Appendix A
Matrix Algebra

A.1 Introduction

This appendix gives (i) a summary of basic definitions and results in matrix algebra with comments and (ii) details of those results and proofs which are used in this book but normally not treated in undergraduate Mathematics courses. It is designed as a convenient source of reference to be used in the rest of the book. A geometrical interpretation of some of the results is also given. If the reader is unfamiliar with any of the results not proved here he should consult a text such as Graybill (1969, especially pp. 4–52, 163–196, and 222–235) or Rao (1973, pp. 1–78). For the computational aspects of matrix operations see for example Wilkinson (1965).

Definition *A matrix* **A** *is a rectangular array of numbers. If* **A** *has n rows and p columns we say it is of* order $n \times p$. *For example, n observations on p random variables are arranged in this way.*

Notation 1 We write matrix **A** of order $n \times p$ as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} = (a_{ij}), \tag{A.1.1}$$

where a_{ij} is the element in row i and column j of the matrix **A**, $i = 1, \ldots, n$; $j = 1, \ldots, p$. Sometimes, we write $(\mathbf{A})_{ij}$ for a_{ij}.

We may write the matrix **A** as $\mathbf{A}(n \times p)$ to emphasize the row and column order. In general, matrices are represented by boldface upper case letters throughout this book, e.g. **A**, **B**, **X**, **Y**, **Z**. Their elements are represented by small letters with subscripts.

Definition *The* transpose *of a matrix* **A** *is formed by interchanging the rows and columns:*

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{pmatrix}.$$

Definition *A matrix with column-order one is called a* column vector. *Thus*

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

is a column vector with n components.

In general, boldface lower case letters represent column vectors. Row vectors are written as column vectors transposed, i.e.

$$\mathbf{a}' = (a_1, \ldots, a_n).$$

Notation 2 We write the columns of the matrix **A** as $\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \ldots, \mathbf{a}_{(p)}$ and the rows (if written as column vectors) as $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ so that

$$\mathbf{A} = (\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \ldots, \mathbf{a}_{(p)}) = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_n' \end{bmatrix}, \tag{A.1.2}$$

where

$$\mathbf{a}_{(j)} = \begin{bmatrix} a_{1j} \\ \cdot \\ \cdot \\ \cdot \\ a_{nj} \end{bmatrix}, \qquad \mathbf{a}_i = \begin{bmatrix} a_{i1} \\ \cdot \\ \cdot \\ \cdot \\ a_{ip} \end{bmatrix}.$$

Definition *A matrix written in terms of its sub-matrices is called a* partitioned matrix.

Notation 3 Let $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$, and $\mathbf{A}_{22}$ be submatrices such that $\mathbf{A}_{11}(r\times s)$ has elements a_{ij}, $i=1,\dots,r$; $j=1,\dots,s$ and so on. Then we write

$$\mathbf{A}(n\times p)=\begin{bmatrix}\mathbf{A}_{11}(r\times s) & \mathbf{A}_{12}(r\times(p-s))\\ \mathbf{A}_{21}((n-r)\times s) & \mathbf{A}_{22}((n-r)\times(p-s))\end{bmatrix}.$$

Obviously, this notation can be extended to contain further partitions of $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, etc.

A list of some important types of particular matrices is given in Table A.1.1. Another list which depends on the next section appears in Table A.3.1.

Table A.1.1 Particular matrices and types of matrix (List 1). For List 2 see Table A.3.1.

Name	Definiton	Notation	Trivial Examples
1 Scalar	$p=n=1$	a, b	(1)
2*a* Column vector	$p=1$	$\mathbf{a}, \mathbf{b}, \dots$	$\begin{pmatrix}1\\2\end{pmatrix}$
2b Unit vector	$(1,\dots,1)'$	$\mathbf{1}$ or $\mathbf{1}_p$	$\begin{pmatrix}1\\1\end{pmatrix}$
3 Rectangular	$p\times n$	$\mathbf{A}(n\times p)$	
4 Square	$p=n$	$\mathbf{A}(p\times p)$	$\begin{pmatrix}1&3\\4&5\end{pmatrix}$
4a Diagonal	$p=n,\ a_{ij}=0,\ i\neq j$	$\mathrm{diag}\,(a_{ii})$	$\begin{pmatrix}2&0\\0&1\end{pmatrix}$
4b Identity	$\mathrm{diag}\,(\mathbf{1})$	$\mathbf{I}$ or $\mathbf{I}_p$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$
4c Symmetric	$a_{ij}=a_{ji}$		$\begin{pmatrix}3&2\\2&5\end{pmatrix}$
4d Unit matrix	$p=n,\ a_{ij}=1$	$\mathbf{J}_p=\mathbf{11}'$	$\begin{pmatrix}1&1\\1&1\end{pmatrix}$
4e Triangular matrix (upper)	$a_{ij}=0$ below the diagonal	Δ'	
Triangular matrix (lower)	$a_{ij}=0$ above the diagonal	Δ	$\begin{pmatrix}1&0&0\\2&2&0\\3&2&5\end{pmatrix}$
5 Asymmetric	$a_{ij}\neq a_{ji}$		$\begin{pmatrix}1&1\\2&3\end{pmatrix}$
6 Null	$a_{ij}=0$	$\mathbf{0}$	$\begin{pmatrix}0&0&0\\0&0&0\end{pmatrix}$

As shown in Table A.1.1 a square matrix $\mathbf{A}(p\times p)$ is *diagonal* if $a_{ij}=0$ for all $i\neq j$. There are two convenient ways to construct diagonal matrices. If $\mathbf{a}=(a_1,\dots,a_p)'$ is any vector and $\mathbf{B}(p\times p)$ is any square matrix then

$$\mathrm{diag}\,(\mathbf{a})=\mathrm{diag}\,(a_i)=\mathrm{diag}\,(a_1,\dots,a_p)=\begin{pmatrix}a_1&\cdots&0\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \cdot&&\cdot\\ 0&\cdots&a_p\end{pmatrix}$$

and

$$\mathrm{Diag}\,(\mathbf{B})=\begin{pmatrix}b_{11}&\cdots&0\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \cdot&&\cdot\\ 0&\cdots&b_{pp}\end{pmatrix}$$

each defines a diagonal matrix.

A.2 Matrix Operations

Table A.2.1. gives a summary of various important matrix operations. We deal with some of these in detail, assuming the definitions in the table.

Table A.2.1 Basic matrix operations

Operation	Restrictions	Definitions	Remarks
1 Addition	$\mathbf{A}, \mathbf{B}$ of the same order	$\mathbf{A}+\mathbf{B}=(a_{ij}+b_{ij})$	
2 Subtraction	$\mathbf{A}, \mathbf{B}$ of the same order	$\mathbf{A}-\mathbf{B}=(a_{ij}-b_{ij})$	
3a Scalar multiplication		$c\mathbf{A}=(ca_{ij})$	
3b Inner product	$\mathbf{a}, \mathbf{b}$ of the same order	$\mathbf{a}'\mathbf{b}=\sum a_ib_i$	
3c Multiplication	Number of columns of $\mathbf{A}$ equals number of rows of $\mathbf{B}$	$\mathbf{AB}=(\mathbf{a}_i'\mathbf{b}_{(j)})$	$\mathbf{AB}\neq\mathbf{BA}$
4 Transpose		$\mathbf{A}'=(\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_n)$	Section A.2.1.
5 Trace	$\mathbf{A}$ square	$\mathrm{tr}\,\mathbf{A}=\sum a_{ii}$	Section A.2.2.
6 Determinant	$\mathbf{A}$ square	$\lvert\mathbf{A}\rvert$	Section A.2.3.
7 Inverse	$\mathbf{A}$ square and $\lvert\mathbf{A}\rvert\neq0$	$\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$	$(\mathbf{A}+\mathbf{B})^{-1}\neq\mathbf{A}^{-1}+\mathbf{B}^{-1}$, Section A.2.4
8 g-inverse $(\mathbf{A}^-)$	$\mathbf{A}(n\times p)$	$\mathbf{AA}^-\mathbf{A}=\mathbf{A}$	Section A.8

A.2.1 Transpose

The transpose satisfies the simple properties

$$(\mathbf{A}')'=\mathbf{A},\qquad(\mathbf{A}+\mathbf{B})'=\mathbf{A}'+\mathbf{B}',\qquad(\mathbf{AB})'=\mathbf{B}'\mathbf{A}'.\qquad(\text{A.2.1})$$

For partitioned **A**,

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

If **A** is a symmetric matrix, $a_{ij} = a_{ji}$, so that

$$\mathbf{A}' = \mathbf{A}.$$

A.2.2 Trace

The trace function, $\text{tr}\,\mathbf{A} = \Sigma a_{ii}$, satisfies the following properties for $\mathbf{A}(p \times p)$, $\mathbf{B}(p \times p)$, $\mathbf{C}(p \times n)$, $\mathbf{D}(n \times p)$, and scalar α:

$$\text{tr}\,\alpha = \alpha, \qquad \text{tr}\,\mathbf{A} \pm \mathbf{B} = \text{tr}\,\mathbf{A} \pm \text{tr}\,\mathbf{B}, \qquad \text{tr}\,\alpha\mathbf{A} = \alpha\,\text{tr}\,\mathbf{A} \quad \text{(A.2.2a)}$$

$$\text{tr}\,\mathbf{CD} = \text{tr}\,\mathbf{DC} = \sum_{i,j} c_{ij} d_{ji}, \quad \text{(A.2.2b)}$$

$$\sum \mathbf{x}_i'\mathbf{A}\mathbf{x}_i = \text{tr}\,(\mathbf{AT}), \quad \text{where} \quad \mathbf{T} = \sum \mathbf{x}_i\mathbf{x}_i'. \quad \text{(A.2.2c)}$$

To prove this last property, note that since $\sum \mathbf{x}_i'\mathbf{A}\mathbf{x}_i$ is a scalar, the left-hand side of (A.2.2c) is

$$\begin{aligned} \text{tr} \sum \mathbf{x}_i'\mathbf{A}\mathbf{x}_i &= \sum \text{tr}\,\mathbf{x}_i'\mathbf{A}\mathbf{x}_i \quad \text{by (A.2.2a)} \\ &= \sum \text{tr}\,\mathbf{A}\mathbf{x}_i\mathbf{x}_i' \quad \text{by (A.2.2b)} \\ &= \text{tr}\,\mathbf{A} \sum \mathbf{x}_i\mathbf{x}_i' \quad \text{by (A.2.2a).} \end{aligned}$$

As a special case of (A.2.2b) note that

$$\text{tr}\,\mathbf{CC}' = \text{tr}\,\mathbf{C}'\mathbf{C} = \sum c_{ij}^2. \quad \text{(A.2.2d)}$$

A.2.3 Determinants and cofactors

Definition *The determinant of a square matrix* **A** *is defined as*

$$|\mathbf{A}| = \sum (-1)^{|\tau|} a_{1\tau(1)} \cdots a_{p\tau(p)}, \quad \text{(A.2.3a)}$$

where the summation is taken over all permutations τ *of* $(1, 2, \ldots, p)$, *and* $|\tau|$ *equals* $+1$ *or* -1, *depending on whether* τ *can be written as the product of an even or odd number of transpositions.*

For $p = 2$,

$$|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}. \quad \text{(A.2.3b)}$$

Definition *The* cofactor *of* a_{ij} *is defined by* $(-1)^{i+j}$ *times the minor of* a_{ij}, *where the* minor *of* a_{ij} *is the value of the determinant obtained after deleting the ith row and the jth column of* **A**.

