Chapter 2
A Short Excursion into Matrix Algebra

This chapter serves as a reminder of basic concepts of matrix algebra, which are particularly useful in multivariate analysis. It also introduces the notations used in this book for vectors and matrices. Eigenvalues and eigenvectors play an important role in multivariate techniques. In Sections 2.2 and 2.3, we present the spectral decomposition of matrices and consider the maximisation (minimisation) of quadratic forms given some constraints.

In analyzing the multivariate normal distribution, partitioned matrices appear naturally. Some of the basic algebraic properties are given in Section 2.5. These properties will be heavily used in Chapters 4 and 5.

The geometry of the multinormal and the geometric interpretation of the multivariate techniques (Part III) intensively uses the notion of angles between two vectors, the projection of a point on a vector and the distances between two points. These ideas are introduced in Section 2.6.

2.1 Elementary Operations

A matrix $\mathcal{A}$ is a system of numbers with n rows and p columns:

$$\mathcal{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & \dots & a_{1p} \\ \vdots & a_{22} & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & \dots & \dots & a_{np} \end{pmatrix}.$$

We also write (a_{ij}) for $\mathcal{A}$ and $\mathcal{A}(n \times p)$ to indicate the numbers of rows and columns. Vectors are matrices with one column and are denoted as x or $x(p \times 1)$. Special matrices and vectors are defined in Table 2.1. Note that we use small letters for scalars as well as for vectors.

W.K. Härdle, L. Simar, *Applied Multivariate Statistical Analysis*,
DOI 10.1007/978-3-642-17229-8_2, © Springer-Verlag Berlin Heidelberg 2012

Table 2.1 Special matrices and vectors

Name	Definition	Notation	Example
scalar	$p = n = 1$	a	3
column vector	$p = 1$	a	$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$
row vector	$n = 1$	$a^\top$	$\begin{pmatrix} 1 & 3 \end{pmatrix}$
vector of ones	$\underbrace{(1, \ldots, 1)}_{n}^\top$	1_n	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
vector of zeros	$\underbrace{(0, \ldots, 0)}_{n}^\top$	0_n	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
square matrix	$n = p$	$\mathcal{A}(p \times p)$	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$
diagonal matrix	$a_{ij} = 0, i \neq j, n = p$	$\mathrm{diag}(a_{ii})$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
identity matrix	$\mathrm{diag}\underbrace{(1, \ldots, 1)}_{p}$	$\mathcal{I}_p$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
unit matrix	$a_{ij} = 1, n = p$	$1_n 1_n^\top$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
symmetric matrix	$a_{ij} = a_{ji}$		$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$
null matrix	$a_{ij} = 0$	0	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
upper triangular matrix	$a_{ij} = 0, i < j$		$\begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$
idempotent matrix	$\mathcal{A}\mathcal{A} = \mathcal{A}$		$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$
orthogonal matrix	$\mathcal{A}^\top \mathcal{A} = \mathcal{I} = \mathcal{A}\mathcal{A}^\top$		$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$

Matrix Operations

Elementary operations are summarised below:

$$\mathcal{A}^\top = (a_{ji})$$
$$\mathcal{A} + \mathcal{B} = (a_{ij} + b_{ij})$$

$$\mathcal{A} - \mathcal{B} = (a_{ij} - b_{ij})$$
$$c \cdot \mathcal{A} = (c \cdot a_{ij})$$
$$\mathcal{A} \cdot \mathcal{B} = \mathcal{A}(n \times p)\, \mathcal{B}(p \times m) = \mathcal{C}(n \times m) = \left(\sum_{j=1}^{p} a_{ij} b_{jk} \right).$$

