{q}'\mathbf{b}^0$  is an unbiased estimator of  $\mathbf{q}'\mathbf{b}$ . Thus it is that any linear combination  $\mathbf{q}'\mathbf{b}^0$  of elements of the solution vector  $\mathbf{b}^0$  that is invariant to  $\mathbf{b}^0$  is an unbiased estimator of the same combination,  $\mathbf{q}'\mathbf{b}$ , of elements of the parameter vector,  $\mathbf{b}$ . This is precisely the property of estimability: such a function  $\mathbf{q}'\mathbf{b}$  is called an *estimable function* and  $\mathbf{q}'\mathbf{b}^0$  is an unbiased estimator of  $\mathbf{q}'\mathbf{b}$ .

More than that,  $\mathbf{q}'\mathbf{b}^0$  is the b.l.u.e. (best, linear, unbiased estimator) of  $\mathbf{q}'\mathbf{b}$ . This means that not only is it unbiased, as in (47), and linear in the elements of  $\mathbf{y}$  (because  $\mathbf{q}'\mathbf{b}^0 = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{y}$ ), but it is also "best" in the sense that among all linear (in  $\mathbf{y}$ ) unbiased estimators of  $\mathbf{q}'\mathbf{b}$  it is the one with smallest variance (see, for example, Searle, 1971, p. 182). This variance has a very

simple form, obtained by noticing that  $\mathbf{q}' = \mathbf{w}'\mathbf{H}$  implies  $\mathbf{q}' = \mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{X}$ ; i.e.,

$$\mathbf{q}' = \mathbf{t}'\mathbf{X} \quad (48)$$

for some  $\mathbf{t}'$ , namely  $\mathbf{t}' = \mathbf{w}'\mathbf{G}\mathbf{X}'$ . Then, denoting the variance of the scalar  $\mathbf{q}'\mathbf{b}^0$  by  $v(\mathbf{q}'\mathbf{b}^0)$ , its value is, using (5) of Section 13.1,

$$\begin{aligned} v(\mathbf{q}'\mathbf{b}^0) &= \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{q}\sigma^2, && \text{from (22)} \\ &= \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{X}'\mathbf{t}\sigma^2, && \text{from (48)} \\ &= \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{t}\sigma^2, && \text{from (11)} \\ &= \mathbf{q}'\mathbf{G}\mathbf{q}\sigma^2, && \text{from (48)}. \end{aligned} \quad (49)$$

Notice that with  $\mathbf{q}' = \mathbf{t}'\mathbf{X}$  of (48), the variance in (49) is also  $v(\mathbf{q}'\mathbf{b}^0) = \mathbf{t}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{t}\sigma^2$ , which is invariant to  $\mathbf{G}$ . This is the kind of result alluded to following (22). Whereas  $\mathbf{b}^0$  is not invariant to  $\mathbf{G}$  and is accordingly of little interest in itself,  $\mathbf{q}'\mathbf{b}^0$  for  $\mathbf{q}'\mathbf{b}$  being estimable is not only invariant to  $\mathbf{G}$ , it is also of interest, being the b.l.u.e. of  $\mathbf{q}'\mathbf{b}$ , and its variance  $v(\mathbf{q}'\mathbf{b}^0) = \mathbf{q}'\mathbf{G}\mathbf{q}\sigma^2$  is also invariant to  $\mathbf{G}$ .

We use the traditional "hat" over an expression to indicate the b.l.u.e. of an estimable function, and so when, for some  $\mathbf{t}'$ ,

$$\mathbf{q}' = \mathbf{t}'\mathbf{X}, \text{ or equivalently } \mathbf{q}' = \mathbf{q}'\mathbf{H} \text{ or } \mathbf{q}' = \mathbf{w}'\mathbf{H}, \quad (50)$$

then

$$\mathbf{q}'\mathbf{b} \text{ is an estimable function} \quad (51)$$

and

$$\text{the b.l.u.e. of } \mathbf{q}'\mathbf{b} \text{ is } \widehat{\mathbf{q}'\mathbf{b}} = \mathbf{q}'\mathbf{b}^0 \quad (52)$$

with variance

$$v(\mathbf{q}'\mathbf{b}^0) = \mathbf{q}'\mathbf{G}\mathbf{q}\sigma^2. \quad (53)$$

Two further aspects of estimable functions are very useful in application:

$$\begin{aligned} &\text{The expected value of any observation} \\ &\text{is an estimable function;} \end{aligned} \quad (54)$$

$$\begin{aligned} &\text{Linear combinations of estimable} \\ &\text{functions are estimable functions.} \end{aligned} \quad (55)$$

These, as available alternatives to (50), provide ready opportunity for developing estimable functions in any particular model, as is illustrated in Section 13.

Details of many aspects of estimable functions are given in Searle (1971, Section 5.4).

**Example** (continued). The model is  $E(y_{ij}) = \mu + \alpha_i$  of (1). Therefore, by (54),  $\mu + \alpha_i$  is estimable and so by (52)

$$\widehat{\mu + \alpha_i} = \mu^0 + \alpha_i^0.$$

Examples from (17) are

$$\widehat{\mu + \alpha_1} = 0 + 100 = 100 \quad \text{and} \quad \widehat{\mu + \alpha_2} = 0 + 86 = 86$$

and from (18)

$$\widehat{\mu + \alpha_1} = 86 + 14 = 100 \quad \text{and} \quad \widehat{\mu + \alpha_2} = 86 + 0 = 86,$$

so demonstrating the invariance of  $\mu^0 + \alpha_i^0$  to the value used for  $\mathbf{b}^0$ . Similarly, since the estimable function  $\mu + \alpha_1$  is

$$\mu + \alpha_1 = [1 \ 1 \ 0 \ 0]\mathbf{b},$$

using (16) in (53) gives

$$v(\widehat{\mu + \alpha_1}) = [1 \ 1 \ 0 \ 0] \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \sigma^2 = \frac{1}{3}\sigma^2,$$

and using (18) in (53) gives the same result:

$$\begin{aligned} v(\widehat{\mu + \alpha_1}) &= [1 \ 1 \ 0 \ 0] \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{5}{6} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & 1\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \sigma^2 \\ &= (\frac{1}{2} + \frac{5}{6} - 1)\sigma^2 = \frac{1}{3}\sigma^2. \end{aligned}$$

Further, from (55), because  $\mu + \alpha_1$  and  $\mu + \alpha_2$  are estimable, so is  $(\mu + \alpha_1) - (\mu + \alpha_2) = \alpha_1 - \alpha_2$ ; i.e.,  $\alpha_1 - \alpha_2$  is estimable. Therefore by (52) the b.l.u.e. of  $\alpha_1 - \alpha_2$  is

$$\widehat{\alpha_1 - \alpha_2} = \alpha_1^0 - \alpha_2^0 = 100 - 86 = 14,$$

using (17); and it is  $14 - 0 = 14$  using (18). Likewise, using (17) in (53)

$$\begin{aligned} v(\widehat{\alpha_1 - \alpha_2}) &= [0 \ 1 \ -1 \ 0] \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \sigma^2 \\ &= (\frac{1}{3} + \frac{1}{2})\sigma^2 = \frac{5}{6}\sigma^2 \end{aligned}$$

and the same value will be obtained using  $\mathbf{G}$  of (18) in place of  $\mathbf{G}$  of (17).

## 12. TESTING LINEAR HYPOTHESES

We now consider the calculations required for testing linear hypotheses, a linear hypothesis being a hypothesis that some linear function of the parameters equals some arbitrary constant. The discussion is confined to linear hypotheses involving estimable functions, since these are the only ones that can be tested. Thus in Section 11, the function  $\alpha_1 - \alpha_2$  is estimable and the hypothesis that  $\alpha_1 - \alpha_2$  equals some constant,  $m_1$  say, can be tested. We would write this hypothesis as

$$H: \alpha_1 - \alpha_2 = m_1.$$

More generally we might want to test the hypothesis

$$H: \begin{cases} \alpha_1 - \alpha_2 = m_1 \\ \alpha_2 - \alpha_3 = m_2 \end{cases} \quad (56)$$

which can also be written as

$$H: \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}. \quad (57)$$

This is a particular case of a general way of writing a linear hypothesis as

$$H: \mathbf{K}\mathbf{b} = \mathbf{m}, \quad (58)$$

where  $\mathbf{b}$  is, of course, the vector of parameters in the model  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ , and  $\mathbf{m}$  is the vector of known constants that form part of the hypothesis. For

example, if (56) were the hypothesis

$$H: \begin{cases} \alpha_1 - \alpha_2 = 6 \\ \alpha_2 - \alpha_3 = 5 \end{cases}$$

then on rewriting this as (58),  $\mathbf{m}$  would be  $\mathbf{m}' = [6 \ 5]$ .