We denote the cofactor of a_{ij} by A_{ij}. Thus for $p = 3$,

$$A_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \qquad A_{12} = -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \qquad A_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}. \quad \text{(A.2.3c)}$$

Definition *A square matrix is* non-singular *if* $|\mathbf{A}| \neq 0$; *otherwise it is singular.*

We have the following results:

(I) $|\mathbf{A}| = \sum_{j=1}^{p} a_{ij}A_{ij} = \sum_{i=1}^{p} a_{ij}A_{ij}$, any i, j, (A.2.3d)

but

$$\sum_{k=1}^{p} a_{ik}A_{jk} = 0, \qquad i \neq j. \quad \text{(A.2.3e)}$$

(II) If **A** is triangular or diagonal,

$$|\mathbf{A}| = \prod a_{ii}. \quad \text{(A.2.3f)}$$

(III) $|c\mathbf{A}| = c^p\,|\mathbf{A}|$. (A.2.3g)

(IV) $|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|$. (A.2.3h)

(V) For square submatrices $\mathbf{A}(p \times p)$ and $\mathbf{B}(q \times q)$,

$$\begin{vmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{0} & \mathbf{B} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{B}|. \quad \text{(A.2.3i)}$$

(VI)

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}|\,|\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| = |\mathbf{A}_{22}|\,|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|, \quad \text{(A.2.3j)}$$

$$\begin{vmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{a}' & b \end{vmatrix} = |\mathbf{A}|\,(b - \mathbf{a}'\,\mathbf{A}^{-1}\,\mathbf{a}).$$

(VII) For $\mathbf{B}(p \times n)$ and $\mathbf{C}(n \times p)$, and non-singular $\mathbf{A}(p \times p)$,

$$|\mathbf{A} + \mathbf{BC}| = |\mathbf{A}|\,|\mathbf{I}_p + \mathbf{A}^{-1}\mathbf{BC}| = |\mathbf{A}|\,|\mathbf{I}_n + \mathbf{CA}^{-1}\mathbf{B}|, \quad \text{(A.2.3k)}$$

$$|\mathbf{A} + \mathbf{b}'\mathbf{a}| = |\mathbf{A}|\,(1 + \mathbf{b}'\mathbf{A}^{-1}\,\mathbf{a}).$$

Remarks (1) Properties (I)–(III) follow easily from the definition (A.2.3a). As an application of (I), from (A.2.3b), (A.2.3c), and (A.2.3d), we have, for $p = 3$,

$$|\mathbf{A}| = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}).$$

(2) To prove (V), note that the only permutations giving non-zero terms in the summation (A.2.3a) are those taking $\{1, \ldots, p\}$ to $\{1, \ldots, p\}$ and $\{p+1, \ldots, p+q\}$ to $\{p+1, \ldots, p+q\}$.

(3) To prove (VI), simplify $\mathbf{BAB}'$ and then take its determinant where

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

From (VI), we deduce, after putting $\mathbf{A}^{11} = \mathbf{A}$, $\mathbf{A}^{12} = \mathbf{x}'$, etc.,

$$\begin{vmatrix} \mathbf{A} & \mathbf{x} \\ \mathbf{x} & c \end{vmatrix} = |\mathbf{A}|\,\{c - \mathbf{x}'\mathbf{A}^{-1}\mathbf{x}\}. \quad \text{(A.2.3l)}$$

(4) To prove the second part of (VII), simplify

$$\begin{vmatrix} \mathbf{I}_p & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{C} & \mathbf{I}_n \end{vmatrix}$$

using (VI). As special cases of (VII) we see that, for non-singular $\mathbf{A}$,

$$|\mathbf{A}+\mathbf{bb}'| = |\mathbf{A}|\,(1+\mathbf{b}'\mathbf{A}^{-1}\mathbf{b}), \tag{A.2.3m}$$

and that, for $\mathbf{B}(p\times n)$ and $\mathbf{C}(n\times p)$,

$$|\mathbf{I}_p+\mathbf{BC}| = |\mathbf{I}_n+\mathbf{CB}|. \tag{A.2.3n}$$

In practice, we can simplify determinants using the property that the value of a determinant is unaltered if a linear combination of some of the columns (rows) is added to another column (row).

(5) Determinants are usually evaluated on computers as follows. $\mathbf{A}$ is decomposed into upper and lower triangular matrices $\mathbf{A}=\mathbf{LU}$. If $\mathbf{A}>0$, then the Cholesky decomposition is used (i.e. $\mathbf{U}=\mathbf{L}'$ so $\mathbf{A}=\mathbf{LL}'$). Otherwise the Crout decomposition is used where the diagonal elements of $\mathbf{U}$ are ones.

A.2.4 Inverse

Definition *As already defined in Table* A.1.1, *the* inverse *of* $\mathbf{A}$ *is the unique matrix* $\mathbf{A}^{-1}$ *satisfying*

$$\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}. \tag{A.2.4a}$$

The inverse exists if and only if $\mathbf{A}$ *is non-singular, that is, if and only if* $|\mathbf{A}|\neq 0$.

We write the (i, j)th element of $\mathbf{A}^{-1}$ by a^{ij}. For partitioned $\mathbf{A}$, we write

$$\mathbf{A}^{-1}=\begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}.$$

The following properties hold:

(I) $\mathbf{A}^{-1}=\dfrac{1}{|\mathbf{A}|}(A_{ij})'$. (A.2.4b)

(II) $(c\mathbf{A})^{-1}=c^{-1}\mathbf{A}^{-1}$. (A.2.4c)

(III) $(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$. (A.2.4d)

(IV) The unique solution of $\mathbf{Ax}=\mathbf{b}$ is $\mathbf{x}=\mathbf{A}^{-1}\mathbf{b}$. (A.2.4e)

(V) If all the necessary inverses exist, then for $\mathbf{A}(p\times p)$, $\mathbf{B}(p\times n)$, $\mathbf{C}(n\times n)$, and $\mathbf{D}(n\times p)$,

$$(\mathbf{A}+\mathbf{BCD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}, \tag{A.2.4f}$$
$$(\mathbf{A}+\mathbf{a}\,\mathbf{b}')^{-1}=\mathbf{A}^{-1}-\{(\mathbf{A}^{-1}\mathbf{a})(\mathbf{b}'\mathbf{A}^{-1})(1+\mathbf{b}'\mathbf{A}^{-1}\mathbf{a})^{-1}\}.$$

(VI) If all the necessary inverses exist, then for partitioned $\mathbf{A}$, the elements of $\mathbf{A}^{-1}$ are

$$\left.\begin{aligned} \mathbf{A}^{11}&=(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}, & \mathbf{A}^{22}&=(\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1},\\ \mathbf{A}^{12}&=-\mathbf{A}^{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}, & \mathbf{A}^{21}&=-\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}^{11}. \end{aligned}\right\} \tag{A.2.4g}$$

Alternatively, $\mathbf{A}^{12}$ and $\mathbf{A}^{21}$ can be defined by

$$\mathbf{A}^{12}=-\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}^{22}, \qquad \mathbf{A}^{21}=-\mathbf{A}^{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}.$$

(VII) For symmetrical matrices $\mathbf{A}$ and $\mathbf{D}$, we have, if all necessary inverses exist

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{pmatrix}^{-1}=\begin{pmatrix} \mathbf{A}^{-1} & 0 \\ 0 & 0 \end{pmatrix}+\begin{pmatrix} -\mathbf{E} \\ \mathbf{I} \end{pmatrix}(\mathbf{D}-\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}(-\mathbf{E}',\mathbf{I})$$

where $\mathbf{E}=\mathbf{A}^{-1}\mathbf{B}$.

Remarks (1) The result (I) follows on using (A.2.3d), (A.2.3e). As a simple application, note that, for $p=2$, we have

$$\mathbf{A}^{-1}=\frac{1}{a_{11}a_{22}-a_{12}a_{21}}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

(2) Formulae (II)–(VI) can be verified by checking that the product of the matrix and its inverse reduces to the identity matrix, e.g. to verify (III), we proceed

$$(\mathbf{AB})^{-1}(\mathbf{AB})=\mathbf{B}^{-1}\mathbf{A}^{-1}(\mathbf{AB})=\mathbf{B}^{-1}\mathbf{IB}=\mathbf{I}.$$

(3) We have assumed $\mathbf{A}$ to be a square matrix with $|\mathbf{A}|\neq 0$ in defining $\mathbf{A}^{-1}$. For $\mathbf{A}(n\times p)$, a generalized inverse is defined in Section A.8.

(4) In computer algorithms for evaluating $\mathbf{A}^{-1}$, the following methods are commonly used. If $\mathbf{A}$ is symmetric, the Cholesky method is used, namely, decomposing $\mathbf{A}$ into the form $\mathbf{LL}'$ where $\mathbf{L}$ is lower triangular and then using $\mathbf{A}^{-1}=(\mathbf{L}^{-1})'\mathbf{L}^{-1}$. For non-symmetric matrices, Crout's method is used, which is a modification of Gaussian elimination.

A.2.5 Kronecker products

Definition *Let* $\mathbf{A}=(a_{ij})$ *be an* $(m\times n)$ *matrix and* $\mathbf{B}=(b_{kl})$ *be a* $(p\times q)$ *matrix. Then the* Kronecker product *of* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as*

$$\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix},$$

which is an $(mp\times nq)$ *matrix. It is denoted by* $\mathbf{A}\otimes\mathbf{B}$.

Definition *If* $\mathbf{X}$ *is an* $(n\times p)$ *matrix let* $\mathbf{X}^V$ *denote the np-vector obtained by "vectorizing"* $\mathbf{X}$; *that is, by stacking the columns of* $\mathbf{X}$ *on top of one another so that*

$$\mathbf{X}^V = \begin{bmatrix} \mathbf{x}_{(1)} \\ \mathbf{x}_{(2)} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_{(p)} \end{bmatrix}.$$

From these definitions the elementary properties given below easily follow:

(I) $\alpha(\mathbf{A}\otimes\mathbf{B}) = (\alpha\mathbf{A})\otimes\mathbf{B} = \mathbf{A}\otimes(\alpha\mathbf{B})$ for all scalar α, and hence can be written without ambiguity as $\alpha\mathbf{A}\otimes\mathbf{B}$.
(II) $\mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C}) = (\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C}$. Hence this can be written as $\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}$.
(III) $(\mathbf{A}\otimes\mathbf{B})' = \mathbf{A}'\otimes\mathbf{B}'$.
(IV) $(\mathbf{A}\otimes\mathbf{B})(\mathbf{F}\otimes\mathbf{G}) = (\mathbf{AF})\otimes(\mathbf{BG})$. Here parentheses are necessary.
(V) $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$ for non-singular $\mathbf{A}$ and $\mathbf{B}$.
(VI) $(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes\mathbf{C}+\mathbf{B}\otimes\mathbf{C}$.
(VII) $\mathbf{A}\otimes(\mathbf{B}+\mathbf{C}) = \mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}$.
(VIII) $(\mathbf{AXB})^V = (\mathbf{B}'\otimes\mathbf{A})\mathbf{X}^V$.
(IX) $\operatorname{tr}(\mathbf{A}\otimes\mathbf{B}) = (\operatorname{tr}\mathbf{A})(\operatorname{tr}\mathbf{B})$.

A.3 Further Particular Matrices and Types of Matrix

Table A.3.1 gives another list of some important types of matrices. We consider a few in more detail.

A.3.1 Orthogonal matrices

A square matrix $\mathbf{A}(n\times n)$ is *orthogonal* if $\mathbf{AA}' = \mathbf{I}$. The following properties hold:

(I) $\mathbf{A}^{-1} = \mathbf{A}'$.
(II) $\mathbf{A}'\mathbf{A} = \mathbf{I}$.
(III) $|\mathbf{A}| = \pm 1$.
(IV) $\mathbf{a}_i'\mathbf{a}_j = 0,\ i\neq j;\ \mathbf{a}_i'\mathbf{a}_i = 1.\ \mathbf{a}_{(i)}'\mathbf{a}_{(j)} = 0,\ i\neq j;\ \mathbf{a}_{(i)}'\mathbf{a}_{(i)} = 1$.
(V) $\mathbf{C} = \mathbf{AB}$ is orthogonal if $\mathbf{A}$ and $\mathbf{B}$ are orthogonal.

Remarks (1) All of these properties follow easily from the definition $\mathbf{AA}' = \mathbf{I}$. Result (IV) states that the sum of squares of the elements in each

Table A.3.1 Particular types of matrices (List 2)

Name	Definition	Examples	Details in
Non-singular	$\lvert\mathbf{A}\rvert\neq 0$	$\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}$	Section A.2.3.
Singular	$\lvert\mathbf{A}\rvert = 0$	$\begin{bmatrix}1 & 2\\ 1 & 2\end{bmatrix}$	Section A.2.3.
Orthogonal	$\mathbf{AA}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$	$\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$	Section A.3.1.
Equicorrelation	$\mathbf{E} = (1-\rho)\mathbf{I}+\rho\mathbf{J}$	$\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$	Section A.3.2.
Idempotent	$\mathbf{A}^2 = \mathbf{A}$	$\frac{1}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$	
Centring matrix, H_n	$\mathbf{H}_n = \mathbf{I}_n - n^{-1}\mathbf{J}_n$		Section A.3.3.
Positive definite (p.d.)	$\mathbf{x}'\mathbf{Ax}>0$ for all $\mathbf{x}\neq\mathbf{0}$	$x_1^2+x_2^2$	Section A.7.
Positive semi-definite (p.s.d.)	$\mathbf{x}'\mathbf{Ax}\geq 0$ for all $\mathbf{x}\neq\mathbf{0}$	$(x_1-x_2)^2$	Section A.7.

row (column) is unity whereas the sum of the cross-products of the elements in any two rows (columns) is zero.