Properties of Matrix Operations

$$\begin{aligned} \mathcal{A} + \mathcal{B} &= \mathcal{B} + \mathcal{A} \\ \mathcal{A}(\mathcal{B} + \mathcal{C}) &= \mathcal{A}\mathcal{B} + \mathcal{A}\mathcal{C} \\ \mathcal{A}(\mathcal{B}\mathcal{C}) &= (\mathcal{A}\mathcal{B})\mathcal{C} \\ (\mathcal{A}^\top)^\top &= \mathcal{A} \\ (\mathcal{A}\mathcal{B})^\top &= \mathcal{B}^\top \mathcal{A}^\top \end{aligned}$$

Matrix Characteristics

Rank

The *rank*, $\text{rank}(\mathcal{A})$, of a matrix $\mathcal{A}(n \times p)$ is defined as the maximum number of linearly independent rows (columns). A set of k rows a_j of $\mathcal{A}(n \times p)$ are said to be linearly independent if $\sum_{j=1}^{k} c_j a_j = 0_p$ implies $c_j = 0, \forall j$, where $c_1, \ldots, c_k$ are scalars. In other words no rows in this set can be expressed as a linear combination of the $(k - 1)$ remaining rows.

Trace

The *trace* of a matrix is the sum of its diagonal elements

$$\text{tr}(\mathcal{A}) = \sum_{i=1}^{p} a_{ii}.$$

Determinant

The *determinant* is an important concept of matrix algebra. For a square matrix $\mathcal{A}$, it is defined as:

$$\det(\mathcal{A}) = |\mathcal{A}| = \sum (-1)^{|\tau|}\, a_{1\tau(1)} \ldots a_{p\tau(p)},$$

the summation is over all permutations τ of $\{1, 2, \ldots, p\}$, and $|\tau| = 0$ if the permutation can be written as a product of an even number of transpositions and $|\tau| = 1$ otherwise.

Example 2.1 In the case of $p = 2$, $\mathcal{A} = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$ and we can permute the digits "1" and "2" once or not at all. So,

$$|\mathcal{A}| = a_{11}\, a_{22} - a_{12}\, a_{21}.$$

Transpose

For $\mathcal{A}(n \times p)$ and $\mathcal{B}(p \times n)$

$$(\mathcal{A}^\top)^\top = \mathcal{A}, \quad \text{and} \quad (\mathcal{A}\mathcal{B})^\top = \mathcal{B}^\top \mathcal{A}^\top.$$

Inverse

If $|\mathcal{A}| \neq 0$ and $\mathcal{A}(p \times p)$, then the inverse $\mathcal{A}^{-1}$ exists:

$$\mathcal{A}\, \mathcal{A}^{-1} = \mathcal{A}^{-1}\, \mathcal{A} = \mathcal{I}_p.$$

For small matrices, the inverse of $\mathcal{A} = (a_{ij})$ can be calculated as

$$\mathcal{A}^{-1} = \frac{\mathcal{C}}{|\mathcal{A}|},$$

where $\mathcal{C} = (c_{ij})$ is the adjoint matrix of $\mathcal{A}$. The elements c_{ji} of $\mathcal{C}^\top$ are the co-factors of $\mathcal{A}$:

$$c_{ji} = (-1)^{i+j} \begin{vmatrix} a_{11} & \ldots & a_{1(j-1)} & a_{1(j+1)} & \ldots & a_{1p} \\ \vdots & & & & & \\ a_{(i-1)1} & \ldots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \ldots & a_{(i-1)p} \\ a_{(i+1)1} & \ldots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \ldots & a_{(i+1)p} \\ \vdots & & & & & \\ a_{p1} & \ldots & a_{p(j-1)} & a_{p(j+1)} & \ldots & a_{pp} \end{vmatrix}.$$

G-inverse

A more general concept is the *G-inverse* (Generalised Inverse) $\mathcal{A}^-$ which satisfies the following:

$$\mathcal{A}\, \mathcal{A}^- \mathcal{A} = \mathcal{A}.$$

Later we will see that there may be more than one G-inverse.