The matrix  $\mathbf{K}'$  in the general statement (58) of a linear hypothesis must satisfy two simple requirements. Each is stated mathematically but has important practical ramifications. First,  $\mathbf{K}'$  must be such that each element of  $\mathbf{K}'\mathbf{b}$  is an estimable function. This is required because, in essence, the only linear hypotheses that can be tested are those which involve estimable functions. Second,  $\mathbf{K}'$  must have full row rank. This is a very practical requirement. It means that hypotheses such as

$$H_1: \begin{cases} \alpha_1 - \alpha_2 = 6 \\ 2\alpha_1 - 2\alpha_2 = 12 \end{cases} \quad \text{or} \quad H_2: \begin{cases} \alpha_1 - \alpha_2 = 6 \\ \alpha_2 - \alpha_3 = 5 \\ \alpha_1 - \alpha_3 = 11 \end{cases}$$

cannot be considered. This is very reasonable, because in each of these there is a redundant statement that need not be there. In  $H_1$  only the first statement  $\alpha_1 - \alpha_2 = 6$  need be considered, and in  $H_2$  any two of the statements will suffice, for the third one is the sum of the other two.

Provided  $\mathbf{K}'$  meets the two requirements just specified, and on assuming  $\mathbf{y} \sim N(\mathbf{Xb}, \sigma^2\mathbf{I})$ , the  $F$ -statistic for testing

$$H: \mathbf{K}'\mathbf{b} = \mathbf{m} \quad (58)$$

is calculated as follows. First, calculate the sum of squares

$$Q = (\mathbf{K}'\mathbf{b}^0 - \mathbf{m})'(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}(\mathbf{K}'\mathbf{b}^0 - \mathbf{m}). \quad (59)$$

Then

$$F(H) = Q/\hat{\sigma}^2 r_{\mathbf{K}} \quad (60)$$

is the  $F$ -statistic with  $r_{\mathbf{K}}$  and  $N - r_{\mathbf{X}}$  degrees of freedom. Details of deriving this  $F$ -statistic are beyond the scope of this book and are to be found in Searle (1971, Section 5.5b, with numerous ramifications considered in succeeding sections). It suffices to state here that the requirements placed on  $\mathbf{K}'$  in the preceding paragraph ensure that  $r_{\mathbf{K}}$ , the rank of  $\mathbf{K}$ , equals the number of rows in  $\mathbf{K}'$ , that the inverse  $(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}$  in (59) always exists, and that (59), and hence (60), is invariant to  $\mathbf{G}$ .

The importance of (58), (59) and (60) is their broad generality, limited only by the practical need for  $\mathbf{K}'$  to be of full row rank, with every element of  $\mathbf{K}'\mathbf{b}$  being an estimable function. Within this framework, (58), (59) and (60) provide the entire calculation procedure for the  $F$ -statistic for testing any linear hypothesis about  $\mathbf{b}$  in the model  $\mathbf{y} \sim N(\mathbf{Xb}, \sigma^2\mathbf{I})$ . This includes analysis of variance models with either equal or unequal numbers of observations in the subclasses of the data, be they survey data or data from well-designed and executed experiments. It also includes regression and analysis of variance in the presence of covariates. Thus the applicability of (58), (59) and (60) is very broad indeed.

**Example (continued).** We have seen that  $\alpha_1 - \alpha_2$  and  $\alpha_2 - \alpha_3$  are both estimable functions. We consider testing

$$H: \begin{cases} \alpha_1 - \alpha_2 = 10 \\ \alpha_2 - \alpha_3 = 50 \end{cases}, \quad \text{ie.,} \quad H: \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \mathbf{b} = \begin{bmatrix} 10 \\ 50 \end{bmatrix}. \quad (61)$$

Hence

$$\mathbf{K}' = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \text{with } r_{\mathbf{K}} = 2$$

and so for  $Q$  of (59)

$$\mathbf{K}'\mathbf{b} - \mathbf{m} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix} - \begin{bmatrix} 10 \\ 50 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

and

$$\begin{aligned} (\mathbf{K}'\mathbf{G}\mathbf{K})^{-1} &= \left\{ \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \right\}^{-1} \\ &= \begin{bmatrix} \frac{5}{6} & -\frac{1}{2} \\ -\frac{1}{2} & 1\frac{1}{2} \end{bmatrix}^{-1} = \frac{1}{6} \begin{bmatrix} 9 & 3 \\ 3 & 5 \end{bmatrix}. \quad (62) \end{aligned}$$

Therefore (59) is

$$Q = [4 \ 4] \frac{1}{6} \begin{bmatrix} 9 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = (16/6)(9 + 6 + 5) = 53\frac{1}{3} \quad (63)$$

14. SOME PARTICULAR MODELS

The general results of the preceding sections are now used in three particular models of widespread application.

a. The one-way classification

This is the model (1),  $E(y_{ij}) = \mu + \alpha_i$ , of the example used to this point. The general form of the normal equations (8) is

$$\begin{bmatrix} N & n_1 & n_2 & \cdots & n_a \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_a & 0 & \cdots & 0 & n_a \end{bmatrix} \begin{bmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \vdots \\ \alpha_a^0 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{bmatrix} \quad (66)$$

where  $a$  is the number of groups or classes, corresponding to the three types of plant in Table 15.1, where

$$N = \sum_{i=1}^a n_i, y_{i.} = \sum_{j=1}^{n_i} y_{ij} \text{ and } \bar{y}_{i.} = y_{i.}/n_i \text{ and } y_{..} = \sum_{i=1}^a y_{i.}$$

The solution to the normal equations corresponding to (17) is

$$\mu^0 = 0 \text{ and } \alpha_i^0 = \bar{y}_{i.} \quad (67)$$

The basic estimable function from (54) is  $\mu + \alpha_i$ , and from (52) its b.l.u.e. is

$$\widehat{\mu + \alpha_i} = \mu^0 + \alpha_i^0 = 0 + \bar{y}_{i.} = \bar{y}_{i.} \quad (68)$$

Any linear combination of the  $(\mu + \alpha_i)$ -expressions is, from (55), also estimable, with its b.l.u.e. being that same linear combination of the  $\bar{y}_{i.}$ 's.

Of many features of this model (as considered, for example, in Searle, 1971, Section 6.2), only two will be mentioned, and but briefly. First, from (68) and the sentence that follows it,  $\mu$  is not estimable. In contrast, two other commonly used functions, which are estimable and could each be taken as definitions of some kind of overall mean, are functions  $A$  and  $B$  shown in Table 15.6. The function  $A$  is the simple mean of means, and  $B$  is the weighted mean using the numbers of observations in the group as weights.  $A$  and  $B$  are both special cases of  $C$ , a general weighted mean using weights  $\lambda_i/\sum \lambda_i$  (see footnote of Table 15.6). Each function is followed by an example.

TABLE 15.6. ESTIMABLE FUNCTIONS IN THE 1-WAY CLASSIFICATION THAT COULD BE USED FOR ESTIMATING AN OVERALL MEAN

Function <sup>a</sup>	Estimable Function <sup>b</sup>	b.l.u.e. of q'b	Variance of b.l.u.e.
	q'b	$\widehat{q'b} = q'b^0$	$v(q'b^0) = q'Cqq^0$
A.	$\mu + \frac{1}{a}\sum \alpha_i$	$\frac{1}{a}\sum \bar{y}_{i.}$	$\frac{1}{a}\sum \sigma^2/n_i$
e.g.	$\mu + \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3)$	$\frac{1}{3}(100 + 86 + 32) = 72\frac{2}{3}$	$\frac{1}{3}(\frac{1}{3} + \frac{1}{2} + \frac{1}{1})\sigma^2 = \frac{11}{6}\sigma^2$
B.	$\mu + \frac{1}{N}\sum n_i \alpha_i$	$\frac{1}{N}\sum n_i \bar{y}_{i.} = \bar{y}_{..}$	$\frac{1}{N}\sigma^2$
e.g.	$\mu + \frac{1}{8}(3\alpha_1 + 2\alpha_2 + \alpha_3)$	$\frac{1}{8}(504) = 63$	$\frac{1}{8}\sigma^2$
C.	$\mu + \sum_{i=1}^a \lambda_i \alpha_i / \sum_{i=1}^a \lambda_i$	$(\sum \lambda_i \bar{y}_{i.}) / \sum \lambda_i$	$(\sum \lambda_i^2/n_i)\sigma^2 / (\sum \lambda_i)^2$
e.g.	$\mu + \frac{(\lambda_1 \alpha_1 + \lambda_2 \alpha_2 + \lambda_3 \alpha_3)}{\lambda_1 + \lambda_2 + \lambda_3}$	$\frac{100\lambda_1 + 86\lambda_2 + 32\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$	$\frac{\frac{1}{3}\lambda_1^2 + \frac{1}{2}\lambda_2^2 + \lambda_3^2}{(\lambda_1 + \lambda_2 + \lambda_3)^2}\sigma^2$

<sup>a</sup>A = mean of means, B = weighted mean, C = a general weighted mean, where  $\lambda_i = 1$  gives A, and  $\lambda_i = n_i$  gives B.  
<sup>b</sup>All summations are over  $i = 1, \dots, a$ .