(2) The *Helmert matrix* is a particular orthogonal matrix whose columns are defined by

$$\mathbf{a}_{(1)}' = (n^{-1/2},\ldots,n^{-1/2}),$$
$$\mathbf{a}_{(j)}' = (d_j,\ldots,d_j,-(j-1)d_j,0,\ldots,0),\qquad j=2,\ldots,n,$$

where $d_j = \{j(j-1)\}^{-1/2}$, is repeated $j-1$ times.

(3) Orthogonal matrices can be used to represent a change of basis, or rotation. See Section A.5.

A.3.2 Equicorrelation matrix

Consider the $(p\times p)$ matrix defined by

$$\mathbf{E} = (1-\rho)\mathbf{I}+\rho\mathbf{J}, \tag{A.3.2a}$$

where ρ is any real number. Then $e_{ii}=1$, $e_{ij}=\rho$, for $i\neq j$. For statistical purposes this matrix is most useful for $-(p-1)^{-1}<\rho<1$, when it is called the *equicorrelation* matrix.

Direct verification shows that, provided $\rho\neq 1, -(p-1)^{-1}$, then $\mathbf{E}^{-1}$ exists and is given by

$$\mathbf{E}^{-1} = (1-\rho)^{-1}[\mathbf{I}-\rho\{1+(p-1)\rho\}^{-1}\mathbf{J}]. \tag{A.3.2b}$$

Its determinant is given by

$$|\mathbf{E}| = (1-\rho)^{p-1}\{1+\rho(p-1)\}. \tag{A.3.2c}$$

This formula is most easily verified using the eigenvalues given in Remark 6 of Section A.6.

A.3.3 Centring matrix

The $(n \times n)$ centring matrix is defined by $\mathbf{H} = \mathbf{H}_n = \mathbf{I} - n^{-1}\mathbf{J}$, where $\mathbf{J} = \mathbf{1}\mathbf{1}'$. Then

(I) $\mathbf{H}' = \mathbf{H}$, $\mathbf{H}^2 = \mathbf{H}$.
(II) $\mathbf{H1} = \mathbf{0}$, $\mathbf{HJ} = \mathbf{JH} = \mathbf{0}$.
(III) $\mathbf{Hx} = \mathbf{x} - \bar{x}\mathbf{1}$, where $\bar{x} = n^{-1}\sum x_i$.
(IV) $\mathbf{x}'\mathbf{Hx} = n^{-1}\sum (x_i - \bar{x})^2$.

Remark (1) Property (I) states that $\mathbf{H}$ is symmetric and idempotent.

(2) Property (III) is most important in data analysis. The ith element of $\mathbf{Hx}$ is $x_i - \bar{x}$. Therefore, premultiplying a column vector by $\mathbf{H}$ has the effect of re-expressing the elements of the vector as *deviations from the mean*. Similarly, premultiplying a matrix by $\mathbf{H}$ re-expresses each element of the matrix as a *deviation from its column mean*, i.e. $\mathbf{HX}$ has its (i, j)th element $x_{ij} - \bar{x}_j$, where $\bar{x}_j$ is the mean of the jth column of $\mathbf{X}$. This "centring" property explains the nomenclature for $\mathbf{H}$.

A.4 Vector Spaces, Rank, and Linear Equations

A.4.1 Vector spaces

The set of vectors in R^n satisfies the following properties. For all $\mathbf{x}, \mathbf{y} \in R^n$ and all $\lambda, \mu \in R$,

(1) $\lambda(\mathbf{x}+\mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$,
(2) $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$,
(3) $(\lambda\mu)\mathbf{x} = \lambda(\mu\mathbf{x})$,
(4) $1\mathbf{x} = \mathbf{x}$.

Thus R^n can be considered as a *vector space* over the real numbers R.

Definition *If W is a subset of R^n such that for all $\mathbf{x}, \mathbf{y} \in W$ and $\lambda \in R$*

$$\lambda(\mathbf{x}+\mathbf{y}) \in W,$$

then W is called a vector subspace *of R^n.*

Two simple examples of subspaces of R^n are $\{\mathbf{0}\}$ and R^n itself.

Definition *Vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are called* linearly dependent *if there exist numbers $\lambda_1, \ldots, \lambda_k$, not all zero, such that*

$$\lambda_1\mathbf{x}_1 + \ldots + \lambda_k\mathbf{x}_k = \mathbf{0}.$$

Otherwise the k vectors are linearly independent.

Definition *Let W be a subspace of R^n. Then a* basis *of W is a maximal linearly independent set of vectors.*

The following properties hold for a basis of W:

(I) Every basis of W contains the same (finite) number of elements. This number is called the *dimension* of W and denoted dim W. In particular dim $R^n = n$.
(II) If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is a basis for W then every element $\mathbf{x}$ in W can be expressed as a linearly combination of $\mathbf{x}_1, \ldots, \mathbf{x}_k$; that is, $\mathbf{x} = \lambda_1\mathbf{x}_1 + \ldots + \lambda_k\mathbf{x}_k$ for some numbers $\lambda_1, \ldots, \lambda_k$.

Definition *The* inner (*or* scalar *or* dot) product *between two vectors $\mathbf{x}, \mathbf{y} \in R^n$ is defined by*

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}'\mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

The vectors $\mathbf{x}$ *and* $\mathbf{y}$ *are called* orthogonal *if* $\mathbf{x} \cdot \mathbf{y} = 0$.

Definition *The* norm *of a vector $\mathbf{x} \in R^n$ is given by*

$$\|\mathbf{x}\| = (\mathbf{x} \cdot \mathbf{x})^{1/2} = \left(\sum x_i^2\right)^{1/2}.$$

Then the distance *between two vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *is given by*

$$\|\mathbf{x} - \mathbf{y}\|.$$

Definition *A basis $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of a subspace W of R^n is called* orthonormal *if all the elements have norm 1 and are orthogonal to one another; that is, if*

$$\mathbf{x}_i'\mathbf{x}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

In particular, if $\mathbf{A}(n \times n)$ is an orthogonal matrix then the columns of $\mathbf{A}$ form an orthonormal basis of R^n.

A.4.2 Rank

Definition *The* rank *of a matrix $\mathbf{A}(n \times p)$ is defined as the maximum number of linearly independent rows (columns) in $\mathbf{A}$.*

We denote it by $r(\mathbf{A})$ or rank $(\mathbf{A})$.

The following properties hold:

(I) $0 \leq r(\mathbf{A}) \leq \min(n, p)$. (A.4.2a)

(II) $r(\mathbf{A}) = r(\mathbf{A}')$. (A.4.2b)

(III) $r(\mathbf{A}+\mathbf{B}) \leq r(\mathbf{A}) + r(\mathbf{B})$. (A.4.2c)

(IV) $r(\mathbf{AB}) \leq \min\{r(\mathbf{A}), r(\mathbf{B})\}$. (A.4.2d)

(V) $r(\mathbf{A}'\mathbf{A}) = r(\mathbf{AA}') = r(\mathbf{A})$. (A.4.2e)

(VI) If $\mathbf{B}(n\times n)$ and $\mathbf{C}(p\times p)$ are non-singular then $r(\mathbf{BAC}) = r(\mathbf{A})$. (A.4.2f)

(VII) If $n = p$ then $r(\mathbf{A}) = p$ if and only if $\mathbf{A}$ is non-singular. (A.4.2g)

Table A.4.1 gives the ranks of some particular matrices.

Remarks (1) Another definition of $r(\mathbf{A})$ is $r(\mathbf{A})$ = the largest order of those (square) submatrices which have non-vanishing determinants.

(2) If we define $M(\mathbf{A})$ as the vector subspace in R^n spanned by the columns of $\mathbf{A}$, then $r(\mathbf{A}) = \dim M(\mathbf{A})$ and we may choose linearly independent columns of $\mathbf{A}$ as a basis for $M(\mathbf{A})$. Note that for any p-vector $\mathbf{x}$, $\mathbf{Ax} = x_1\mathbf{a}_{(1)} + \ldots + x_p\mathbf{a}_{(p)}$ is a linear combination of the columns of $\mathbf{A}$ and hence $\mathbf{Ax}$ lies in $M(\mathbf{A})$.

(3) Define the *null space* of $\mathbf{A}(n\times p)$ by

$$N(\mathbf{A}) = \{\mathbf{x} \in R^p : \mathbf{Ax} = \mathbf{0}\}.$$

Then $N(\mathbf{A})$ is a vector subspace of R^p of dimension k, say. Let $\mathbf{e}_1, \ldots, \mathbf{e}_p$ be a basis of R^p for which $\mathbf{e}_1, \ldots, \mathbf{e}_k$ are a basis of $N(\mathbf{A})$. Then $\mathbf{Ae}_{k+1}, \ldots, \mathbf{Ae}_p$ form a maximally linearly independent set of vectors in $M(\mathbf{A})$, and hence are a basis for $M(\mathbf{A})$. Thus, we get the important result

$$\dim N(\mathbf{A}) + \dim M(\mathbf{A}) = p. \quad \text{(A.4.2h)}$$

(4) To prove (V) note that if $\mathbf{Ax} = \mathbf{0}$, then $\mathbf{A}'\mathbf{Ax} = \mathbf{0}$; conversely if $\mathbf{A}'\mathbf{Ax} = \mathbf{0}$ then $\mathbf{x}'\mathbf{A}'\mathbf{Ax} = \|\mathbf{Ax}\|^2 = 0$ and so $\mathbf{Ax} = \mathbf{0}$. Thus $N(\mathbf{A}) = N(\mathbf{A}'\mathbf{A})$. Since $\mathbf{A}$ and $\mathbf{A}'\mathbf{A}$ each have p columns, we see from (A.4.2h) that $\dim M(\mathbf{A}) = \dim M(\mathbf{A}'\mathbf{A})$ so that $r(\mathbf{A}) = r(\mathbf{A}'\mathbf{A})$.

Table A.4.1 Rank of some matrices

Matrix	Rank
Non-singular $\mathbf{A}(p\times p)$	p
diag (a_i)	Number of non-zero a_i
$\mathbf{H}_n$	$n-1$
Idempotent $\mathbf{A}$	tr $\mathbf{A}$
$\mathbf{CAB}$, non-singular $\mathbf{B}$, $\mathbf{C}$	$r(\mathbf{A})$

(5) If $\mathbf{A}$ is symmetric, its rank equals the number of non-zero eigenvalues of $\mathbf{A}$. For general $\mathbf{A}(n\times p)$, the rank is given by the number of non-zero eigenvalues of $\mathbf{A}'\mathbf{A}$. See Section A.6.

A.4.3 Linear equations

For the n linear equations

$$x_1\mathbf{a}_{(1)} + \ldots + x_p\mathbf{a}_{(p)} = \mathbf{b} \quad \text{(A.4.3a)}$$

or

$$\mathbf{Ax} = \mathbf{b} \quad \text{(A.4.3b)}$$

with the coefficient matrix $\mathbf{A}(n\times p)$, we note the following results:

(I) If $n = p$ and $\mathbf{A}$ is non-singular, the unique solution is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \frac{1}{|\mathbf{A}|}[A_{ij}]'\mathbf{b}. \quad \text{(A.4.3c)}$$

(II) The equation is *consistent* (i.e. admits at least one solution) if and only if

$$r(\mathbf{A}) = r[(\mathbf{A}, \mathbf{b})]. \quad \text{(A.4.3d)}$$

(III) For $\mathbf{b} = \mathbf{0}$, there exists a non-trivial solution (i.e. $\mathbf{x} \neq \mathbf{0}$) if and only if $r(\mathbf{A}) < p$.

(IV) The equation $\mathbf{A}'\mathbf{A} = \mathbf{A}'\mathbf{b}$ is always consistent. (A.4.3e)

Remarks (1) To prove (II) note that the vector $\mathbf{Ax}$ is a linear combination of the columns of $\mathbf{A}$. Thus the equation $\mathbf{Ax} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ can be expressed as a linear combination of the columns of $\mathbf{A}$.

(2) The proof of (III) is immediate from the definition of rank.

(3) To prove (IV) note that $M(\mathbf{A}'\mathbf{A}) \subseteq M(\mathbf{A}')$ because $\mathbf{A}'\mathbf{A}$ is a matrix whose columns are linear combinations of the columns of $\mathbf{A}'$. From Remark 4 of Section A.4.2 we see that $\dim M(\mathbf{A}'\mathbf{A}) = \dim M(\mathbf{A}) = \dim M(\mathbf{A}')$ and hence $M(\mathbf{A}'\mathbf{A}) = M(\mathbf{A}')$. Thus, $\mathbf{A}'\mathbf{b} \in M(\mathbf{A}'\mathbf{A})$, and so $r(\mathbf{A}'\mathbf{A}) = r(\mathbf{A}'\mathbf{A}, \mathbf{A}'\mathbf{b})$.