Example 2.2 The generalised inverse can also be calculated for singular matrices. We have:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

which means that the generalised inverse of $\mathcal{A} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ is $\mathcal{A}^- = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ even though the inverse matrix of $\mathcal{A}$ does not exist in this case.

Eigenvalues, Eigenvectors

Consider a $(p \times p)$ matrix $\mathcal{A}$. If there a scalar λ and a vector γ exists such as

$$\mathcal{A}\gamma = \lambda\gamma, \tag{2.1}$$

then we call

λ an eigenvalue
γ an eigenvector.

It can be proven that an eigenvalue λ is a root of the p-th order polynomial $|\mathcal{A} - \lambda I_p| = 0$. Therefore, there are up to p eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $\mathcal{A}$. For each eigenvalue λ_j, a corresponding eigenvector γ_j exists given by equation (2.1). Suppose the matrix $\mathcal{A}$ has the eigenvalues $\lambda_1, \ldots, \lambda_p$. Let $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$.

The determinant $|\mathcal{A}|$ and the trace $\text{tr}(\mathcal{A})$ can be rewritten in terms of the eigenvalues:

$$|\mathcal{A}| = |\Lambda| = \prod_{j=1}^{p} \lambda_j \tag{2.2}$$

$$\text{tr}(\mathcal{A}) = \text{tr}(\Lambda) = \sum_{j=1}^{p} \lambda_j. \tag{2.3}$$

An idempotent matrix $\mathcal{A}$ (see the definition in Table 2.1) can only have eigenvalues in $\{0, 1\}$ therefore $\text{tr}(\mathcal{A}) = \text{rank}(\mathcal{A}) =$ number of eigenvalues $\neq 0$.

Example 2.3 Let us consider the matrix $\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. It is easy to verify that $\mathcal{A}\mathcal{A} = \mathcal{A}$ which implies that the matrix $\mathcal{A}$ is idempotent.

We know that the eigenvalues of an idempotent matrix are equal to 0 or 1. In this case, the eigenvalues of $\mathcal{A}$ are $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 = 0$ since

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} = 1 \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix} = 0 \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Using formulas (2.2) and (2.3), we can calculate the trace and the determinant of $\mathcal{A}$ from the eigenvalues: $\operatorname{tr}(\mathcal{A}) = \lambda_1 + \lambda_2 + \lambda_3 = 2$, $|\mathcal{A}| = \lambda_1\lambda_2\lambda_3 = 0$, and $\operatorname{rank}(\mathcal{A}) = 2$.

Properties of Matrix Characteristics

$\mathcal{A}(n \times n)$, $\mathcal{B}(n \times n)$, $c \in \mathbb{R}$

$$\operatorname{tr}(\mathcal{A} + \mathcal{B}) = \operatorname{tr}\mathcal{A} + \operatorname{tr}\mathcal{B} \tag{2.4}$$

$$\operatorname{tr}(c\mathcal{A}) = c \operatorname{tr}\mathcal{A} \tag{2.5}$$

$$|c\mathcal{A}| = c^n|\mathcal{A}| \tag{2.6}$$

$$|\mathcal{A}\mathcal{B}| = |\mathcal{B}\mathcal{A}| = |\mathcal{A}||\mathcal{B}| \tag{2.7}$$

$\mathcal{A}(n \times p)$, $\mathcal{B}(p \times n)$

$$\operatorname{tr}(\mathcal{A}\cdot\mathcal{B}) = \operatorname{tr}(\mathcal{B}\cdot\mathcal{A}) \tag{2.8}$$

$$\operatorname{rank}(\mathcal{A}) \leq \min(n, p)$$

$$\operatorname{rank}(\mathcal{A}) \geq 0 \tag{2.9}$$

$$\operatorname{rank}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}^\top) \tag{2.10}$$

$$\operatorname{rank}(\mathcal{A}^\top A) = \operatorname{rank}(\mathcal{A}) \tag{2.11}$$

$$\operatorname{rank}(\mathcal{A} + \mathcal{B}) \leq \operatorname{rank}(\mathcal{A}) + \operatorname{rank}(\mathcal{B}) \tag{2.12}$$

$$\operatorname{rank}(\mathcal{A}\mathcal{B}) \leq \min\{\operatorname{rank}(\mathcal{A}), \operatorname{rank}(\mathcal{B})\} \tag{2.13}$$