From function  $A$  of Table 15.6 can be seen the effect of using the "convenient restraint"  $\alpha_1 + \alpha_2 + \alpha_3 = 0$  as part of the model:  $\mu$  then becomes estimable with b.l.u.e.  $(1/a)\sum \bar{y}_i$ , the mean of sample means. But if a different "convenient restraint"  $\sum n_i \alpha_i = 0$  is used, as in  $B$ , then again  $\mu$  becomes estimable, but with a different b.l.u.e., namely  $\bar{y}$ , the average of all observations. This apparent occurrence of two different b.l.u.e.'s for  $\mu$  arises because  $\mu$  in the two cases is not the same: it is  $\mu$  in two different models. In the first case it is a model that includes  $\sum \alpha_i = 0$ , and in the second it is a different model, one that includes  $\sum n_i \alpha_i = 0$ .

A second feature of the model is its analysis of variance. Using (66) and (67) in (34) gives (44) as

$$R(\mu, \alpha) = SSR = 0(y_{...}) + \sum_{i=1}^a \bar{y}_i \cdot y_i = \sum_{i=1}^a n_i \bar{y}_i^2 = \sum_{i=1}^a y_i^2 / n_i \quad (69)$$

and, as usual,  $R(\mu) = N\bar{y}^2$  of (45). Using these and  $SSR_m$  of (46) in the first part of Table 15.5 gives the analysis of variance shown in Table 15.7. Part (b) of Table 15.5 is an example of Table 15.7. The  $F$ -statistic  $F(\alpha | \mu)$  tests the hypothesis  $H: \alpha_1 = \alpha_2 = \dots = \alpha_a$ . Full description of this analysis of variance and of other details of this model are to be found in Searle (1971, Section 7.2).

TABLE 15.7. ANALYSIS OF VARIANCE FOR THE 1-WAY CLASSIFICATION

Source of Variation	d.f.	Sum of Squares	$F$ -Statistic
Mean	1	$R(\mu) = N\bar{y}^2$	
Classes, <sup>a</sup> a.f.m.	$a - 1$	$R(\alpha   \mu) = R(\mu, \alpha) - N\bar{y}^2$	$F(\alpha   \mu) = \frac{R(\alpha   \mu) / (a - 1)}{SSE / (N - a)}$
Error	$N - a$	$SSE = y'y - R(\mu, \alpha)$	
Total	$N$	$SST = y'y = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2$	

<sup>a</sup>a.f.m. = adjusted for mean.

**b. Two-way classification, no interactions, balanced data**

Countless experiments are undertaken each year in agriculture and the plant sciences to investigate the effect on growth and yield of various fertilizer treatments applied to different varieties of a species. Suppose we have data from six plants, representing three varieties being tested in

TABLE 15.8. YIELDS OF PLANTS

Variety	Treatment		Total
	1	2	
1	$y_{11}$	$y_{12}$	$y_{1\cdot}$
2	$y_{21}$	$y_{22}$	$y_{2\cdot}$
3	$y_{31}$	$y_{32}$	$y_{3\cdot}$
Total	$y_{\cdot 1}$	$y_{\cdot 2}$	$y_{\cdot \cdot}$

combination with two fertilizer treatments. Although the experiment would not necessarily be conducted by growing the plants in two rows of three plants each, it is convenient to visualize the data as in Table 15.8. The entries in the table are such that  $y_{ij}$  represents the yield of the plant of variety  $i$  that received fertilizer treatment  $j$ . If  $\mu$  is a general mean for plant yield, if  $\alpha_i$  represents the effect on yield due to variety  $i$  and  $\beta_j$  the effect due to treatment  $j$ , the equation of the linear model for  $y_{ij}$  is

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (70)$$

where  $e_{ij}$  is a random error term peculiar to  $y_{ij}$ . This is the model for the 2-way classification without interaction. It is a straightforward extension of the model given in (1) for the 1-way classification. As before, it is assumed that the  $e$ 's are independent with zero means and variance  $\sigma^2$ .

On writing each observation in terms of (70) we get

$$\begin{aligned} y_{11} &= \mu + \alpha_1 && + \beta_1 && + e_{11} \\ y_{12} &= \mu + \alpha_1 && + \beta_2 && + e_{12} \\ y_{21} &= \mu && + \alpha_2 && + \beta_1 && + e_{21} \\ y_{22} &= \mu && + \alpha_2 && + \beta_2 && + e_{22} \\ y_{31} &= \mu && + \alpha_3 && + \beta_1 && + e_{31} \\ y_{32} &= \mu && + \alpha_3 && + \beta_2 && + e_{32} \end{aligned}$$

which can easily be written in matrix form as

$$y = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} + e \quad (71)$$

where  $y$  and  $e$  are the vectors of observations and error terms, respectively. With  $X$  representing the matrix of 0's and 1's, and  $b$  the vector of parameters, this equation is the familiar form  $y = Xb + e$ . Accordingly we find

$$X'y = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ y_{.1} \\ y_{.2} \end{bmatrix}, \text{ and } X'X = \begin{bmatrix} 6 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 0 \\ 3 & 1 & 1 & 1 & 0 & 3 \end{bmatrix}$$

for which a generalized inverse of  $X'X$  is

$$G = \frac{1}{6} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 1 & 1 & -2 & 0 \\ 0 & 1 & 4 & 1 & -2 & 0 \\ 0 & 1 & 1 & 4 & -2 & 0 \\ 0 & -2 & -2 & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

With  $w' = [w_0 \ w_1 \ w_2 \ w_3 \ w_4 \ w_5]$ , we find, from (50) and (51) after calculating  $H = GX'X$ , that for any values of the  $w$ 's estimable functions are

$$q'b = w'Hb = (w_1 + w_2 + w_3)\mu + w_1\alpha_1 + w_2\alpha_2 + w_3\alpha_3 + w_4\beta_1 + (w_1 + w_2 + w_3 - w_4)\beta_2 \tag{72}$$

and their b.l.u.e.'s are, from (52),

$$q'b^0 = w'GX'y \tag{73}$$

$$= w_1(4y_{1.} + y_{2.} + y_{3.} - 2y_{.1})/6 + w_2(y_{1.} + 4y_{2.} + y_{3.} - 2y_{.1})/6$$

$$+ w_3(y_{1.} + y_{2.} + 4y_{3.} - 2y_{.1})/6 - w_4(2y_{1.} + 2y_{2.} + 2y_{3.} - 4y_{.1})/6.$$

The general estimable function (72) provides particular estimable functions, by giving special values to the  $w$ 's. It can also be used to see if particular functions of interest are estimable (by ascertaining whether or not appropriate values can be chosen for the  $w$ 's). For example,  $\alpha_1 - \alpha_2$  is seen

to be estimable, because by putting  $w_1 = 1, w_2 = -1$ , and  $w_3 = 0 = w_4$  in (72) it reduces to  $\alpha_1 - \alpha_2$ . The same values in (73) give the b.l.u.e. of  $\alpha_1 - \alpha_2$ . On the other hand,  $\mu + \alpha_1$  is not estimable; no values can be chosen for the  $w$ 's that reduce (72) to  $\mu + \alpha_1$ .

There is, of course, an infinite number of sets of  $w$ 's that could be used in (72), and so there is an infinite number of estimable functions. But each is a function  $w'Hb$  for some  $w$ . Since each  $w$  is arbitrary of order  $p$  (the number of elements in  $b$ ), we can always find  $p$  linearly independent values of  $w$ , represented by  $W$ , say. Then  $W^{-1}$  exists, and the corresponding estimable functions are  $WHb$ . But  $r_{WH} = r_H = r_X$ . Therefore there are no more than  $r_X$  linearly independent estimable functions in  $WHb$ . Hence no matter how many particular estimable functions we generate from (72) by using a different set of values for the  $w$ 's, no more than  $r_X$  (in this case 4) of them will be linearly independent. Table 15.9 shows such a set of linearly independent estimable functions, and the b.l.u.e. of each. These are all familiar results; nevertheless, it is interesting to demonstrate their derivation from the general equations (51) and (52), showing again that no "convenient restraints" such as  $\alpha_1 + \alpha_2 + \alpha_3 = 0$  or  $\beta_1 + \beta_2 = 0$  are needed to obtain them.