A.5 Linear Transformations

Definitions *The transformation from* $\mathbf{x}(p\times 1)$ *to* $\mathbf{y}(n\times 1)$ *given by*

$$\mathbf{y} = \mathbf{Ax} + \mathbf{b}, \quad \text{(A.5.1)}$$

where $\mathbf{A}$ *is an* $(n\times p)$ *matrix is called a* linear transformation. *For* $n = p$, *the transformation is called* non-singular *if* $\mathbf{A}$ *is non-singular and in this case the inverse transformation is*

$$\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}).$$

An orthogonal *transformation is defined by*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{A.5.2}$$

where $\mathbf{A}$ *is an orthogonal matrix. Geometrically, an orthogonal matrix represents a rotation of the coordinate axes. See Section* A.10.

A.6 Eigenvalues and Eigenvectors

A.6.1 General results

If $\mathbf{A}(p \times p)$ is any square matrix then

$$q(\lambda) = |\mathbf{A} - \lambda\mathbf{I}| \tag{A.6.1}$$

is a pth order polynomial in λ. The p roots of $q(\lambda)$, $\lambda_1, \ldots, \lambda_p$, possibly complex numbers, are called *eigenvalues* of $\mathbf{A}$. Some of the λ_i will be equal if $q(\lambda)$ has multiple roots.

For each $i = 1, \ldots, p$, $|\mathbf{A} - \lambda_i\mathbf{I}| = 0$, so $\mathbf{A} - \lambda_i\mathbf{I}$ is singular. Hence, there exists a non-zero vector $\boldsymbol{\gamma}$ satisfying

$$\mathbf{A}\boldsymbol{\gamma} = \lambda_i\boldsymbol{\gamma}. \tag{A.6.2}$$

Any vector satisfying (A.6.2) is called a (*right*) *eigenvector* of $\mathbf{A}$ for the eigenvalue λ_i. If λ_i is complex, then $\boldsymbol{\gamma}$ may have complex entries. An eigenvector $\boldsymbol{\gamma}$ with real entries is called *standardized* if

$$\boldsymbol{\gamma}'\boldsymbol{\gamma} = 1. \tag{A.6.3}$$

If $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors for λ_i and $\alpha \in R$, then $\mathbf{x} + \mathbf{y}$ and $\alpha\mathbf{x}$ are also eigenvectors for λ_i. Thus, the set of all eigenvectors for λ_i forms a subspace which is called the *eigenspace* of $\mathbf{A}$ for λ_i.

Since the coefficient of λ^p in $q(\lambda)$ is $(-1)^p$, we can write $q(\lambda)$ in terms of its roots as

$$q(\lambda) = \prod_{i=1}^{p} (\lambda_i - \lambda). \tag{A.6.4}$$

Setting $\lambda = 0$ in (A.6.1) and (A.6.4) gives

$$|\mathbf{A}| = \prod \lambda_i; \tag{A.6.5}$$

that is, $|\mathbf{A}|$ is the product of the eigenvalues of $\mathbf{A}$. Similarly, matching the coefficient of λ in (A.6.1) and (A.6.4) gives

$$\sum a_{ii} = \operatorname{tr} \mathbf{A} = \sum \lambda_i; \tag{A.6.6}$$

that is, $\operatorname{tr} \mathbf{A}$ is the sum of the eigenvalues of $\mathbf{A}$.

Let $\mathbf{C}(p \times p)$ be a non-singular matrix. Then

$$|\mathbf{A} - \lambda\mathbf{I}| = |\mathbf{C}|\,|\mathbf{A} - \lambda\mathbf{C}^{-1}\mathbf{C}|\,|\mathbf{C}^{-1}| = |\mathbf{C}\mathbf{A}\mathbf{C}^{-1} - \lambda\mathbf{I}|. \tag{A.6.7}$$

Thus $\mathbf{A}$ and $\mathbf{C}\mathbf{A}\mathbf{C}^{-1}$ have the same eigenvalues. Further, if $\boldsymbol{\gamma}$ is an eigenvector of $\mathbf{A}$ for λ_i, then $\mathbf{C}\mathbf{A}\mathbf{C}^{-1}(\mathbf{C}\boldsymbol{\gamma}) = \lambda_i\mathbf{C}\boldsymbol{\gamma}$, so that

$$\boldsymbol{\nu} = \mathbf{C}\boldsymbol{\gamma}$$

is an eigenvector of $\mathbf{C}\mathbf{A}\mathbf{C}^{-1}$ for λ_i.

Let $\alpha \in R$. Then $|\mathbf{A} + \alpha\mathbf{I} - \lambda\mathbf{I}| = |\mathbf{A} - (\lambda - \alpha)\mathbf{I}|$, so that $\mathbf{A} + \alpha\mathbf{I}$ has eigenvalues $\lambda_i + \alpha$. Further, if $\mathbf{A}\boldsymbol{\gamma} = \lambda_i\boldsymbol{\gamma}$, then $(\mathbf{A} + \alpha\mathbf{I})\boldsymbol{\gamma} = (\lambda_i + \alpha)\boldsymbol{\gamma}$, so that $\mathbf{A}$ and $\mathbf{A} + \alpha\mathbf{I}$ have the same eigenvectors.

Bounds on the dimension of the eigenspace of $\mathbf{A}$ for λ_i are given by the following theorem.

Theorem A.6.1 *Let* λ_1 *denote any particular eigenvalue of* $\mathbf{A}(p \times p)$, *with eigenspace* H *of dimension* r. *If* k *denotes the multiplicity of* λ_1 *in* $q(\lambda)$, *then* $1 \leq r \leq k$.

Proof Since λ_1 is an eigenvalue, there is at least one non-trivial eigenvector so $r \geq 1$.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_r$ be an orthonormal basis of H and extend it so that $\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_{p-r}$ is an orthonormal basis of R^p. Write $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_r)$, $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_{p-r})$. Then $(\mathbf{E}, \mathbf{F})$ is an orthogonal matrix so that $\mathbf{I}_p = (\mathbf{E}, \mathbf{F})(\mathbf{E}, \mathbf{F})' = \mathbf{E}\mathbf{E}' + \mathbf{F}\mathbf{F}'$ and $|(\mathbf{E}, \mathbf{F})| = 1$. Also $\mathbf{E}'\mathbf{A}\mathbf{E} = \lambda_1\mathbf{E}'\mathbf{E} = \lambda_1\mathbf{I}_r$, $\mathbf{F}'\mathbf{F} = \mathbf{I}_{p-r}$, and $\mathbf{F}'\mathbf{A}\mathbf{E} = \lambda_1\mathbf{F}'\mathbf{E} = \mathbf{0}$. Thus

$$\begin{aligned} q(\lambda) &= |\mathbf{A} - \lambda\mathbf{I}| = |(\mathbf{E}, \mathbf{F})'|\,|\mathbf{A} - \lambda\mathbf{I}|\,|(\mathbf{E}, \mathbf{F})| \\ &= |(\mathbf{E}, \mathbf{F})'[\mathbf{A}\mathbf{E}\mathbf{E}' + \mathbf{A}\mathbf{F}\mathbf{F}' - \lambda\mathbf{E}\mathbf{E}' - \lambda\mathbf{F}\mathbf{F}'](\mathbf{E}, \mathbf{F})| \\ &= \begin{vmatrix} (\lambda_1 - \lambda)\mathbf{I}_r & \mathbf{E}'\mathbf{A}\mathbf{F} \\ \mathbf{0} & \mathbf{F}'\mathbf{A}\mathbf{F} - \lambda\mathbf{I}_{p-r} \end{vmatrix} \\ &= (\lambda_1 - \lambda)^r q_1(\lambda), \text{ say,} \end{aligned}$$

using (A.2.3i). Thus the multiplicity of λ_1 as a root of $q(\lambda)$ is at least r.

Remarks (1) If $\mathbf{A}$ is symmetric then $r = k$; see Section A.6.2. However, if $\mathbf{A}$ is not symmetric, it is possible that $r < k$. For example,

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

has eigenvalue 0 with multiplicity 2; however, the corresponding eigenspace which is generated by $(1, 0)'$ only has dimension 1.

(2) If $r = 1$, then the eigenspace for λ_1 has dimension 1 and the standardized eigenvector for λ_1 is unique (up to sign).

Now let $\mathbf{A}(n \times p)$ and $\mathbf{B}(p \times n)$ be any two matrices and suppose $n \geq p$. Then from (A.2.3j)

$$\begin{vmatrix} -\lambda \mathbf{I}_n & -\mathbf{A} \\ \mathbf{B} & \mathbf{I}_p \end{vmatrix} = (-\lambda)^{n-p} |\mathbf{BA} - \lambda \mathbf{I}_p| = |\mathbf{AB} - \lambda \mathbf{I}_n|. \qquad \text{(A.6.8)}$$

Hence the n eigenvalues of $\mathbf{AB}$ equal the p eigenvalues of $\mathbf{BA}$, plus the eigenvalue 0, $n-p$ times. The following theorem describes the relationship between the eigenvectors.

Theorem A.6.2 *For $\mathbf{A}(n \times p)$ and $\mathbf{B}(p \times n)$, the non-zero eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ are the same and have the same multiplicity. If $\mathbf{x}$ is a non-trivial eigenvector of $\mathbf{AB}$ for an eigenvalue $\lambda \neq 0$, then $\mathbf{y} = \mathbf{Bx}$ is a non-trivial eigenvector of $\mathbf{BA}$.*

Proof The first part follows from (A.6.8). For the second part substituting $\mathbf{y} = \mathbf{Bx}$ in the equation $\mathbf{B}(\mathbf{ABx}) = \lambda \mathbf{Bx}$ gives $\mathbf{BAy} = \lambda \mathbf{y}$. The vector $\mathbf{x}$ is non-trivial if $\mathbf{x} \neq \mathbf{0}$. Since $\mathbf{Ay} = \mathbf{ABx} = \lambda \mathbf{x} \neq \mathbf{0}$, it follows that $\mathbf{y} \neq \mathbf{0}$ also. ■

Corollary A.6.2.1 *For $\mathbf{A}(n \times p)$, $\mathbf{B}(q \times n)$, $\mathbf{a}(p \times 1)$, and $\mathbf{b}(q \times 1)$, the matrix $\mathbf{Aab'B}$ has rank at most 1. The non-zero eigenvalue, if present, equals $\mathbf{b'BAa}$, with eigenvector $\mathbf{Aa}$.*

Proof The non-zero eigenvalue of $\mathbf{Aab'B}$ equals that of $\mathbf{b'BAa}$, which is a scalar, and hence is its own eigenvalue. The fact that $\mathbf{Aa}$ is a corresponding eigenvector is easily checked. ■

A.6.2 Symmetric matrices

If $\mathbf{A}$ is symmetric, it is possible to give more detailed information about its eigenvalues and eigenvectors.

Theorem A.6.3 *All the eigenvalues of a symmetric matrix $\mathbf{A}(p \times p)$ are real.*

Proof If possible, let

$$\boldsymbol{\gamma} = \mathbf{x} + i\mathbf{y}, \qquad \lambda = a + ib, \qquad \boldsymbol{\gamma} \neq \mathbf{0}. \qquad \text{(A.6.9)}$$

From (A.6.2), after equating real and imaginary parts, we have

$$\mathbf{Ax} = a\mathbf{x} - b\mathbf{y}, \qquad \mathbf{Ay} = b\mathbf{x} + a\mathbf{y}.$$

On premultiplying by $\mathbf{y}'$ and $\mathbf{x}'$, respectively, and subtracting, we obtain $b = 0$. Hence from (A.6.9), λ is real. ■

In the above discussion, we can choose $\mathbf{y} = \mathbf{0}$ so we can assume $\boldsymbol{\gamma}$ to be real.

Theorem A.6.4 (Spectral decomposition theorem, or Jordan decomposition theorem) *Any symmetric matrix $\mathbf{A}(p \times p)$ can be written as*

$$\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}' = \sum \lambda_i \boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}', \qquad \text{(A.6.10)}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\mathbf{A}$, and $\boldsymbol{\Gamma}$ is an orthogonal matrix whose columns are standardized eigenvectors.