$\mathcal{A}(n \times p)$, $\mathcal{B}(p \times q)$, $\mathcal{C}(q \times n)$

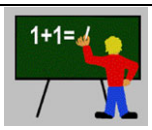
$$\begin{aligned} \operatorname{tr}(\mathcal{A}\mathcal{B}\mathcal{C}) &= \operatorname{tr}(\mathcal{B}\mathcal{C}\mathcal{A}) \\ &= \operatorname{tr}(\mathcal{C}\mathcal{A}\mathcal{B}) \end{aligned} \tag{2.14}$$

$$\operatorname{rank}(\mathcal{A}\mathcal{B}\mathcal{C}) = \operatorname{rank}(\mathcal{B}) \quad \text{for nonsingular } \mathcal{A}, \mathcal{C} \tag{2.15}$$

$\mathcal{A}(p \times p)$

$$|\mathcal{A}^{-1}| = |\mathcal{A}|^{-1} \tag{2.16}$$

$$\operatorname{rank}(\mathcal{A}) = p \quad \text{if and only if } \mathcal{A} \text{ is nonsingular.} \tag{2.17}$$



Summary

- ↪ The determinant $|\mathcal{A}|$ is the product of the eigenvalues of $\mathcal{A}$.
- ↪ The inverse of a matrix $\mathcal{A}$ exists if $|\mathcal{A}| \neq 0$.
- ↪ The trace $\text{tr}(\mathcal{A})$ is the sum of the eigenvalues of $\mathcal{A}$.
- ↪ The sum of the traces of two matrices equals the trace of the sum of the two matrices.
- ↪ The trace $\text{tr}(\mathcal{AB})$ equals $\text{tr}(\mathcal{BA})$.
- ↪ The $\text{rank}(\mathcal{A})$ is the maximal number of linearly independent rows (columns) of $\mathcal{A}$.

2.2 Spectral Decompositions

The computation of eigenvalues and eigenvectors is an important issue in the analysis of matrices. The spectral decomposition or Jordan decomposition links the structure of a matrix to the eigenvalues and the eigenvectors.

Theorem 2.1 (Jordan Decomposition) *Each symmetric matrix $\mathcal{A}(p \times p)$ can be written as*

$$\mathcal{A} = \Gamma\ \Lambda\ \Gamma^\top = \sum_{j=1}^{p} \lambda_j \gamma_j \gamma_j^\top \tag{2.18}$$

where

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$$

and where

$$\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$$

is an orthogonal matrix consisting of the eigenvectors γ_j of $\mathcal{A}$.

Example 2.4 Suppose that $\mathcal{A} = \left(\begin{smallmatrix} 1 & 2 \\ 2 & 3 \end{smallmatrix}\right)$. The eigenvalues are found by solving $|\mathcal{A} - \lambda\mathcal{I}| = 0$. This is equivalent to

$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & 3-\lambda \end{vmatrix} = (1-\lambda)(3-\lambda) - 4 = 0.$$

Hence, the eigenvalues are $\lambda_1 = 2 + \sqrt{5}$ and $\lambda_2 = 2 - \sqrt{5}$. The eigenvectors are $\gamma_1 = (0.5257, 0.8506)^\top$ and $\gamma_2 = (0.8506, -0.5257)^\top$. They are orthogonal since $\gamma_1^\top \gamma_2 = 0$.