The estimable functions in Table 15.9 could also be derived from (54) and (55), as was done at (68). Since from (70), the expected value of  $y_{ij}$  is

$$E(y_{ij}) = \mu + \alpha_i + \beta_j,$$

any linear combination of the expressions  $\mu + \alpha_i + \beta_j$  is estimable. For example,

$$\mu + \alpha_1 + \beta_1 - (\mu + \alpha_2 + \beta_1) = \alpha_1 - \alpha_2$$

is estimable; and, of course, its b.l.u.e. is  $\widehat{\alpha_1 - \alpha_2} = \alpha_1^0 - \alpha_2^0$ , which, from calculating  $b^0 = GX'y$ , is found to be  $\bar{y}_{1.} - \bar{y}_{2.}$ .

TABLE 15.9. ESTIMATORS OF ESTIMABLE FUNCTIONS (IN THE 2-WAY CLASSIFICATION, NO INTERACTION MODEL, BALANCED DATA)

Values of $w$ 's				Estimable Function, $q'b$ Equation (72)	b.l.u.e., $q'b^0$ Equation (73) <sup>a</sup>
$w_1$	$w_2$	$w_3$	$w_4$		
1	-1	0	0	$\alpha_1 - \alpha_2$	$\bar{y}_{1.} - \bar{y}_{2.}$
1	0	-1	0	$\alpha_1 - \alpha_3$	$\bar{y}_{1.} - \bar{y}_{3.}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\mu + \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3) + \frac{1}{3}(\beta_1 + \beta_2)$	$\bar{y}_{..}$
0	0	0	1	$\beta_1 - \beta_2$	$\bar{y}_{.1} - \bar{y}_{.2}$

<sup>a</sup> $\bar{y}_{.1} - \bar{y}_{.2} \equiv -(2y_{1.} + 2y_{2.} + 2y_{3.} - 4y_{.1})/6.$

Once again it is appropriate to reiterate what has been said before in these pages. Although these matrix procedures may at first appear cumbersome when applied to the simple, familiar cases just considered, it is the universality of their application to situations not so familiar that commands our attention. With balanced data (the same number of observations in each of the smallest subclassifications of the data), the customary procedures used for computing parameter estimates are suitable *solely* because the matrix  $X'X$  has a simple form in such cases. As a result, the matrix methods that have been described lead to these customary procedures. But when data are not balanced, as is so often the case with experimental and field survey data, the expressions for estimates of functions of parameters are not at all standard or well known, and they depend a great deal on the nature of the data available. The matrix methods are then a means of deriving the estimable functions and their estimators. We illustrate this with another example.

**c. Two-way classification, no interactions, unbalanced data**

Suppose in the previous example there had been two plants of variety 1 receiving treatment 1, and none of variety 3 receiving treatment 2. The numbers of plants and total yields might then be as shown in Table 15.10. The yields are hypothetical and are solely for purposes of illustration; and they have been chosen partly with an eye to easy arithmetic.

The model is essentially the same as before, as are the vectors  $b$  and  $X'y$ . For the normal equations,  $X'X$  and a generalized inverse of it are:

$$X'X = \begin{bmatrix} 6 & 3 & 2 & 1 & 4 & 2 \\ 3 & 3 & 0 & 0 & 2 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 4 & 2 & 1 & 1 & 4 & 0 \\ 2 & 1 & 1 & 0 & 0 & 2 \end{bmatrix}$$

$$G = \frac{1}{7} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 2 & 4 & -4 & 0 \\ 0 & 2 & 5 & 3 & -3 & 0 \\ 0 & 4 & 3 & 13 & -6 & 0 \\ 0 & -4 & -3 & -6 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (74)$$

$H = GX'X$  turns out to be the same as previously, and so the estimable functions are the same as in (72). But the b.l.u.e.'s of these estimable

TABLE 15.10. NUMBERS OF PLANTS AND THEIR YIELDS

Variety	Number of Plants		Yield of Plants		Total	
	Treatment		Treatment			
	1	2	1	2		
1	2	1	3	3, 7	8	18 = $y_1$
2	1	1	2	34	2	36 = $y_2$
3	1	—	1	12	—	12 = $y_3$
Total	4	2	6	56 = $y_{.1}$	10 = $y_{.2}$	66 = $y_{..}$

functions,  $q'b^0 = w'GX'y$  as in (73), are now

$$q'b^0 = w'G[y_1, y_1, y_2, y_3, y_{.1}, y_{.2}]' \quad (75)$$

$$= w_1(5y_1 + 2y_2 + 4y_3 - 4y_{.1})/7 + w_2(2y_1 + 5y_2 + 3y_3 - 3y_{.1})/7$$

$$+ w_3(4y_1 + 3y_2 + 13y_3 - 6y_{.1})/7 - w_4(4y_1 + 3y_2 + 6y_3 - 6y_{.1})/7$$

As a function of  $y_1, y_2, y_3,$  and  $y_{.1}$ , this expression for  $q'b^0$  is quite different from its counterpart (73) in the balanced data case, even though

TABLE 15.11. ESTIMATORS OF THE FOUR EXAMPLES OF ESTIMABLE FUNCTIONS SHOWN IN TABLE 15.9

Example	Estimable function	Estimator	
		Balanced Data (Table 15.8)	Unbalanced Data (Table 15.10)
		As in Table 15.9, from equation (73)	Equation (75)
1	$\alpha_1 - \alpha_2$	$\bar{y}_1 - \bar{y}_2$	$(3y_1 - 3y_2 + y_3 - y_{.1})/7$
2	$\alpha_1 - \alpha_3$	$\bar{y}_1 - \bar{y}_3$	$(y_1 - y_2 - 9y_3 + 2y_{.1})/7$
3	$\mu + \bar{\alpha} + \bar{\beta}^a$	$\bar{y}_{..}$	$(10y_1 + 11y_2 + 22y_3 - 8y_{.1})/42$
4	$\beta_1 - \beta_2$	$\bar{y}_{.1} - \bar{y}_{.2}$	$-(4y_1 + 3y_2 + 6y_3 - 6y_{.1})/7$

<sup>a</sup> $\bar{\alpha} = \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3)$  and  $\bar{\beta} = \frac{1}{2}(\beta_1 + \beta_2)$ .



the expression for the estimable function is the same, namely (72). This emphasizes how even a minor difference in the nature of the data available, such as envisaged here, can lead to quite substantial differences in the estimators. Further evidence of this is seen in Table 15.11, which shows the same estimable functions as Table 15.9, with the b.l.u.e.'s from that table, and with their b.l.u.e.'s from (75). The difference between the two sets of estimators is striking.

As an example of unbalanced data in the 2-way classification, that discussed here is about as simple as one can get. The literature abounds with more extensive examples and with description of more general difficulties with this model that are beyond the scope of this book. For further reading, a reference such as Searle, Speed and Henderson (1981), and the sources it uses, provides a starting point.

15. THE  $R(\cdot | \cdot)$  NOTATION (continued)

We now extend the ideas implicit in the sum of squares

$$R(\alpha | \mu) = R(\mu, \alpha) - R(\mu)$$

introduced in (46) and used in the analysis of variance table of Table 15.7. This, it will be recalled, is the sum of squares due to fitting the model  $E(y_{ij}) = \mu + \alpha_i$  over and above that due to fitting  $E(y_{ij}) = \mu$ . The extension is simple. First, in parallel with  $R(\alpha | \mu)$ , we have

$$R(\beta | \mu) = R(\mu, \beta) - R(\mu),$$

which is the additional sum of squares due to fitting  $E(y_{ij}) = \mu + \beta_j$  after fitting  $E(y_{ij}) = \mu$ . Just like (69), for  $b$  levels of the  $\beta$ -factor,  $j = 1, \dots, b$

$$R(\mu, \beta) = \sum_{j=1}^b n_{.j} \bar{y}_{.j}^2 = \sum_{j=1}^b y_{.j}^2 / n_{.j} \tag{76}$$

Second, we have

$$R(\alpha | \mu, \beta) = R(\mu, \alpha, \beta) - R(\mu, \beta),$$

the sum of squares due to  $E(y_{ij}) = \mu + \alpha_i + \beta_j$  after fitting  $E(y_{ij}) = \mu + \beta_j$

and in parallel with this there is

$$R(\beta | \mu, \alpha) = R(\mu, \alpha, \beta) - R(\mu, \alpha),$$

the sum of squares due to  $E(y_{ij}) = \mu + \alpha_i + \beta_j$  after fitting  $E(y_{ij}) = \mu + \alpha_i$ . Since, for the model  $E(y_{ij}) = \mu + \alpha_i + \beta_j$ , the error sum of squares is

$$SSE = SST - R(\mu, \alpha, \beta)$$

with SST being the total sum of squares  $y'y$ , these differences between sums of squares can be summarized as partitionings of SST, as shown in Table 15.12. It is similar to, but an extension of, Table 15.7. In calculating these differences,  $R(\mu, \alpha)$  and  $R(\mu, \beta)$  come from (69) and (76), respectively, and  $R(\mu, \alpha, \beta)$  is the reduction in sum of squares for fitting the model

$$E(y_{ij}) = \mu + \alpha_i + \beta_j, \tag{77}$$

i.e., it is the SSR for this model. By (34) it is therefore

$$R(\mu, \alpha, \beta) = \mathbf{b}^0' \mathbf{X}' \mathbf{y}, \text{ for (77) written as } E(\mathbf{y}) = \mathbf{X} \mathbf{b}. \tag{78}$$