Proof Suppose we can find orthonormal vectors $\boldsymbol{\gamma}_{(1)}, \ldots, \boldsymbol{\gamma}_{(p)}$ such that $\mathbf{A}\boldsymbol{\gamma}_{(i)} = \lambda_i \boldsymbol{\gamma}_{(i)}$ for some numbers λ_i. Then

$$\boldsymbol{\gamma}_{(i)}'\mathbf{A}\boldsymbol{\gamma}_{(j)} = \lambda_j \boldsymbol{\gamma}_{(i)}'\boldsymbol{\gamma}_{(j)} = \begin{cases} \lambda_i, & i = j, \\ 0, & i \neq j, \end{cases}$$

or in matrix form

$$\boldsymbol{\Gamma}'\mathbf{A}\boldsymbol{\Gamma} = \boldsymbol{\Lambda}. \qquad \text{(A.6.11)}$$

Pre- and post-multiplying by $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$ gives (A.6.10). From (A.6.7), $\mathbf{A}$ and $\boldsymbol{\Lambda}$ have the same eigenvalues, so the elements of $\boldsymbol{\Lambda}$ are exactly the eigenvalues of $\mathbf{A}$ with the same multiplicities.

Thus we must find an orthonormal basis of eigenvectors. Note that if $\lambda_i \neq \lambda_j$ are distinct eigenvalues with eigenvectors $\mathbf{x}+\mathbf{y}$, respectively, then $\lambda_i \mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{Ay} = \mathbf{y}'\mathbf{Ax} = \lambda_j \mathbf{y}'\mathbf{x}$, so that $\mathbf{y}'\mathbf{x} = 0$. *Hence for a symmetric matrix, eigenvectors corresponding to distinct eigenvalues are orthogonal to one another.*

Suppose there are k distinct eigenvalues of $\mathbf{A}$ with corresponding eigenspaces $H_1, \ldots, H_k$ of dimensions $r_1, \ldots, r_k$. Let

$$r = \sum_{j=1}^{k} r_j.$$

Since distinct eigenspaces are orthogonal, there exists an orthonormal set of vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$ such that the vectors labelled

$$\sum_{i=1}^{j-1} r_i + 1, \ldots, \sum_{i=1}^{j} r_i$$

form a basis for H_j. From Theorem A.6.1, r_j is less than or equal to the multiplicity of the corresponding eigenvalue. Hence by re-ordering the eigenvalues λ_i if necessary, we may suppose

$$\mathbf{Ae}_i = \lambda_i \mathbf{e}_i, \qquad i = 1, \ldots, r,$$

and $r \leq p$. (If all p eigenvalues are distinct, then we know from Theorem A.6.1 that $r = p$).

If $r = p$, set $\boldsymbol{\gamma}_{(i)} = \mathbf{e}_i$ and the proof follows. We shall show that the situation $r < p$ leads to a contradiction, and therefore cannot arise.

Without loss of generality we may suppose that all of the eigenvalues of $\mathbf{A}$ are strictly positive. (If not, we can replace $\mathbf{A}$ by $\mathbf{A} + \alpha\mathbf{I}$ for a suitable α, because $\mathbf{A}$ and $\mathbf{A} + \alpha\mathbf{I}$ have the same eigenvectors). Set

$$\mathbf{B} = \mathbf{A} - \sum_{i=1}^{r} \lambda_i \mathbf{e}_i \mathbf{e}_i'.$$

Then

$$\operatorname{tr} \mathbf{B} = \operatorname{tr} \mathbf{A} - \sum_{i=1}^{r} \lambda_i (\mathbf{e}_i'\mathbf{e}_i) = \sum_{i=r+1}^{p} \lambda_i > 0,$$

since $r < p$. Thus $\mathbf{B}$ has at least one non-zero eigenvalue, say θ. Let $\mathbf{x} \neq \mathbf{0}$ be a corresponding eigenvector. Then for $1 \leq j \leq r$,

$$\theta \mathbf{e}_j'\mathbf{x} = \mathbf{e}_j'\mathbf{B}\mathbf{x} = \left\{\lambda_j \mathbf{e}_j' - \sum_{i=1}^{r} \lambda_i (\mathbf{e}_j'\mathbf{e}_i)\mathbf{e}_i'\right\}\mathbf{x} = 0,$$

so that $\mathbf{x}$ is orthogonal to $\mathbf{e}_j$, $j = 1, \ldots, r$. Therefore,

$$\theta \mathbf{x} = \mathbf{B}\mathbf{x} = \left(\mathbf{A} - \sum \lambda_i \mathbf{e}_i \mathbf{e}_i'\right)\mathbf{x} = \mathbf{A}\mathbf{x} - \sum \lambda_i (\mathbf{e}_i'\mathbf{x})\mathbf{e}_i = \mathbf{A}\mathbf{x}$$

so that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ also. Thus $\theta = \lambda_i$ for some i and $\mathbf{x}$ is a linear combination of some of the $\mathbf{e}_i$, which contradicts the orthogonality between $\mathbf{x}$ and the $\mathbf{e}_i$. ■

Corollary A.6.4.1 *If $\mathbf{A}$ is a non-singular symmetric matrix, then for any integer n,*

$$\boldsymbol{\Lambda}^n = \operatorname{diag}(\lambda_i^n) \quad \text{and} \quad \mathbf{A}^n = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^n\boldsymbol{\Gamma}'. \tag{A.6.12}$$

If all the eigenvalues of $\mathbf{A}$ are positive then we can define the rational powers

$$\mathbf{A}^{r/s} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{r/s}\boldsymbol{\Gamma}', \quad \text{where} \quad \boldsymbol{\Lambda}^{r/s} = \operatorname{diag}(\lambda_i^{r/s}), \tag{A.6.13}$$

for integers $s > 0$ and r. If some of the eigenvalues of $\mathbf{A}$ are zero, then (A.6.12) and (A.6.13) hold if the exponents are restricted to be non-negative.

Proof Since

$$\mathbf{A}^2 = (\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}')^2 = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}'$$

and

$$\mathbf{A}^{-1} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}', \qquad \boldsymbol{\Lambda}^{-1} = \operatorname{diag}(\lambda_i^{-1}),$$

we see that (A.6.12) can be easily proved by induction. To check that rational powers make sense note that

$$(\mathbf{A}^{r/s})^s = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{r/s}\boldsymbol{\Gamma}' \ldots \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{r/s}\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^r\boldsymbol{\Gamma}' = \mathbf{A}^r. \quad ■$$

Motivated by (A.6.13), we can define powers of $\mathbf{A}$ for real-valued exponents.

Important special cases of (A.6.13) are

$$\mathbf{A}^{1/2} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Gamma}', \qquad \boldsymbol{\Lambda}^{1/2} = \operatorname{diag}(\lambda_i^{1/2}) \tag{A.6.14}$$

when $\lambda_i \geq 0$ for all i and

$$\mathbf{A}^{-1/2} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Gamma}', \qquad \boldsymbol{\Lambda}^{-1/2} = \operatorname{diag}(\lambda_i^{-1/2}) \tag{A.6.15}$$

when $\lambda_i > 0$ for all i. The decomposition (A.6.14) is called the *symmetric square root decomposition* of $\mathbf{A}$.

Corollary A.6.4.2 *The rank of $\mathbf{A}$ equals the number of non-zero eigenvalues.*

Proof By (A.4.2f), $r(\mathbf{A}) = r(\boldsymbol{\Lambda})$, whose rank is easily seen to equal the number of non-zero diagonal elements. ■

Remarks (1) Theorem A.6.4 shows that a symmetric matrix $\mathbf{A}$ is uniquely determined by its eigenvalues and eigenvectors, or more specifically by its distinct eigenvalues and corresponding eigenspaces.

(2) Since $\mathbf{A}^{1/2}$ has the same eigenvectors as $\mathbf{A}$ and has eigenvalues which are given functions of the eigenvalues of $\mathbf{A}$, we see that the symmetric square root is uniquely defined.

(3) If the λ_i are all distinct and written in decreasing order say, then $\boldsymbol{\Gamma}$ is uniquely determined, up to the signs of its columns.

(4) If $\lambda_{k+1} = \ldots = \lambda_p = 0$ then (A.6.10) can be written more compactly as

$$\mathbf{A} = \boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_1' = \sum_{i=1}^{k} \lambda_i \boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}',$$

where $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ and $\boldsymbol{\Gamma}_1 = (\boldsymbol{\gamma}_{(1)}, \ldots, \boldsymbol{\gamma}_{(k)})$.

(5) A symmetric matrix $\mathbf{A}$ has rank 1 if and only if

$$\mathbf{A} = \mathbf{x}\mathbf{x}'$$

for some $\mathbf{x}$. Then the only non-zero eigenvalue of $\mathbf{A}$ is given by

$$\operatorname{tr} \mathbf{A} = \operatorname{tr} \mathbf{x}\mathbf{x}' = \mathbf{x}'\mathbf{x}$$

and the corresponding eigenspace is generated by $\mathbf{x}$.

(6) Since $\mathbf{J} = \mathbf{1}\mathbf{1}'$ has rank 1 with eigenvalue p and corresponding eigenvector $\mathbf{1}$, we see that the equicorrelation matrix $\mathbf{E} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ has

eigenvalues $\lambda_1 = 1+(p-1)\rho$ and $\lambda_2 = \ldots = \lambda_p = 1-\rho$, and the same eigenvectors as $\mathbf{J}$. For the eigenvectors $\boldsymbol{\gamma}_{(2)}, \ldots, \boldsymbol{\gamma}_{(p)}$, we can select any standardized set of vectors orthogonal to $\mathbf{1}$ and each other. A possible choice for $\boldsymbol{\Gamma}$ is the Helmert matrix of Section A.3.1. Multiplying the eigenvalues together yields the formula for $|\mathbf{E}|$ given in (A.3.2c).

(7) If $\mathbf{A}$ is symmetric and idempotent (that is, $\mathbf{A}=\mathbf{A}'$ and $\mathbf{A}^2=\mathbf{A}$), then $\lambda_i = 0$ or 1 for all i, because $\mathbf{A}=\mathbf{A}^2$ implies $\boldsymbol{\Lambda}=\boldsymbol{\Lambda}^2$.

(8) If $\mathbf{A}$ is symmetric and idempotent then $r(\mathbf{A}) = \operatorname{tr} \mathbf{A}$. This result follows easily from (A.6.6) and Corollary A.6.4.2.

(9) As an example, consider

$$\mathbf{A} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{A.6.16}$$

The eigenvalues of $\mathbf{A}$ from (A.6.1) are the solutions of

$$\begin{vmatrix} 1-\lambda & \rho \\ \rho & 1-\lambda \end{vmatrix} = 0,$$

namely, $\lambda_1 = 1+\rho$ and $\lambda_2 = 1-\rho$. Thus,

$$\boldsymbol{\Lambda} = \operatorname{diag}(1+\rho, 1-\rho). \tag{A.6.17}$$

For $\rho \neq 0$, the eigenvector corresponding to $\lambda_1 = 1+\rho$ from (A.6.2) is

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (1+\rho)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

which leads to $x_1 = x_2$, therefore the first standardized eigenvector is

$$\boldsymbol{\gamma}_{(1)} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

Similarly, the eigenvector corresponding to $\lambda_2 = 1-\rho$ is

$$\boldsymbol{\gamma}_{(2)} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}.$$

Hence,

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}. \tag{A.6.18}$$

If $\rho = 0$ then $\mathbf{A} = \mathbf{I}$ and any orthonormal basis will do.

(10) Formula (A.6.14) suggests a method for calculating the symmetric *square root* of a matrix. For example, for the matrix in (A.6.16) with $\rho^2 < 1$, we find on using $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ from (A.6.11) and (A.6.14) that

$$\mathbf{A}^{1/2} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Gamma} = \begin{pmatrix} a & b \\ b & a \end{pmatrix},$$

where

$$2a = (1+\rho)^{1/2} + (1-\rho)^{1/2}, \qquad 2b = (1+\rho)^{1/2} - (1-\rho)^{1/2}.$$

*(11) The following methods are commonly used to calculate eigenvalues and eigenvectors on computers. For symmetric matrices, the Householder reduction to tri-diagonal form (i.e. $a_{ij} = 0$, for $i \geq j+2$ and $i \leq j-2$) is used followed by the QL algorithm. For non-symmetric matrices, reduction to upper Hessenberg form (i.e. $a_{ij} = 0$ for $i \geq j+2$) is used followed by the QR algorithm.

(12) For general matrices $\mathbf{A}(n \times p)$, we can use the spectral decomposition theorem to derive the following result.