Using spectral decomposition, we can define powers of a matrix $\mathcal{A}(p \times p)$. Suppose $\mathcal{A}$ is a symmetric matrix with positive eigenvalues. Then by Theorem 2.1

$$\mathcal{A} = \Gamma \Lambda \Gamma^\top,$$

and we define for some $\alpha \in \mathbb{R}$

$$\mathcal{A}^\alpha = \Gamma \Lambda^\alpha \Gamma^\top, \tag{2.19}$$

where $\Lambda^\alpha = \text{diag}(\lambda_1^\alpha, \ldots, \lambda_p^\alpha)$. In particular, we can easily calculate the inverse of the matrix $\mathcal{A}$. Suppose that the eigenvalues of $\mathcal{A}$ are positive. Then with $\alpha = -1$, we obtain the inverse of $\mathcal{A}$ from

$$\mathcal{A}^{-1} = \Gamma \Lambda^{-1} \Gamma^\top. \tag{2.20}$$

Another interesting decomposition which is later used is given in the following theorem.

Theorem 2.2 (Singular Value Decomposition) *Each matrix $\mathcal{A}(n \times p)$ with rank r can be decomposed as*

$$\mathcal{A} = \Gamma \ \Lambda \ \Delta^\top,$$

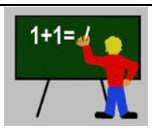
where $\Gamma(n \times r)$ and $\Delta(p \times r)$. Both Γ and Δ are column orthonormal, i.e., $\Gamma^\top \Gamma = \Delta^\top \Delta = \mathcal{I}_r$ and $\Lambda = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2})$, $\lambda_j > 0$. The values $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of the matrices $\mathcal{A}\mathcal{A}^\top$ and $\mathcal{A}^\top \mathcal{A}$. Γ and Δ consist of the corresponding r eigenvectors of these matrices.

This is obviously a generalisation of Theorem 2.1 (Jordan decomposition). With Theorem 2.2, we can find a G-inverse $\mathcal{A}^-$ of $\mathcal{A}$. Indeed, define $\mathcal{A}^- = \Delta \ \Lambda^{-1} \ \Gamma^\top$. Then $\mathcal{A} \ \mathcal{A}^- \ \mathcal{A} = \Gamma \ \Lambda \ \Delta^\top = \mathcal{A}$. Note that the G-inverse is not unique.

Example 2.5 In Example 2.2, we showed that the generalised inverse of $\mathcal{A} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ is $\mathcal{A}^- \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. The following also holds

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

which means that the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 8 \end{smallmatrix}\right)$ is also a generalised inverse of $\mathcal{A}$.



Summary

- ↪ The Jordan decomposition gives a representation of a symmetric matrix in terms of eigenvalues and eigenvectors.
- ↪ The eigenvectors belonging to the largest eigenvalues indicate the "main direction" of the data.
- ↪ The Jordan decomposition allows one to easily compute the power of a symmetric matrix $\mathcal{A}$: $\mathcal{A}^\alpha = \Gamma \Lambda^\alpha \Gamma^\top$.
- ↪ The singular value decomposition (SVD) is a generalisation of the Jordan decomposition to non-quadratic matrices.

2.3 Quadratic Forms

A quadratic form $Q(x)$ is built from a symmetric matrix $\mathcal{A}(p \times p)$ and a vector $x \in \mathbb{R}^p$:

$$Q(x) = x^\top \mathcal{A}\, x = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j. \tag{2.21}$$

Definiteness of Quadratic Forms and Matrices

$$\begin{array}{lll} Q(x) > 0 & \text{for all } x \neq 0 & \textit{positive definite} \\ Q(x) \geq 0 & \text{for all } x \neq 0 & \textit{positive semidefinite} \end{array}$$

A matrix $\mathcal{A}$ is called positive definite (semidefinite) if the corresponding quadratic form $Q(.)$ is positive definite (semidefinite). We write $\mathcal{A} > 0$ (≥ 0).

Quadratic forms can always be diagonalized, as the following result shows.