TABLE 15.12. TWO PARTITIONINGS OF TOTAL SUM OF SQUARES IN THE TWO-WAY CLASSIFICATION WITHOUT INTERACTION

Fitting $\mu + \alpha_i$ before $\mu + \alpha_i + \beta_j$	Fitting $\mu + \beta_j$ before $\mu + \alpha_i + \beta_j$
$R(\mu) = N\bar{y}^2$	$R(\mu) = N\bar{y}^2$
$R(\alpha   \mu) = R(\mu, \alpha) - R(\mu)$	$R(\beta   \mu) = R(\mu, \beta) - R(\mu)$
$R(\beta   \mu, \alpha) = R(\mu, \alpha, \beta) - R(\mu, \alpha)$	$R(\alpha   \mu, \beta) = R(\mu, \alpha, \beta) - R(\mu, \beta)$
SSE = $y'y - R(\mu, \alpha, \beta)$	SSE = $y'y - R(\mu, \alpha, \beta)$
SST = $y'y$	SST = $y'y$

**Example.** For the data of Table 15.10

$$\begin{aligned} R(\mu) &= 6(11^2) &&= 726 \\ R(\mu, \alpha) &= 18^2/3 + 36^2/2 + 12^2/1 &&= 900 \\ R(\mu, \beta) &= 56^2/4 + 10^2/2 &&= 834 \\ SST &= 3^2 + 7^2 + 8^2 + 34^2 + 2^2 + 12^2 &&= 1426. \end{aligned}$$

Then, using  $\mathbf{G}$  of (74) and

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} y_{..} \\ y_{.1} \\ y_{.2} \\ y_{.3} \\ y_{.1} \\ y_{.2} \end{bmatrix} = \begin{bmatrix} 66 \\ 18 \\ 36 \\ 12 \\ 56 \\ 10 \end{bmatrix} \quad \text{gives} \quad \mathbf{b}^0 = \mathbf{GX}'\mathbf{y} = \begin{bmatrix} 0 \\ -2 \\ 12 \\ 0 \\ 12 \\ 0 \end{bmatrix}$$

so that from (78)

$$R(\mu, \alpha, \beta) = 0(66) + (-2)(18) + 12(36) + 0(12) + 12(56) + 0(10) \\ = 1068.$$

TABLE 15.13. DEGREES OF FREEDOM AND SUMS OF SQUARES FOR ANALYSIS OF VARIANCE TABLES CORRESPONDING TO TABLE 15.12

Fitting $\mu + \alpha_i$ before $\mu + \alpha_i + \beta_j$		Fitting $\mu + \beta_j$ before $\mu + \alpha_i + \beta_j$	
d.f. <sup>a</sup>	Sum of Squares	d.f. <sup>a</sup>	Sum of Squares
1 = 1	$R(\mu) = 726$	1 = 1	$R(\mu) = 726$
$a - 1 = 2$	$R(\alpha   \mu) = 174$	$b - 1 = 1$	$R(\beta   \mu) = 108$
$b - 1 = 1$	$R(\beta   \mu, \alpha) = 168$	$a - 1 = 2$	$R(\alpha   \mu, \beta) = 234$
$N' = 2$	$SSE = 358$	$N' = 2$	$SSE = 358$
$N = 6$	$SST = 1426$	$N = 6$	$SST = 1426$

$$^a N' = N - a - b + 1 = 6 - 2 - 3 + 1 = 2.$$

The sums of squares of Table 15.12 are shown in Table 15.13. Alongside each is shown the corresponding degrees of freedom used when normality is assumed, and when  $F$ -statistics are calculated such as

$$F(\alpha | \mu) = \frac{R(\alpha | \mu) / (a - 1)}{SSE / N'} = \frac{174/2}{358/2} = \frac{87}{179}$$

and

$$F(\beta | \mu, \alpha) = \frac{R(\beta | \mu, \alpha) / (b - 1)}{SSE / N'} = \frac{168/2}{358/2} = \frac{84}{179}$$

Much has been written about  $F$ -statistics such as these, and this book is not the place for any extensive discussion. That can be found, for example, in Searle (1971, Chapter 7). It suffices to say, as is shown there (p. 283), that

$$F(\alpha | \mu) \text{ does not test } H: \alpha_i \text{'s all equal.} \quad (79)$$

It tests

$$H: \left( \alpha_i + \sum_{j=1}^b n_{ij} \beta_j / n_i \right) \text{ equal for all } i,$$

which in the case of this example, is

$$H: \alpha_1 + \frac{2}{3}\beta_1 + \frac{1}{3}\beta_2 = \alpha_2 + \frac{1}{2}\beta_1 + \frac{1}{2}\beta_2 = \alpha_3 + \beta_1. \quad (80)$$

In contrast,

$$F(\beta | \mu, \alpha) \text{ tests } H: \beta_j \text{'s all equal,}$$

and, correspondingly,

$$F(\alpha | \mu, \beta) \text{ tests } H: \alpha_i \text{'s all equal.}$$

The existence of distinctly different partitionings of the total sum of squares as shown in Table 15.12 and illustrated in Table 15.13 arises solely from there being unequal numbers of observations in the subclasses of the data. When there is the same number in every subclass, as in Table 15.8, both partitionings reduce, algebraically, to the familiar forms shown in Table 15.14 (wherein all double summations are over  $i = 1, \dots, a = 3$  and  $j = 1, \dots, b = 2$ ). For example, with balanced data, both  $R(\alpha | \mu)$  and  $R(\alpha | \mu, \beta)$  simplify to be  $\sum \sum (\bar{y}_{i.} - \bar{y}_{..})^2$ . This simplification is due *solely* to the form of the various  $\mathbf{X}$ -matrices involved when calculating each reduction in sum of squares as  $SSR = \mathbf{b}^0' \mathbf{X}' \mathbf{y} = \mathbf{y}' \mathbf{X} \mathbf{G} \mathbf{X}' \mathbf{y}$  of (34) when data are balanced.

TABLE 15.14. ANALYSIS OF VARIANCE FOR BALANCED DATA

Source of Variation	d.f.	Sum of Squares
Mean	1	$N\bar{y}_{..}^2$
Rows ( $\alpha$ 's)	2	$\sum \sum (\bar{y}_{i.} - \bar{y}_{..})^2$
Columns ( $\beta$ 's)	1	$\sum \sum (\bar{y}_{.j} - \bar{y}_{..})^2$
Error	2	$\sum \sum (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$
Total	6	$\sum \sum y_{ij}^2$

The matrix development presented here is always appropriate for unbalanced data (even in the presence of covariates—see Searle, 1971, Section 8.1). It also applies to balanced data (in that, for example, both parts of Table 15.13 simplify to the same thing, namely Table 15.14). There is therefore some convenience in thinking of the analysis of balanced data as being a special case of unbalanced data and not vice versa.

The whole subject of analyzing unbalanced data has been touched on here in terms of just the simplest of examples. But this is an effective introduction to the general ideas involved. They can readily be extended to more complex situations, but to do so here would be beyond the scope of this book. The current trend is to eliminate many of the complexities of these over-parameterized (more parameters to be estimated than there are means to estimate them from) models by returning to the basic model as, for example,  $E(y_{ijk}) = \mu_{ij}$ . This is called the  $\mu_{ij}$ -model, or cell means model—but it is another subject altogether.