Theorem A.6.5 (Singular value decomposition theorem) *If* $\mathbf{A}$ *is an* $(n \times p)$ *matrix of rank* r*, then* $\mathbf{A}$ *can be written as*

$$\mathbf{A} = \mathbf{U}\mathbf{L}\mathbf{V}' \tag{A.6.19}$$

where $\mathbf{U}(n \times r)$ *and* $\mathbf{V}(p \times r)$ *are column orthonormal matrices* ($\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$) *and* $\mathbf{L}$ *is a diagonal matrix with positive elements.*

Proof Since $\mathbf{A}'\mathbf{A}$ is a symmetric matrix which also has rank r, we can use the spectral decomposition theorem to write

$$\mathbf{A}'\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}', \tag{A.6.20}$$

where $\mathbf{V}(p \times r)$ is a column orthonormal matrix of eigenvectors of $\mathbf{A}'\mathbf{A}$ and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ contains the non-zero eigenvalues. Note that all the λ_i are positive because $\lambda_i = \mathbf{v}_{(i)}'\mathbf{A}'\mathbf{A}\mathbf{v}_{(i)} = \|\mathbf{A}\mathbf{v}_{(i)}\|^2 > 0$. Let

$$l_i = \lambda_i^{1/2}, \qquad i = 1, \ldots, r, \tag{A.6.21}$$

and set $\mathbf{L} = \operatorname{diag}(l_1, \ldots, l_r)$. Define $\mathbf{U}(n \times r)$ by

$$\mathbf{u}_{(i)} = l_i^{-1}\mathbf{A}\mathbf{v}_{(i)}, \qquad i = 1, \ldots, r. \tag{A.6.22}$$

Then

$$\mathbf{u}_{(j)}'\mathbf{u}_{(i)} = l_i^{-1}l_j^{-1}\mathbf{v}_{(j)}'\mathbf{A}'\mathbf{A}\mathbf{v}_{(i)} = \lambda_i l_i^{-1}l_j^{-1}\mathbf{v}_{(j)}'\mathbf{v}_{(i)} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Thus $\mathbf{U}$ is also a column orthonormal matrix.

Any p-vector $\mathbf{x}$ can be written as $\mathbf{x} = \sum \alpha_i \mathbf{v}_{(i)} + \mathbf{y}$ where $\mathbf{y} \in N(\mathbf{A})$, the null space of $\mathbf{A}$. Note that $N(\mathbf{A}) = N(\mathbf{A}'\mathbf{A})$ is the eigenspace of $\mathbf{A}'\mathbf{A}$ for the eigenvalue 0, so that $\mathbf{y}$ is orthogonal to the eigenvectors $\mathbf{v}_{(i)}$. Let $\mathbf{e}_i$

denote the r-vector with 1 in the ith place and 0 elsewhere. Then

$$\begin{aligned}\mathbf{ULV'x} &= \sum \alpha_i \mathbf{ULe}_i + \mathbf{0} \\ &= \sum \alpha_i l_i \mathbf{u}_{(i)} + \mathbf{0} \\ &= \sum \alpha_i \mathbf{Av}_{(i)} + \mathbf{Ay} = \mathbf{Ax}.\end{aligned}$$

Since this formula holds for all $\mathbf{x}$ it follows that $\mathbf{ULV'} = \mathbf{A}$. ■

Note that the columns of $\mathbf{U}$ are eigenvectors of $\mathbf{AA'}$ and the columns of $\mathbf{V}$ are eigenvectors of $\mathbf{A'A}$. Also, from Theorem A.6.2, the eigenvalues of $\mathbf{AA'}$ and $\mathbf{A'A}$ are the same.

A.7 Quadratic Forms and Definiteness

Definition A quadratic form *in the vector* $\mathbf{x}$ *is a function of the form*

$$Q(\mathbf{x}) \equiv \mathbf{x'Ax} = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j, \tag{A.7.1}$$

where $\mathbf{A}$ *is a symmetric matrix; that is,*

$$Q(\mathbf{x}) = a_{11}x_1^2 + \ldots + a_{pp}x_p^2 + 2a_{12}x_1x_2 + \ldots + 2a_{p-1,p}x_{p-1}x_p.$$

Clearly, $Q(\mathbf{0}) = 0$.

Definition (1) $Q(\mathbf{x})$ *is called a* positive definite (p.d.) quadratic form *if* $Q(\mathbf{x}) > 0$ *for all* $\mathbf{x} \neq \mathbf{0}$.

(2) $Q(\mathbf{x})$ *is called a* positive semi-definite (p.s.d) quadratic form *if* $Q(\mathbf{x}) \geq 0$ *for all* $\mathbf{x} \neq \mathbf{0}$.

(3) *A symmetric matrix* $\mathbf{A}$ *is called p.d. (p.s.d) if* $Q(\mathbf{x})$ *is p.d. (p.s.d.) and we write* $\mathbf{A} > 0$ *or* $\mathbf{A} \geq 0$ *for* $\mathbf{A}$ *positive definite or positive semi-definite, respectively.*

Negative definite and negative semi-definite quadratic forms are similarly defined.

For $p = 2$, $Q(\mathbf{x}) = x_1^2 + x_2^2$ *is p.d. while* $Q(\mathbf{x}) = (x_1 - x_2)^2$ *is p.s.d.*

Canonical form Any quadratic form can be converted into a weighted sum of squares without cross-product terms with the help of the following theorem.

Theorem A.7.1 *For any symmetric matrix* $\mathbf{A}$*, there exists an orthogonal transformation*

$$\mathbf{y} = \mathbf{\Gamma'x} \tag{A.7.2}$$

such that

$$\mathbf{x'Ax} = \sum \lambda_i y_i^2. \tag{A.7.3}$$

Proof *Consider the spectral decomposition given in Theorem* A.6.4:

$$\mathbf{A} = \mathbf{\Gamma\Lambda\Gamma'}. \tag{A.7.4}$$

From (A.7.2),

$$\mathbf{x'Ax} = \mathbf{y'\Gamma'A\Gamma y} = \mathbf{y'\Gamma'\Gamma\Lambda\Gamma'\Gamma y} = \mathbf{y'\Lambda y}.$$

Hence (A.7.3) follows. ■

It is important to recall that $\mathbf{\Gamma}$ has as its columns the eigenvectors of $\mathbf{A}$ and that $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{A}$. Using this theorem, we can deduce the following results for a matrix $\mathbf{A} > 0$.

Theorem A.7.2 *If* $\mathbf{A} > 0$ *then* $\lambda_i > 0$ *for* $i = 1, \ldots, p$. *If* $\mathbf{A} \geq 0$, *then* $\lambda_i \geq 0$.

Proof If $\mathbf{A} > 0$, we have, for all $\mathbf{x} \neq \mathbf{0}$,

$$0 < \mathbf{x'Ax} = \lambda_1 y_1^2 + \ldots + \lambda_p y_p^2.$$

From (A.7.2), $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{y} \neq \mathbf{0}$. Choosing $y_1 = 1$, $y_2 = \ldots = y_p = 0$, we deduce that $\lambda_1 > 0$. Similarly $\lambda_i > 0$ for all i. If $\mathbf{A} \geq 0$ the above inequalities are weak. ■

Corollary A.7.2.1 *If* $\mathbf{A} > 0$, *then* $\mathbf{A}$ *is non-singular and* $|\mathbf{A}| > 0$.

Proof Use the determinant of (A.7.4) with $\lambda_i > 0$. ■

Corollary A.7.2.2 *If* $\mathbf{A} > 0$, *then* $\mathbf{A}^{-1} > 0$.

Proof From (A.7.3), we have

$$\mathbf{x'A^{-1}x} = \sum y_i^2/\lambda_i. \quad ■ \tag{A.7.5}$$

Corollary A.7.2.3 (Symmetric decomposition) *Any matrix* $\mathbf{A} \geq 0$ *can be written as*

$$\mathbf{A} = \mathbf{B}^2, \tag{A.7.6}$$

where $\mathbf{B}$ *is a symmetric matrix.*

Proof Take $\mathbf{B} = \mathbf{\Gamma\Lambda}^{1/2}\mathbf{\Gamma'}$ in (A.7.4). ■

Theorem A.7.3 *If* $\mathbf{A} \geq 0$ *is a* $(p \times p)$ *matrix, then for any* $(p \times n)$ *matrix* $\mathbf{C}$, $\mathbf{C'AC} \geq 0$. *If* $\mathbf{A} > 0$ *and* $\mathbf{C}$ *is non-singular (so* $p = n$*), then* $\mathbf{C'AC} > 0$.

Proof If $\mathbf{A} \geq 0$ then for any n-vector $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x'C'ACx} = (\mathbf{Cx})'\mathbf{A}(\mathbf{Cx}) \geq 0, \quad \text{so} \quad \mathbf{C'AC} \geq 0.$$

If $\mathbf{A}>0$ and $\mathbf{C}$ is non-singular, the $\mathbf{Cx}\neq\mathbf{0}$, so $(\mathbf{Cx})'\mathbf{A}(\mathbf{Cx})>0$, and hence $\mathbf{C}'\mathbf{AC}>0$. ■

Corollary A.7.3.1 *If $\mathbf{A}\geq 0$ and $\mathbf{B}>0$ are $(p\times p)$ matrices, then all of the non-zero eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$ are positive.*

Proof Since $\mathbf{B}>0$, $\mathbf{B}^{-1/2}$ exists and, by Theorem A.6.2, $\mathbf{B}^{-1/2}\mathbf{AB}^{-1/2}$, $\mathbf{B}^{-1}\mathbf{A}$, and $\mathbf{AB}^{-1}$ have the same eigenvalues. By Theorem A.7.3, $\mathbf{B}^{-1/2}\mathbf{AB}^{-1/2}\geq 0$, so all of the non-zero eigenvalues are positive. ■

Remarks (1) There are other forms of interest:

(a) *Linear form.* $\mathbf{a}'\mathbf{x}=a_1x_1+\ldots+a_px_p$. Generally called a linear combination.

(b) *Bilinear form.* $\mathbf{x}'\mathbf{Ay}=\sum\sum a_{ij}x_iy_j$.

(2) We have noted in Corollary A.7.2.1 that $|\mathbf{A}|>0$ for $\mathbf{A}>0$. In fact, $|\mathbf{A}_{11}|>0$ for all partitions of $\mathbf{A}$. The proof follows on considering $\mathbf{x}'\mathbf{Ax}>0$ for all $\mathbf{x}$ with $x_{i+1}=\ldots=x_p=0$. The converse is also true.

(3) For

$$\boldsymbol{\Sigma}=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix},\qquad \rho^2<1$$

the transformation (A.7.2) is given by (A.6.18),

$$y_1=(x_1+x_2)/\sqrt{2},\qquad y_2=(x_1-x_2)/\sqrt{2}.$$

Thus, from (A.7.3) and (A.7.5),

$$\mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}=x_1^2+2\rho x_1x_2+x_2^2=(1+\rho)y_1^2+(1-\rho)y_2^2,$$

$$\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}=\frac{1}{(1-\rho^2)}(x_1^2-2\rho x_1x_2+x_2^2)=\frac{y_1^2}{1+\rho}+\frac{y_2^2}{1-\rho}.$$

A geometrical interpretation of these results will be found in Section A.10.4.

(4) Note that the centring matrix $\mathbf{H}\geq 0$ because $\mathbf{x}'\mathbf{Hx}=\sum(x_i-\bar{x})^2\geq 0$.

(5) For any matrix $\mathbf{A}$, $\mathbf{AA}'\geq 0$ and $\mathbf{A}'\mathbf{A}\geq 0$. Further, $r(\mathbf{AA}')=r(\mathbf{A}'\mathbf{A})=r(\mathbf{A})$.

*A.8 Generalized Inverse

We now consider a method of defining an inverse for any matrix.

Definition *For a matrix $\mathbf{A}(n\times p)$, $\mathbf{A}^-$ is called a* g-inverse (generalized inverse) *of $\mathbf{A}$ if*

$$\mathbf{AA}^-\mathbf{A}=\mathbf{A}. \qquad (A.8.1)$$

A generalized inverse always exists although in general it is not unique. ■

Methods of construction

(1) Using the singular value decomposition theorem, (Theorem A.6.5) for $\mathbf{A}(n\times p)$, write $\mathbf{A}=\mathbf{ULV}'$. Then it is easily checked that

$$\mathbf{A}^-=\mathbf{VL}^{-1}\mathbf{U}' \qquad (A.8.2)$$

defines a g-inverse.

(2) If $r(\mathbf{A})=r$, re-arrange the rows and columns of $\mathbf{A}(n\times p)$ and partition $\mathbf{A}$ so that $\mathbf{A}_{11}$ is an $(r\times r)$ non-singular matrix. Then it can be verified that

$$\mathbf{A}^-=\begin{pmatrix}\mathbf{A}_{11}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} \qquad (A.8.3)$$

is a g-inverse.