Theorem 2.3 *If $\mathcal{A}$ is symmetric and $Q(x) = x^\top \mathcal{A} x$ is the corresponding quadratic form, then there exists a transformation $x \mapsto \Gamma^\top x = y$ such that*

$$x^\top \mathcal{A}\, x = \sum_{i=1}^{p} \lambda_i y_i^2,$$

where λ_i are the eigenvalues of $\mathcal{A}$.

Proof $\mathcal{A} = \Gamma\ \Lambda\ \Gamma^\top$. By Theorem 2.1 and $y = \Gamma^\top \alpha$ we have that $x^\top \mathcal{A} x = x^\top \Gamma \Lambda \Gamma^\top x = y^\top \Lambda y = \sum_{i=1}^{p} \lambda_i y_i^2$. □

Positive definiteness of quadratic forms can be deduced from positive eigenvalues.

Theorem 2.4 *$\mathcal{A} > 0$ if and only if all $\lambda_i > 0, i = 1, \ldots, p$.*

Proof $0 < \lambda_1 y_1^2 + \cdots + \lambda_p y_p^2 = x^\top \mathcal{A} x$ for all $x \neq 0$ by Theorem 2.3. □

Corollary 2.1 *If $\mathcal{A} > 0$, then $\mathcal{A}^{-1}$ exists and $|\mathcal{A}| > 0$.*

Example 2.6 The quadratic form $Q(x) = x_1^2 + x_2^2$ corresponds to the matrix $\mathcal{A} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ with eigenvalues $\lambda_1 = \lambda_2 = 1$ and is thus positive definite. The quadratic form $Q(x) = (x_1 - x_2)^2$ corresponds to the matrix $\mathcal{A} = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$ with eigenvalues $\lambda_1 = 2, \lambda_2 = 0$ and is positive semidefinite. The quadratic form $Q(x) = x_1^2 - x_2^2$ with eigenvalues $\lambda_1 = 1, \lambda_2 = -1$ is indefinite.

In the statistical analysis of multivariate data, we are interested in maximising quadratic forms given some constraints.

Theorem 2.5 *If $\mathcal{A}$ and $\mathcal{B}$ are symmetric and $\mathcal{B} > 0$, then the maximum of $\frac{x^\top \mathcal{A} x}{x^\top \mathcal{B} x}$ is given by the largest eigenvalue of $\mathcal{B}^{-1}\mathcal{A}$. More generally,*

$$\max_x \frac{x^\top \mathcal{A} x}{x^\top \mathcal{B} x} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_x \frac{x^\top \mathcal{A} x}{x^\top \mathcal{B} x},$$

where $\lambda_1, \ldots, \lambda_p$ denote the eigenvalues of $\mathcal{B}^{-1}\mathcal{A}$. The vector which maximises (minimises) $\frac{x^\top \mathcal{A} x}{x^\top \mathcal{B} x}$ is the eigenvector of $\mathcal{B}^{-1}\mathcal{A}$ which corresponds to the largest (smallest) eigenvalue of $\mathcal{B}^{-1}\mathcal{A}$. If $x^\top \mathcal{B} x = 1$, we get

$$\max_x x^\top \mathcal{A} x = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_x x^\top \mathcal{A} x.$$

Proof By definition, $\mathcal{B}^{1/2} = \Gamma_{\mathcal{B}} \Lambda_{\mathcal{B}}^{1/2} \Gamma_{\mathcal{B}}^\top$ is symmetric. Then $x^\top \mathcal{B} x = \|x^\top \mathcal{B}^{1/2}\|^2 = \|\mathcal{B}^{1/2} x\|^2$. Set $y = \frac{\mathcal{B}^{1/2} x}{\|\mathcal{B}^{1/2} x\|}$, then

$$\max_x \frac{x^\top \mathcal{A} x}{x^\top \mathcal{B} x} = \max_{\{y : y^\top y = 1\}} y^\top \mathcal{B}^{-1/2} \mathcal{A} \mathcal{B}^{-1/2} y. \tag{2.22}$$