## REFERENCES

- Aitken, A. C. (1948). *Determinants and Matrices*, 5th ed. Oliver and Boyd, Edinburgh.
- Albert, A. (1972). *Regression and the Moore-Penrose Inverse*. Academic Press, New York.
- Alling, David W. (1958). The after-history of pulmonary tuberculosis: A stochastic model. *Biometrics*, **14**, 527–547.
- Anderson, V. L. and Kempthorne, O. (1954). A model for the study of quantitative inheritance. *Genetics*, **39**, 883–898.
- Arnold, S. (1981). *The Theory of Linear Models and Multivariate Analysis*. John Wiley and Sons, New York.
- Ben-Israel, A. and Greville, T. N. E. (1974). *Generalized Inverses: Theory and Applications*. John Wiley and Sons, New York.
- Bernadelli, H. (1941). Population waves. *J. Burma Res. Soc.*, **31**, 1–18.
- Deemer, W. L. and Olkin, I. (1951). The Jacobians of certain matrix transformations useful in multivariate analysis. Based on lectures of P. L. Hsu at the University of North Carolina, 1947. *Biometrika*, **38**, 345–367.
- Draper, N. and Smith, H. (1981). *Applied Regression Analysis*, 2nd ed. John Wiley and Sons, New York.
- Dwyer, P. S. (1967). Some applications of matrix derivations in multivariate analysis. *J. Amer. Stat. Assoc.*, **62**, 607–625.
- Falconer, D. S. (1981). *Introduction to Quantitative Genetics*, 2nd ed. Longman, New York.
- Federer, W. T. (1955). *Experimental Design*. Macmillan, New York.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed. John Wiley and Sons, New York.
- Ferrar, W. L. (1941). *Algebra, A Text-Book of Determinants, Matrices and Algebraic Forms*. Oxford University Press, Oxford.
- Fisher, R. A. (1949). *The Theory of Inbreeding*. Oliver and Boyd, Edinburgh.
- Fisher, R. A. (1951). *The Design of Experiments*, 6th ed. Oliver and Boyd, Edinburgh.
- Frazer, R. A., Duncan, W. J. and Collar, A. R. (1938). *Elementary Matrices*. Cambridge University Press, Cambridge.
- Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, Vol. 1. McGraw-Hill, New York.

- Graybill, F. A. (1976). *Theory and Applications of the Linear Model*. Duxbury Press, North Scituate, Massachusetts.
- Hayman, B. I. and Mather, K. (1955). The description of gene interaction in continuous variation. *Biometrics*, **11**, 69–82.
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, **28**, 476–490.
- Henderson, H. V., Pukelsheim, F. and Searle, S. R. (1981). On the history of the Kronecker product. Paper No. BU-733-M, Biometrics Unit, Cornell University.
- Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Can. J. Stat.*, **7**, 65–81.
- Henderson, H. V. and Searle, S. R. (1981a). On deriving the inverse of a sum of matrices. *SIAM Review*, **23**, 53–60.
- Henderson, H. V. and Searle, S. R. (1981b). The vec-permutation matrix, the vec operator and Kronecker products: A review. *Linear and Multilinear Algebra*, **9**, 271–288.
- Kempthorne, O. (1952). *The Design and Analysis of Experiments*. John Wiley and Sons, New York.
- Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. John Wiley and Sons, New York.
- Kendall, M. G. and Stuart, A. (1974). *The Advanced Theory of Statistics*. Charles Griffin and Co., London.
- Kurkjian, B. and Zelen, M. (1962). A calculus for factorial arrangements. *Ann. Math. Stat.*, **33**, 600–619.
- Lemenager, R. P., Nelson, L. A. and Hendrix, K. S. (1980). Influence of cow size and breed type on energy requirements. *J. Animal Sci.*, **51**, 566–576.
- Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183–212.
- Leslie, P. H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika*, **35**, 213–245.
- Li, C. C. (1955). *Population Genetics*. University of Chicago Press, Chicago.
- Marsaglia, G. and Styan, G. P. H. (1974a). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra*, **2**, 269–292.
- Marsaglia, G. and Styan, G. P. H. (1974b). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā*, **36**, 437–442.
- Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.
- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, **26**, 394–395.
- Nashed, M. Zuhair, Ed. (1976). *Generalized Inverses and Applications*. Proceedings of an Advanced Seminar, University of Wisconsin. Academic Press, New York.
- Nel, D. G. (1980). On matrix differentiation in statistics. *South African J. Stat.*, **14**, 137–193.
- Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Amer. Stat. Assoc.*, **64**, 953–963.
- Newcombe, R. W. (1960). On the simultaneous diagonalization of two semidefinite matrices. *Q. Appl. Math.*, **19**, 144–146.
- Ouellette, D. V. (1981). Schur complements and statistics. *Linear Algebra and Its Applications*, **36**, 187–195.
- Penrose, R. A. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.*, **51**, 406–413.

- Pielou, E. C. (1964). The spatial pattern of two-phase patchworks of vegetation. *Biometrics*, **20**, 156–167.
- Pringle, R. M. and Rayner, A. A. (1971). *Generalized Inverse Matrices with Applications to Statistics*. Charles Griffin and Co., London.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. John Wiley and Sons, New York.
- Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. John Wiley and Sons, New York.
- Robson, D. S. (1959). A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, **15**, 187–191.
- Scheffé, H. (1959). *The Analysis of Variance*. John Wiley and Sons, New York.
- Searle, S. R. (1963). The efficiency of ancestor records in animal breeding. *Heredity*, **18**, 351–360.
- Searle, S. R. (1966). *Matrix Algebra for the Biological Sciences (including Applications in Statistics)*. John Wiley and Sons, New York.
- Searle, S. R. (1971). *Linear Models*. John Wiley and Sons, New York.
- Searle, S. R. and Henderson, H. V. (1979). Dispersion matrices for variance components models. *J. Amer. Stat. Assoc.*, **74**, 465–470.
- Searle, S. R., Speed, F. M. and Henderson, H. V. (1981). Some computational and model equivalences in analyses of variance of unequal-subclass-numbers data. *The Amer. Stat.*, **35**, 16–33.
- Seber, G. A. F. (1977). *Linear Regression Analysis*. John Wiley and Sons, New York.
- Seelye, C. J. (1958). Conditions for a positive definite quadratic form established by induction. *Am. Math. Monthly*, **65**, 355–356.
- Snedecor, G. W. and Cochran, W. G. (1967). *Statistical Methods*, 7th ed. Iowa State University Press, Ames.
- Steel, R. G. D. and Torrie, J. H. (1980). *Principles and Procedures of Statistics: A Biometrical Approach*. McGraw-Hill, New York.
- Stewart, G. W. (1973). *An Introduction to Matrix Computations*. Academic Press, New York.
- Turner, M. E., Monroe, R. J. and Homer, L. D. (1963). Generalized kinetic regression analysis: hypergeometric kinetics. *Biometrics*, **19**, 406–428.
- Urquhart, N. S. (1969). The nature of the lack of uniqueness of generalized inverse matrices. *SIAM Review*, **11**, 268–271.
- Van Loan, C. F. (1976). Generalizing the singular value decomposition. *SIAM J. Numer. Anal.*, **13**, 76–83.
- Winer, B. J. (1971). *Statistical Principles in Experimental Design*. McGraw-Hill, New York.
- Wolkowicz, H. and Styan, G. P. H. (1979). Extensions of Samuelson's inequality. *The Amer. Stat.*, **33**, 143–144.
- Yates, F. (1937). The design and analysis of factorial experiments. *Imp. Bur. Soil. Sci. Tech. Comm.*, **35**, 1–95.
- Zelen, M. and Federer, W. T. (1965). Application of the calculus for factorial arrangements. III. Analysis of factorials with unequal numbers of observations. *Sankhyā: The Indian Journal of Statistics, Series A*, Vol. 27, Parts 2, 3 and 4, pp. 383–400.

## Index

- Addition, 28
- Adjoint matrix, 128
- Adjugate matrix, 128
- $aI + bJ$  matrix, 322
  - inverse, 132
- Aitken equations, 357
- Aitken's integral, 340
- Associative laws of algebra, 50
- Augmented matrix, 232
  
- Basis, 179
- Bilinear form, 76
  
- Calculus results, 334
- Canonical form, 184
  - congruent, 201
  - equivalent, 190
  - orthogonal similar, 291
  - similar, 283
  - summary, 323
- Cauchy-Schwartz inequality, 59
- Cayley-Hamilton theorem, 314
- Centering matrix, 67, 350
- Characteristic equation, 274
  - factoring, 298
- Characteristic root (and vector), 274
- Chi-square distribution, 354
- Cofactor, 93, 125
- Column, 3
  - operations, 185
  - space, 180
  - vector, 15
- Commutative laws of algebra, 51
- Complementary minor, 108
  - principal, 108
- Complex conjugate, 341
  
- Computing inverses, 139
- Conformable:
  - for addition, 29
  - for multiplication, 38
  - for subtraction, 31
- Contrasts among means, 358
- Correlation matrix, 348
- Cramer's rule, 118
- Cumulants, of normal, 356
  
- Data matrix, 349
- Decompositions, 323
- Deficient (defective) matrix, 284
- Determinant, 84, 90
  - expansion, 87
  - geometry, 117
  - order, 85
  - partitioned matrix, 258
  - product, 98
  - transpose, 92
- Diagonability theorem, 305
- Diagonal, 13
  - element, 13
  - matrix, 13, 132
    - product with, 43
  - non-, 13
  - off-, 13
  - secondary, 88
  - sub-, 13
- Diagonal expansion, 106
- Diagonal form, symmetric matrix, 201
- Differential operators, vector of, 327
- Differentiation:
  - determinant, 336
  - inverse, 335
  - trace, 336