The result follows on noting that there exist $\mathbf{B}$ and $\mathbf{C}$ such that

$$\mathbf{A}_{12}=\mathbf{A}_{11}\mathbf{B},\ \mathbf{A}_{21}=\mathbf{C}\,\mathbf{A}_{11}\text{ and }\mathbf{A}_{22}=\mathbf{C}\,\mathbf{A}_{11}\mathbf{B}.$$

(3) If $\mathbf{A}(p\times p)$ is non-singular then $\mathbf{A}^-=\mathbf{A}^{-1}$ is uniquely defined.

(4) If $\mathbf{A}(p\times p)$ is symmetric of rank r, then, using Remark 4 after Theorem A.6.4, $\mathbf{A}$ can be written as $\mathbf{A}=\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_1'$, where $\boldsymbol{\Gamma}_1$ is a column orthonormal matrix of eigenvectors corresponding to the non-zero eigenvalues $\boldsymbol{\Lambda}_1=\operatorname{diag}(\lambda_1,\ldots,\lambda_r)$ of $\mathbf{A}$. Then it is easily checked that

$$\mathbf{A}^-=\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Gamma}_1' \qquad (A.8.4)$$

is a g-inverse.

Applications

(1) *Linear equations.* A particular solution of the consistent equations

$$\mathbf{Ax}=\mathbf{b}, \qquad (A.8.5)$$

is

$$\mathbf{x}=\mathbf{A}^-\mathbf{b}. \qquad (A.8.6)$$

Proof From (A.8.1),

$$\mathbf{AA}^-\mathbf{Ax}=\mathbf{Ax}\Rightarrow\mathbf{A}(\mathbf{A}^-\mathbf{b})=\mathbf{b}$$

which when compared with (A.8.5) leads to (A.8.6). ■

It can be shown that a general solution of a consistent equation is

$$\mathbf{x}=\mathbf{A}^-\mathbf{b}+(\mathbf{I}-\mathbf{G})\mathbf{z},$$

where $\mathbf{z}$ is arbitrary and $\mathbf{G}=\mathbf{A}^-\mathbf{A}$. For $\mathbf{b}=\mathbf{0}$, a general solution is $(\mathbf{I}-\mathbf{G})\mathbf{z}$.

(2) *Quadratic forms.* Let $\mathbf{A}(p \times p)$ be a symmetric matrix of rank $r \leq p$. Then there exists an orthogonal transformation such that for $\mathbf{x}$ restricted to $M(\mathbf{A})$ the subspace spanned by the columns of $\mathbf{A}$, $\mathbf{x}'\mathbf{A}^-\mathbf{x}$ can be written as

$$\mathbf{x}'\mathbf{A}^-\mathbf{x} = \sum u_i^2/\lambda_i, \tag{A.8.7}$$

where $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of $\mathbf{A}$.

Proof First note that if $\mathbf{x}$ lies in $M(\mathbf{A})$ we can write $\mathbf{x} = \mathbf{A}\mathbf{y}$ for some $\mathbf{y}$, so that

$$\mathbf{x}'\mathbf{A}^-\mathbf{x} = \mathbf{y}'\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{y} = \mathbf{y}'\mathbf{A}\mathbf{y}$$

does not depend upon the particular g-inverse chosen. From the spectral decomposition of $\mathbf{A}$ we see that $M(\mathbf{A})$ is spanned by the eigenvectors of $\mathbf{A}$ corresponding to non-zero eigenvalues, say by $(\boldsymbol{\gamma}_{(1)}, \ldots, \boldsymbol{\gamma}_{(r)}) = \boldsymbol{\Gamma}_1$. Then if $\mathbf{x} \in M(\mathbf{A})$, it can be written as $\mathbf{x} = \boldsymbol{\Gamma}_1\mathbf{u}$ for some r-vector $\mathbf{u}$. Defining $\mathbf{A}^-$ by (A.8.4), we see that (A.8.7) follows.

Remarks (1) For the equicorrelation matrix $\mathbf{E}$, if $1+(p-1)\rho = 0$, then $(1-\rho)^{-1}\mathbf{I}$ is a g-inverse of $\mathbf{E}$.

(2) Under the following conditions $\mathbf{A}^-$ is defined uniquely:

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}, \qquad \mathbf{A}\mathbf{A}^- \text{ and } \mathbf{A}^-\mathbf{A} \text{ symmetric}, \qquad \mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-.$$

*(3) For $\mathbf{A} \geq 0$, $\mathbf{A}^-$ is normally computed by using Cholesky decomposition (see Remark 4, Section A.2.4.).

A.9 Matrix Differentiation and Maximization Problems

Let us define the derivative of $f(\mathbf{X})$ with respect to $\mathbf{X}(n \times p)$ as the matrix

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

We have the following results:

$$\text{(I)}\quad \frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}. \tag{A.9.1}$$

$$\text{(II)}\quad \frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}, \quad \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}+\mathbf{A}')\mathbf{x}, \quad \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y}. \tag{A.9.2}$$

$$\text{(III)}\quad \frac{\partial |\mathbf{X}|}{\partial x_{ij}} = X_{ij} \text{ if all elements of } \mathbf{X}(n \times n) \text{ are distinct}$$

$$= \begin{cases} X_{ii}, & i=j \\ 2X_{ij}, & i \neq j \end{cases} \text{ if } \mathbf{X} \text{ is symmetric}, \tag{A.9.3}$$

where X_{ij} is the (i, j)th cofactor of $\mathbf{X}$.

$$\text{(IV)}\quad \frac{\partial \operatorname{tr} \mathbf{X}\mathbf{Y}}{\partial \mathbf{X}} = \mathbf{Y}' \text{ if all elements of } \mathbf{X}(n \times p) \text{ are distinct},$$

$$= \mathbf{Y} + \mathbf{Y}' - \operatorname{Diag}(\mathbf{Y}) \text{ if } \mathbf{X}(n \times n) \text{ is symmetric}. \tag{A.9.4}$$

$$\text{(V)}\quad \frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1}\mathbf{J}_{ij}\mathbf{X}^{-1} \text{ if all elements of } \mathbf{X}(n \times n) \text{ are distinct}$$

$$= \begin{cases} -\mathbf{X}^{-1}\mathbf{J}_{ii}\mathbf{X}^{-1}, & i=j, \\ -\mathbf{X}^{-1}(\mathbf{J}_{ij}+\mathbf{J}_{ji})\mathbf{X}^{-1}, & i \neq j \end{cases} \text{ if } \mathbf{X} \text{ is symmetric}, \tag{A.9.5}$$

where $\mathbf{J}_{ij}$ denotes a matrix with a 1 in the (i, j)th place and zeros elsewhere.

We now consider some applications of these results to some stationary value problems.

Theorem A.9.1 *The vector $\mathbf{x}$ which minimizes*

$$f(\mathbf{x}) = (\mathbf{y} - \mathbf{A}\mathbf{x})'(\mathbf{y} - \mathbf{A}\mathbf{x})$$

is given by

$$\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{A}'\mathbf{y}. \tag{A.9.6}$$

Proof Differentiate $f(\mathbf{x})$ and set the derivative equal to $\mathbf{0}$. Note that the second derivative matrix $2\mathbf{A}'\mathbf{A} \geq 0$ so that the solution to (A.9.6) will give a minimum. Also note that from (A.4.3e), (A.9.6) is a consistent set of equations. ∎

Theorem A.9.2 *Let $\mathbf{A}$ and $\mathbf{B}$ be two symmetric matrices. Suppose that $\mathbf{B} > 0$. Then the maximum (minimum) of $\mathbf{x}'\mathbf{A}\mathbf{x}$ given*

$$\mathbf{x}'\mathbf{B}\mathbf{x} = 1 \tag{A.9.7}$$

is attained when $\mathbf{x}$ is the eigenvector of $\mathbf{B}^{-1}\mathbf{A}$ corresponding to the largest (smallest) eigenvalue of $\mathbf{B}^{-1}\mathbf{A}$. Thus if λ_1 and λ_p are the largest and smallest eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$, then, subject to the constraint (A.9.7),

$$\max_{\mathbf{x}} \mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1, \qquad \min_{\mathbf{x}} \mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_p. \tag{A.9.8}$$

Proof Let $\mathbf{B}^{1/2}$ denote the symmetric square root of $\mathbf{B}$, and let $\mathbf{y} = \mathbf{B}^{1/2}\mathbf{x}$.

Then the maximum of $\mathbf{x}'\mathbf{A}\mathbf{x}$ subject to (A.9.7) can be written as

$$\max_{\mathbf{y}} \mathbf{y}'\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\mathbf{y} \quad \text{subject to} \quad \mathbf{y}'\mathbf{y}=1. \tag{A.9.9}$$

Let $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}=\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'$ be a spectral decomposition of the symmetric matrix $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$. Let $\mathbf{z}=\mathbf{\Gamma}'\mathbf{y}$. Then $\mathbf{z}'\mathbf{z}=\mathbf{y}'\mathbf{\Gamma}\mathbf{\Gamma}'\mathbf{y}=\mathbf{y}'\mathbf{y}$ so that (A.9.9) can be written

$$\max_{\mathbf{z}} \mathbf{z}'\mathbf{\Lambda}\mathbf{z}=\max_{\mathbf{z}} \sum \lambda_i z_i^2 \quad \text{subject to} \quad \mathbf{z}'\mathbf{z}=1. \tag{A.9.10}$$

If the eigenvalues are written in descending order then (A.9.10) satisfies

$$\max \sum \lambda_i z_i^2 \leq \lambda_1 \max \sum z_i^2 = \lambda_1.$$

Further this bound is attained for $\mathbf{z}=(1,0,\ldots,0)'$, that is for $\mathbf{y}=\boldsymbol{\gamma}_{(1)}$, and for $\mathbf{x}=\mathbf{B}^{-1/2}\boldsymbol{\gamma}_{(1)}$. By Theorem A.6.2, $\mathbf{B}^{-1}\mathbf{A}$ and $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ have the same eigenvalues and $\mathbf{x}=\mathbf{B}^{-1/2}\boldsymbol{\gamma}_{(1)}$ is an eigenvector of $\mathbf{B}^{-1}\mathbf{A}$ corresponding to λ_1. Thus the theorem is proved for maximization.

The same technique can be applied to prove the minimization result. ■

Corollary A.9.2.1 *If $R(\mathbf{x})=\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}$ then, for $\mathbf{x}\neq\mathbf{0}$,*

$$\lambda_p \leq R(\mathbf{x}) \leq \lambda_1. \tag{A.9.11}$$

Proof Since $R(\mathbf{x})$ is invariant under changes of scale of $\mathbf{x}$, we can regard the problem as maximizing (minimizing) $\mathbf{x}'\mathbf{A}\mathbf{x}$ given (A.9.7). ■

Corollary A.9.2.2 *The maximum of $\mathbf{a}'\mathbf{x}$ subject to* (A.9.7) *is*

$$(\mathbf{a}'\mathbf{B}^{-1}\mathbf{a})^{1/2}. \tag{A.9.12}$$

Further

$$\max_{\mathbf{x}} \{(\mathbf{a}'\mathbf{x})^2/(\mathbf{x}'\mathbf{B}\mathbf{x})\}=\mathbf{a}'\mathbf{B}^{-1}\mathbf{a} \tag{A.9.13}$$

and the maximum is attained at $\mathbf{x}=\mathbf{B}^{-1}\mathbf{a}/(\mathbf{a}'\mathbf{B}^{-1}\mathbf{a})^{1/2}$.

Proof Apply Theorem A.9.2 with $\mathbf{x}'\mathbf{A}\mathbf{x}=(\mathbf{a}'\mathbf{x})^2=\mathbf{x}'(\mathbf{a}\mathbf{a}')\mathbf{x}$. ■

Remarks (1) A direct method is sometimes instructive. Consider the problem of maximizing the squared distance from the origin

$$x^2+y^2$$

of a point (x, y) on the ellipse

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{A.9.14}$$

When y^2 is eliminated, the problem reduces to finding the maximum of

$$x^2+b^2(x^2/a^2-1), \qquad x\in[-a,a].$$

Setting the derivative equal to 0 yields the stationary point $x=0$ which, from (A.9.14), gives $y=\pm b$. Also, at the endpoints of the interval ($x=\pm a$), we get $y=0$. Hence

$$\max(x^2+y^2)=\max(a^2,b^2).$$

This solution is not as elegant as the proof of Theorem A.9.2, and does not generalize neatly to more complicated quadratic forms.

(2) The results (A.9.1)–(A.9.2) follow by direct substitution, e.g.

$$\frac{\partial}{\partial x_1}\mathbf{a}'\mathbf{x}=\frac{\partial}{\partial x_1}(a_1x_1+\ldots+a_px_p)=a_1$$

proves (A.9.1). For (A.9.3) use (A.2.3d).