From Theorem 2.1, let

$$\mathcal{B}^{-1/2} \mathcal{A} \mathcal{B}^{-1/2} = \Gamma \Lambda \Gamma^\top$$

be the spectral decomposition of $\mathcal{B}^{-1/2} \mathcal{A} \mathcal{B}^{-1/2}$. Set

$$z = \Gamma^\top y, \quad \text{then} \quad z^\top z = y^\top \Gamma \Gamma^\top y = y^\top y.$$

Thus (2.22) is equivalent to

$$\max_{\{z:z^\top z=1\}} z^\top \Lambda z = \max_{\{z:z^\top z=1\}} \sum_{i=1}^{p} \lambda_i z_i^2.$$

But

$$\max_z \sum \lambda_i z_i^2 \le \lambda_1 \underbrace{\max_z \sum z_i^2}_{=1} = \lambda_1.$$

The maximum is thus obtained by $z = (1, 0, \ldots, 0)^\top$, i.e.,

$$y = \gamma_1, \quad \text{hence} \quad x = \mathcal{B}^{-1/2}\gamma_1.$$

Since $\mathcal{B}^{-1}\mathcal{A}$ and $\mathcal{B}^{-1/2}\,\mathcal{A}\,\mathcal{B}^{-1/2}$ have the same eigenvalues, the proof is complete.

To maximise (minimise) $x^\top \mathcal{A}x$ under $x^\top \mathcal{B}x = 1$, below is another proof using the Lagrange method.

$$\max_x x^\top \mathcal{A}x = \max_x [x^\top \mathcal{A}x - \lambda(x^\top \mathcal{B}x - 1)].$$

The first derivative of it in respect to x, is equal to 0:

$$2\mathcal{A}x - 2\lambda\mathcal{B}x = 0,$$

so

$$\mathcal{B}^{-1}\mathcal{A}x = \lambda x.$$

By the definition of eigenvector and eigenvalue, our maximiser x^* is $\mathcal{B}^{-1}\mathcal{A}$'s eigenvector corresponding to eigenvalue λ. So

$$\max_{\{x:x^\top \mathcal{B}x=1\}} x^\top \mathcal{A}x = \max_{\{x:x^\top \mathcal{B}x=1\}} x^\top \mathcal{B}\mathcal{B}^{-1}\mathcal{A}x = \max_{\{x:x^\top \mathcal{B}x=1\}} x^\top \mathcal{B}\lambda x = \max \lambda$$

which is just the maximum eigenvalue of $\mathcal{B}^{-1}\mathcal{A}$, and we choose the corresponding eigenvector as our maximiser x^*. □

Example 2.7 Consider the following matrices

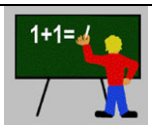
$$\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We calculate

$$\mathcal{B}^{-1}\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

The biggest eigenvalue of the matrix $\mathcal{B}^{-1}\mathcal{A}$ is $2+\sqrt{5}$. This means that the maximum of $x^\top \mathcal{A}x$ under the constraint $x^\top \mathcal{B}x = 1$ is $2 + \sqrt{5}$.

Notice that the constraint $x^\top \mathcal{B}x = 1$ corresponds, with our choice of $\mathcal{B}$, to the points which lie on the unit circle $x_1^2 + x_2^2 = 1$.



Summary

↪ A quadratic form can be described by a symmetric matrix $\mathcal{A}$.

↪ Quadratic forms can always be diagonalised.

↪ Positive definiteness of a quadratic form is equivalent to positiveness of the eigenvalues of the matrix $\mathcal{A}$.

↪ The maximum and minimum of a quadratic form given some constraints can be expressed in terms of eigenvalues.