- Dimension:  
 of matrix, 13  
 of space, 179
- Direct product, 265  
 left, 266  
 right, 266
- Direct sum, 263
- Dispersion matrix, 347
- Distributive laws of algebra, 51
- Dot notation, 11
- Doubly stochastic, 83
- Effective inverse, 216
- Eigen root, 272  
 multiple, 280
- Eigenvalue, 272
- Eigenvector, 272  
 regular, 284
- Element, 4, 12  
 complex numbers, 341  
 diagonal, 13  
 leading, 13  
 non-diagonal, 13  
 off-diagonal, 13  
 sub-diagonal, 13
- Elementary operator matrix, 100
- Elementary operators, 184  
 inverses, 185  
 products, 186  
 rank of matrix, 186  
 transposes, 185
- Elementary row operations, 99  
*see also* Row operations
- Elementary vector, 65
- Equality, 31
- Equations, linear, 227
- Equivalent matrices, 187
- Factorization, 175
- Full column rank, 172
- Full rank, 172  
 factorization, 175, 194, 205  
 non-negative definite matrix, 206
- Full row rank, 171, 208
- Functions of matrices, 325
- Generalized inverse, 212  
 normalized, 216  
 other names, 216  
 partitioned matrix, 261  
 reflexive, 216  
 weak, 216
- G-inverse, 216
- Givens matrix (transformation), 72
- Graph theory, 46
- Hadamard inequality, 260
- Hadamard product, 49
- Helmert matrix, 71
- Hermitian matrix, 341
- Hessian matrix, 341
- Householder matrix, 72
- Idempotent matrix, 68, 133  
 summary, 320
- Identity matrix, 44  
 products equal to, 123
- Ill-conditioned, 146
- Illustrations, 16  
 age distributions, 133  
 of rabbits, 296  
 ancestor records, 258  
 Bernadelli's beetles, 275, 287  
 corporate income, 30  
 correlation matrix, 260  
 correlations, 60  
 games against nature, 56  
 gene effects, 158  
 genetics, 120  
 geometry, 117  
 graph theory, 46  
 inbreeding, 35, 159  
 input-output analysis, 134  
 larvae counts, 119  
 least squares, 330  
 linear difference equations, 288  
 linear models, 392  
 linear programming, 41  
 lunch servings, 28  
 Markov chain, 288  
 genetics, 57, 286  
 taxicab, 14, 41, 46, 50  
 maximum likelihood, 341  
 mean and variance, 67  
 prices and costs, 32  
 pulmonary tuberculosis, 138  
 random mating, 266  
 regression, 363  
 selection index, 331  
 statistics, 346

- see also* Statistics
- sterile cultures, 3
- sum of squares, 77
- transition probabilities, 49
- transition probability matrix, 68
- transportation problems, 115
- Inner product, 33
- Input-output analysis, 134
- Invariance, solutions to equations, 242
- Inverse matrix, 119  
 $aI + bJ$ , 132  
 algebraic simplifications, 136  
 existence of, 129  
 left, 147  
 order 2, 131  
 orthogonal, 132  
 partitioned matrix, 260  
 right, 147
- Jacobian, 338
- Jacobian matrix, 338
- Kronecker product, 265, 334
- Lambda root, 274
- Laplace expansion, 109
- Latent root (and vector), 274
- Laws of algebra:  
 associative, 50  
 commutative, 51  
 distributive, 51
- Leading element, 13
- Least squares, equations, 251, 357  
 generalized, 357  
 ordinary, 357
- Left direct product, 266
- Left inverse, 147
- LIN, 162
- Linear combination of vectors, 155
- Linear dependence, 159
- Linear difference equations, 288
- Linear equations, 227
- Linear independence, 159
- Linearly dependent vectors, 159, 160  
 inverse matrices, 164  
 partitioned matrices, 163  
 zero determinants, 164
- Linearly independent vectors, 160, 166  
 linear combinations of, 167
- Linear models (Chapter 15), 392
- Linear programming, 41
- Linear transformation, 157
- Manifold, 180
- Markov chain, *see* Illustrations
- Matrix, 3, 12  
 adjoint, 128  
 adjugate, 128  
 augmented, 232  
 centering, 67, 350  
 correlation, 348  
 data, 349  
 defective, 284  
 deficient, 284  
 diagonal, 284  
 diagonal, 13  
 dispersion, 347  
 of dominances, 58  
 doubly stochastic, 83  
 effective inverse, 216  
 elementary operator, 100  
 elements all equal, 65  
 full column rank, 208  
 full rank, 208  
 full row rank, 208  
 of functions, 325  
 generalized inverse, 212  
 g-inverse, 216  
 Givens, 72  
 Helmert, 71  
 Hermitian, 341  
 Hessian, 341  
 Householder, 72  
 idempotent, 68, 320  
 identity, 44  
 ill-conditioned, 146  
 inverse, 119  
 Jacobian, 339  
 left inverse, 147  
 lower triangular, 14  
 Moore-Penrose inverse, 212  
 nilpotent, 69, 321  
 non-negative definite (n.n.d.), 77, 309  
 non-singular, 130  
 normalized generalized inverse, 216  
 null, 31  
 orthogonal, 69, 320  
 partitioned, 24  
 pay-off, 56  
 permutation, 173

## Matrix (Continued)

- P-inverse, 216
- positive (semi) definite, 77
- powers of, 46
- real, 4
- reciprocal, 122
- reflexive generalized inverse, 216
- regret, 56
- regular, 284
- right inverse, 147
- scalar, 44
- Schur complement, 261
- singular, 130
- skew symmetric, 65
- square, 13
- sub-, 25
- sums of squares and products, 349
- symmetric, 60
- transition probability, 14, 68
- transpose, 22
- triangular, 14
- unipotent, 69, 321
- unit, 44
- unitary, 342
- upper triangular, 14
- variance-covariance, 347
- vec-permutation, 334
- weak generalized inverse, 216
- Wishart, 352
- zero, 32
- Matrix algebra, 4
  - contrasts with scalars, 52
  - uses, 2
- Minimax, 56
- Minor, 86, 106
  - complementary, 109
  - principal, 108
  - principal, 107
  - signed, 93
- Moore-Penrose inverse, 212
- Multiple (eigen) root, 280
- Multiplication, 32
  - inner product, 33
  - on left, 39
  - of matrices, 36
  - matrix-vector, 32
  - post-, 39
  - pre-, 39
  - on right, 39
- scalar, 29
- Multiplicity, 281
- Nilpotent matrix, 69, 321
- Non-diagonal, 13
- Non-linear equations, 326
- Non-negative definite, 77, 309
  - congruent canonical form, 206
  - generalized inverse, 220
  - summary, 322
- Non-singular matrix, 130
- Norm, 69
- Normal distribution, 352
  - cumulants of, 356
- Normalized generalized inverse, 216
- Normalized vector, 70
- Null:
  - matrix, 32, 43
  - space, 180
  - vector, 161
- Nullity, 180
- Off-diagonal, 13
- "One" vector, 66
- Order:
  - of determinant, 85
  - of matrix, 13
- Orthogonal, 69
  - eigenvectors, 290
  - inverse, 132
  - matrix, 69, 320
  - partitioned matrix, 257
  - similarity, 291
  - solutions of equations, 246
  - vectors, 70
  - vector space, 248
- Orthogonal complement, 249
- Orthonormal vectors, 70
- Partitioned matrix, 24, 163, 257
  - determinant, 258
  - generalized inverse, 261
  - inverse, 260
  - product of, 48
  - transpose, 26
  - into vectors, 26
- Pay-off matrix, 56
- Permutation matrix, 173
- P-inverse, 216
- Positive (semi) definite, 77