A.10 Geometrical Ideas

A.10.1 *n*-dimensional geometry

Let $\mathbf{e}_i$ denote the vector in R^n with 1 in the ith place and zeros elsewhere so that $(\mathbf{e}_1,\ldots,\mathbf{e}_n)$ forms an orthonormal basis of R^n. In terms of this basis, vectors $\mathbf{x}$ can be represented as $\mathbf{x}=\sum x_i\mathbf{e}_i$, and x_i is called the ith *coordinate axis*. A point $\mathbf{a}$ in R^n is represented in terms of these coordinates by $x_1=a_1,\ldots,x_n=a_n$. The point $\mathbf{a}$ can also be interpreted as a directed line segment from $\mathbf{0}$ to $\mathbf{a}$. Some generalizations of various basic concepts of two- and three-dimensional analytic Euclidean geometry are summarized in Table A.10.1.

A.10.2 Orthogonal transformations

Let $\mathbf{\Gamma}$ be an orthogonal matrix. Then $\mathbf{\Gamma}\mathbf{e}_i=\boldsymbol{\gamma}_{(i)}$, $i=1,\ldots,n$, also form an orthonormal basis and points $\mathbf{x}$ can be represented in terms of this new basis as

$$\mathbf{x}=\sum x_i\mathbf{e}_i=\sum y_i\boldsymbol{\gamma}_{(i)},$$

where $y_i=\boldsymbol{\gamma}_{(i)}'\mathbf{x}$ are new coordinates. If $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are two points with new coordinates $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ note that

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

Table A.10.1 Basic concepts in n-dimensional geometry

Concept	Description $\left(\|\mathbf{x}\| = \left(\sum x_i^2\right)^{1/2}\right)$
Point $\mathbf{a}$	$x_1 = a_1, \ldots, x_n = a_n$
Distance between $\mathbf{a}$ and $\mathbf{b}$	$\|\mathbf{a}-\mathbf{b}\| = \left\{\sum (a_i - b_i)^2\right\}^{1/2}$
Line passing through $\mathbf{a}$, $\mathbf{b}$	$\mathbf{x} = \lambda\mathbf{a} + (1-\lambda)\mathbf{b}$ is the equation
Line passing through $\mathbf{0}$, $\mathbf{a}$	$\mathbf{x} = \lambda\mathbf{a}$
Angle between lines from $\mathbf{0}$ to $\mathbf{a}$ and $\mathbf{0}$ to $\mathbf{b}$	θ where $\cos\theta = \mathbf{a}'\mathbf{b}/\{\|\mathbf{a}\|\,\|\mathbf{b}\|\}^{1/2}$, $0 \leqslant \theta \leqslant \pi$
Direction cosine vector of a line from $\mathbf{0}$ to $\mathbf{a}$	$(\cos\gamma_1, \ldots, \cos\gamma_n)$, $\cos\gamma_i = a_i/\|\mathbf{a}\|$; γ_i = angle between line and ith axis
Plane P	$\mathbf{a}'\mathbf{x} = c$ is general equation
Plane through $\mathbf{b}_1, \ldots, \mathbf{b}_k$	$\mathbf{x} = \sum \lambda_i \mathbf{b}_i$, $\sum \lambda_i = 1$
Plane through $\mathbf{0}, \mathbf{b}_1, \ldots, \mathbf{b}_k$	$\mathbf{x} = \sum \lambda_i \mathbf{b}_i$
Hypersphere with centre $\mathbf{a}$ and radius r	$(\mathbf{x}-\mathbf{a})'(\mathbf{x}-\mathbf{a}) = r^2$
Ellipsoid	$(\mathbf{x}-\mathbf{a})'\mathbf{A}^{-1}(\mathbf{x}-\mathbf{a}) = c^2$, $\mathbf{A} > 0$

so that orthogonal transformations preserve distances. An orthogonal transformation represents a rotation of the coordinate axes (plus a reflection if $|\boldsymbol{\Gamma}| = -1$). When $n = 2$ and $|\boldsymbol{\Gamma}| = 1$, $\boldsymbol{\Gamma}$ can be represented as

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

and represents a rotation of the coordinate axes counterclockwise through an angle θ.

A.10.3 Projections

Consider a point $\mathbf{a}$, in n dimensions (see Figure A.10.1). Its projection onto a plane P (or onto a line) through the origin is the point $\hat{\mathbf{a}}$ at the foot of the perpendicular from $\mathbf{a}$ to P. The vector $\hat{\mathbf{a}}$ is called the *orthogonal projection* of the vector $\mathbf{a}$ onto the plane.

Let the plane P pass through points $\mathbf{0}, \mathbf{b}_1, \ldots, \mathbf{b}_k$ so that its equation from Table A.10.1 is

$$\mathbf{x} = \sum \lambda_i \mathbf{b}_i, \qquad \mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_k).$$

Suppose rank $(\mathbf{B}) = k$ so that the plane is a k-dimensional subspace. The

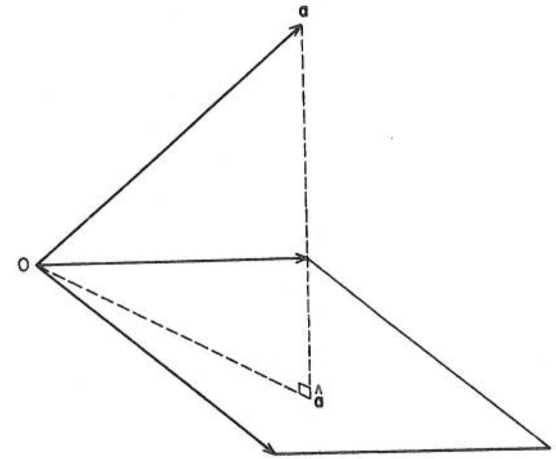

Figure A.10.1 *$\hat{\mathbf{a}}$ is the projection of $\mathbf{a}$ onto the plane P.*

point $\hat{\mathbf{a}}$ is defined by $\mathbf{x} = \sum \hat{\lambda}_i \mathbf{b}_i$, where $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ minimize

$$\left\|\mathbf{a} - \sum \lambda_i \mathbf{b}_i\right\|$$

since $\hat{\mathbf{a}}$ is the point on the plane closest to $\mathbf{a}$. Using Theorem A.9.1, we deduce the following result.

Theorem A.10.1 *The point $\hat{\mathbf{a}}$ is given by*

$$\hat{\mathbf{a}} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{a}. \quad \blacksquare \qquad \text{(A.10.3a)}$$

Note that $\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$ is a symmetric idempotent matrix. In fact, any symmetric idempotent matrix can be used to represent a projection.

A.10.4 Ellipsoids

Let $\mathbf{A}$ be a p.d. matrix. Then

$$(\mathbf{x}-\boldsymbol{\alpha})'\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\alpha}) = c^2 \qquad \text{(A.10.4a)}$$

represents an ellipsoid in n dimensions. We note that the centre of the ellipsoid is at $\mathbf{x} = \boldsymbol{\alpha}$. On shifting the centre to $\mathbf{x} = \mathbf{0}$, the equation becomes

$$\mathbf{x}'\mathbf{A}^{-1}\mathbf{x} = c^2. \qquad \text{(A.10.4b)}$$

Definition *Let $\mathbf{x}$ be a point on the ellipsoid defined by* (A.10.4a) *and let $f(\mathbf{x}) = \|\mathbf{x}-\boldsymbol{\alpha}\|^2$ denote the squared distance between $\boldsymbol{\alpha}$ and $\mathbf{x}$. A line through $\boldsymbol{\alpha}$ and $\mathbf{x}$ for which $\mathbf{x}$ is a stationary point of $f(\mathbf{x})$ is called a* principal axis *of* *the ellipsoid. The distance* $\|\mathbf{x}-\boldsymbol{\alpha}\|$ *is called the length of the principal semi-axis.*

Theorem A.10.2 *Let* $\lambda_1, \ldots, \lambda_n$ *be the eigenvalues of* **A** *satisfying* $\lambda_1 > \lambda_2 > \ldots > \lambda_n$. *Suppose that* $\gamma_{(1)}, \ldots, \gamma_{(n)}$ *are the corresponding eigenvectors. For the ellipsoids* (A.10.4a) *and* (A.10.4b), *we have*

(1) *The direction cosine vector of the ith principal axis is* $\gamma_{(i)}$.
(2) *The length of the ith principal semi-axis is* $c\lambda_i^{1/2}$.

Proof It is sufficient to prove the result for (A.10.4b). The problem reduces to finding the stationary points of $f(\mathbf{x}) = \mathbf{x}'\mathbf{x}$ subject to $\mathbf{x}$ lying on the ellipsoid $\mathbf{x}'\mathbf{A}^{-1}\mathbf{x} = c^2$. The derivative of $\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}$ is $2\mathbf{A}^{-1}\mathbf{x}$. Thus a point $\mathbf{y}$ represents a direction tangent to the ellipsoid at $\mathbf{x}$ if $2\mathbf{y}'\mathbf{A}^{-1}\mathbf{x} = 0$.

The derivative of $f(\mathbf{x})$ is $2\mathbf{x}$ so the directional derivative of $f(\mathbf{x})$ in the direction $\mathbf{y}$ is $2\mathbf{y}'\mathbf{x}$. Then $\mathbf{x}$ is a stationary point if and only if for all points $\mathbf{y}$ representing tangent directions to the ellipsoid at $\mathbf{x}$, we have $2\mathbf{y}'\mathbf{x} = 0$; that is if

$$\mathbf{y}'\mathbf{A}^{-1}\mathbf{x} = 0 \Rightarrow \mathbf{y}'\mathbf{x} = 0.$$

This condition is satisfied if and only if $\mathbf{A}^{-1}\mathbf{x}$ is proportional to $\mathbf{x}$; that is if and only if $\mathbf{x}$ is an eigenvector of $\mathbf{A}^{-1}$.

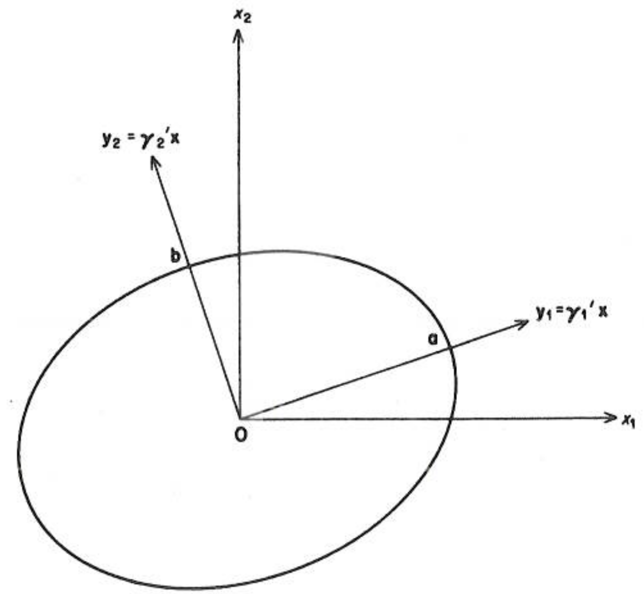

Figure A.10.2 Ellipsoid $\mathbf{x}'\mathbf{A}^{-1}\mathbf{x} = 1$. *Lines defined by* $\mathbf{y}_1$ *and* $\mathbf{y}_2$ *are the first and second principal axes,* $\|\mathbf{a}\| = \lambda_1^{1/2}$, $\|\mathbf{b}\| = \lambda_2^{1/2}$.

Setting $\mathbf{x} = \beta\gamma_{(i)}$ in (A.10.4b) gives $\beta^2/\lambda_i = c^2$, so $\beta = c_i\lambda_i^{1/2}$. Thus, the theorem is proved. ■

If we rotate the coordinate axes with the transformation $\mathbf{y} = \boldsymbol{\Gamma}'\mathbf{x}$, we find that (A.10.4b) reduces to

$$\sum y_i^2/\lambda_i = c^2.$$

Figure A.10.2 gives a pictorial representation.

With $\mathbf{A} = \mathbf{I}$, (A.10.4b) reduces to a hypersphere with $\lambda_1 = \ldots = \lambda_n = 1$ so that the λs are not distinct and the above theorem fails; that is, the position of $\gamma_{(i)}$, $i = 1, \ldots, n$, through the sphere is not unique and any rotation will suffice; that is, all the n components are isometric.

In general, if $\lambda_i = \lambda_{i+1}$, the section of the ellipsoid is *circular* in the plane generated by $\gamma_{(i)}, \gamma_{(i+1)}$. Although we can construct two perpendicular axes for the common root, their position through the circle is not unique. If **A** equals the equicorrelation matrix, there are $p-1$ isotropic principal axes corresponding to the last $p-1$ eigenvalues.