*Applied Probability and Statistics (Continued)*

- DUNN • Basic Statistics: A Primer for the Biomedical Sciences. *Second Edition*
- DUNN and CLARK • Applied Statistics: Analysis of Variance and Regression
- ELANDT-JOHNSON • Probability Models and Statistical Methods in Genetics
- ELANDT-JOHNSON and JOHNSON • Survival Models and Data Analysis
- FLEISS • Statistical Methods for Rates and Proportions. *Second Edition*
- FRANKEN, KÖNIG, ARNDT, and SCHMIDT • Queues and Point Processes
- GALAMBOS • The Asymptotic Theory of Extreme Order Statistics
- GIBBONS, OLKIN, and SOBEL • Selecting and Ordering Populations: A New Statistical Methodology
- GNANADESIKAN • Methods for Statistical Data Analysis of Multivariate Observations
- GOLDBERGER • Econometric Theory
- GOLDSTEIN and DILLON • Discrete Discriminant Analysis
- GROSS and CLARK • Survival Distributions: Reliability Applications in the Biomedical Sciences
- GROSS and HARRIS • Fundamentals of Queueing Theory
- GUPTA and PANCHAPAKESAN • Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations
- GUTTMAN, WILKS, and HUNTER • Introductory Engineering Statistics. *Third Edition*
- HAHN and SHAPIRO • Statistical Models in Engineering
- HALD • Statistical Tables and Formulas
- HALD • Statistical Theory with Engineering Applications
- HAND • Discrimination and Classification
- HARTIGAN • Clustering Algorithms
- HILDEBRAND, LAING, and ROSENTHAL • Prediction Analysis of Cross Classifications
- HOAGLIN, MOSTELLER, and TUKEY • Understanding Robust and Exploratory Data Analysis
- HOEL • Elementary Statistics. *Fourth Edition*
- HOEL and JESSEN • Basic Statistics for Business and Economics. *Third Edition*
- HOLLANDER and WOLFE • Nonparametric Statistical Methods
- IMAN and CONOVER • Modern Business Statistics
- JAGERS • Branching Processes with Biological Applications
- JESSEN • Statistical Survey Techniques
- JOHNSON and KOTZ • Distributions in Statistics
  - Discrete Distributions
  - Continuous Univariate Distributions—1
  - Continuous Univariate Distributions—2
  - Continuous Multivariate Distributions
- JOHNSON and KOTZ • Urn Models and Their Application: An Approach to Modern Discrete Probability Theory
- JOHNSON and LEONE • Statistics and Experimental Design in Engineering and the Physical Sciences, Volumes I and II. *Second Edition*
- JUDGE, HILL, GRIFFITHS, LÜTKEPOHL and LEE • Introduction to the Theory and Practice of Econometrics
- JUDGE, GRIFFITHS, HILL and LEE • The Theory and Practice of Econometrics
- KALBFLEISCH and PRENTICE • The Statistical Analysis of Failure Time Data
- KEENEY and RAIFFA • Decisions with Multiple Objectives
- LAWLESS • Statistical Models and Methods for Lifetime Data
- LEAMER • Specification Searches: Ad Hoc Inference with Nonexperimental Data
- MCNEIL • Interactive Data Analysis

- MANN, SCHAFER and SINGPURWALLA • Methods for Statistical Analysis of Reliability and Life Data  
 MARTZ and WALLER • Bayesian Reliability Analysis  
 MEYER • Data Analysis for Scientists and Engineers  
 MIKÉ and STANLEY • Statistics in Medical Research: Methods and Issues with Applications in Cancer Research  
 MILLER • Survival Analysis  
 MILLER, EFRON, BROWN, and MOSES • Biostatistics Casebook  
 MONTGOMERY and PECK • Introduction to Linear Regression Analysis  
 NELSON • Applied Life Data Analysis  
 OTNES and ENOCHSON • Applied Time Series Analysis: Volume I, Basic Techniques  
 OTNES and ENOCHSON • Digital Time Series Analysis  
 POLLOCK • The Algebra of Econometrics  
 PRENTER • Splines and Variational Methods  
 RAO and MITRA • Generalized Inverse of Matrices and Its Applications  
 RIPLEY • Spatial Statistics  
 SCHUSS • Theory and Applications of Stochastic Differential Equations  
 SEAL • Survival Probabilities: The Goal of Risk Theory  
 SEARLE • Linear Models  
 SEARLE • Matrix Algebra Useful for Statistics  
 SPRINGER • The Algebra of Random Variables  
 UPTON • The Analysis of Cross-Tabulated Data  
 WEISBERG • Applied Linear Regression  
 WHITTLE • Optimization Over Time, Volume I: Dynamic Programming and Stochastic Control  
 WHITTLE • Optimization Under Constraints  
 WILLIAMS • A Sampler on Sampling  
 WONNACOTT and WONNACOTT • Econometrics, *Second Edition*  
 WONNACOTT and WONNACOTT • Introductory Statistics, *Third Edition*  
 WONNACOTT and WONNACOTT • Introductory Statistics for Business and Economics, *Second Edition*  
 WONNACOTT and WONNACOTT • Regression: A Second Course in Statistics  
 WONNACOTT and WONNACOTT • Statistics: Discovering Its Power  
 ZELLNER • An Introduction to Bayesian Inference in Econometrics

## Tracts on Probability and Statistics

- AMBARTZUMIAN • Combinatorial Integral Geometry  
 BARDORFF-NIELSEN • Information and Exponential Families in Statistical Theory  
 BHATTACHARYA and RAO • Normal Approximation and Asymptotic Expansions  
 BIBBY and TOUTENBURG • Prediction and Improved Estimation in Linear Models  
 BILLINGSLEY • Convergence of Probability Measures  
 KELLY • Reversibility and Stochastic Networks  
 RAKTOE, HEDAYAT, and FEDERER • Factorial Designs  
 TOUTENBURG • Prior Information in Linear Models

- Principal minor, 107  
 Product, 32  
   with diagonal matrix, 43  
   existence, 38  
   Hadamard, 49  
   inner, 33  
   matrix-vector, 33  
   with null matrix, 43  
   outer, 40  
   of partitioned matrices, 48  
   with scalar, 42  
   trace of, 45  
   transpose of, 44  
   of two matrices, 36  
 Product notation, 11  
 Pseudo inverse, 216  
  
 Quadratic form, 73, 329  
   chi-square distribution, 354  
   differentiation, 329  
   distribution, 356  
   expected value, 355  
   independence, 356  
   positive definite, 77  
   sum of squares, 206  
  
 Range, 180  
 Rank, 155, 171  
   elementary operators, 186  
   inverse matrices, 172  
   non-zero eigenvalues, 292, 305  
   product matrix, 196  
 Reciprocal matrix, 122  
 Reflexive generalized inverse, 216  
 Regression (Chapter 14), 363  
 Right direct product, 266  
 Right inverse, 147  
 Row, 3  
 Row echelon form, 188  
 Row operations, 99, 142, 184  
 Row sum, 14, 116  
 Row vector, 15  
  
 Scalar, 16  
   matrix, 44  
   multiplication, 29  
 Schur complement, 261  
 Secondary diagonal, 88  
 Similar canonical form, 283  
 Simple (eigen) root, 280  
  
 Singular matrix, 130  
 Singular value decomposition, 316  
 Singular values, 317  
 Skew symmetric matrix, 65  
 Solution space, 249  
 Spanning set, 179  
 Spectral decomposition, 308  
 Square matrix, 13  
 State (probability) vector, 41, 68  
   steady, 59  
 Statistics, 346  
   Aitken equations, 357  
   Aitken's integral, 340  
   chi-square distribution, 354  
   contrasts among means, 358  
   correlation matrix, 348  
   cumulants, of normal, 356  
   data matrix, 349  
   dispersion matrix, 347  
   distribution of quadratic forms, 356  
   least squares, 251, 357  
   linear models, 392  
   Markov chain, 288  
     *see also* Illustrations  
   mean, 67  
   normal distribution, 352  
   regression, 363  
   sum of squares, 77  
   sums of squares and products matrix, 349  
   variance, 67  
   variance-covariance matrix, 347  
   Wishart matrix, 352  
 Steady state vector, 59  
 Sub-diagonal, 13  
 Sub-matrix, 25  
 Subscripts, 5  
   elements of matrix, 5  
 Subtraction, 30  
 Summation notation, 6  
   dot subscript, 11  
 Summing vector, 66  
 Sums of squares and products matrix, 349  
 Symmetric matrix, 60, 308  
   AA' and A'A, 61  
   canonical form, 201  
   congruent form, 201  
   diagonal form, 201  
   eigenvalues, 290  
   eigenvectors, 290  
   generalized inverse, 220



Symmetric matrix (*Continued*)

- inverse, 130
- product of vectors, 63
- products, 61
- of quadratic form, 75
- row and column operations, 199
- skew symmetric, 65
- spectral decomposition, 308
- sum of outer products, 64

## Trace, 27

- of product, 45
- of sum, 29

## Transformation, linear, 157

## Transition probabilities, 14

## Transpose, 22

- of product, 44
- of sum, 29

## Triangular matrix, 14

- lower-, 14
- reduction to, 96
- upper-, 14

## Unipotent matrix, 69, 321

## Unitary matrix, 342

## Unit matrix, 44

## Unit vector, 70

## Upper triangular matrix, 14

Variance-covariance matrix,  
347

## Vec (and vech) operators, 332

## Vec-permutation matrix, 334

## Vectors, 15

- age distribution, 272
- characteristic, 274
- column, 15
- differential operators, 327
- eigen, 274
- elementary, 65
- latent, 274
- linear combinations, 155
- linearly dependent, 159
- linearly independent, 160
- norm of, 69
- normalized, 70
- null, 161
- of ones, 66
- orthogonal, 70
- orthonormal, 70
- products of, 63
- row, 15
- space, 177
  - orthogonal, 248
- state, 68
  - probability, 41
  - steady state probabilities, 288
- subspace, 180
- summing, 66
- unit, 70

## Weak generalized inverse, 216

## Wishart matrix, 352

## Zehfuss product, 265

## Zero matrix, 32