2.4 Derivatives

For later sections of this book, it will be useful to introduce matrix notation for derivatives of a scalar function of a vector x with respect to x. Consider $f : \mathbb{R}^p \to \mathbb{R}$ and a $(p \times 1)$ vector x, then $\frac{\partial f(x)}{\partial x}$ is the column vector of partial derivatives $\{\frac{\partial f(x)}{\partial x_j}\}$, $j = 1, \ldots, p$ and $\frac{\partial f(x)}{\partial x^\top}$ is the row vector of the same derivative ($\frac{\partial f(x)}{\partial x}$ is called the *gradient* of f).

We can also introduce second order derivatives: $\frac{\partial^2 f(x)}{\partial x \partial x^\top}$ is the $(p \times p)$ matrix of elements $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$, $i = 1, \ldots, p$ and $j = 1, \ldots, p$. ($\frac{\partial^2 f(x)}{\partial x \partial x^\top}$ is called the *Hessian* of f.)

Suppose that a is a $(p \times 1)$ vector and that $\mathcal{A} = \mathcal{A}^\top$ is a $(p \times p)$ matrix. Then

$$\frac{\partial a^\top x}{\partial x} = \frac{\partial x^\top a}{\partial x} = a, \tag{2.23}$$

$$\frac{\partial x^\top \mathcal{A} x}{\partial x} = 2\mathcal{A}x. \tag{2.24}$$

The Hessian of the quadratic form $Q(x) = x^\top \mathcal{A} x$ is:

$$\frac{\partial^2 x^\top \mathcal{A} x}{\partial x \partial x^\top} = 2\mathcal{A}. \tag{2.25}$$

Example 2.8 Consider the matrix

$$\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

From formulas (2.24) and (2.25) it immediately follows that the gradient of $Q(x) = x^\top \mathcal{A} x$ is

$$\frac{\partial x^\top \mathcal{A} x}{\partial x} = 2\mathcal{A}x = 2\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} x = \begin{pmatrix} 2x & 4x \\ 4x & 6x \end{pmatrix}$$

and the Hessian is

$$\frac{\partial^2 x^\top \mathcal{A} x}{\partial x \partial x^\top} = 2\mathcal{A} = 2\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 4 & 6 \end{pmatrix}.$$

2.5 Partitioned Matrices

Very often we will have to consider certain groups of rows and columns of a matrix $\mathcal{A}(n \times p)$. In the case of two groups, we have

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix}$$

where $\mathcal{A}_{ij}(n_i \times p_j),\ i, j = 1, 2,\ n_1 + n_2 = n$ and $p_1 + p_2 = p$.

If $\mathcal{B}(n \times p)$ is partitioned accordingly, we have:

$$\mathcal{A} + \mathcal{B} = \begin{pmatrix} \mathcal{A}_{11} + \mathcal{B}_{11} & \mathcal{A}_{12} + \mathcal{B}_{12} \\ \mathcal{A}_{21} + \mathcal{B}_{21} & \mathcal{A}_{22} + \mathcal{B}_{22} \end{pmatrix}$$

$$\mathcal{B}^\top = \begin{pmatrix} \mathcal{B}_{11}^\top & \mathcal{B}_{21}^\top \\ \mathcal{B}_{12}^\top & \mathcal{B}_{22}^\top \end{pmatrix}$$

$$\mathcal{A}\mathcal{B}^\top = \begin{pmatrix} \mathcal{A}_{11}\mathcal{B}_{11}^\top + \mathcal{A}_{12}\mathcal{B}_{12}^\top & \mathcal{A}_{11}\mathcal{B}_{21}^\top + \mathcal{A}_{12}\mathcal{B}_{22}^\top \\ \mathcal{A}_{21}\mathcal{B}_{11}^\top + \mathcal{A}_{22}\mathcal{B}_{12}^\top & \mathcal{A}_{21}\mathcal{B}_{21}^\top + \mathcal{A}_{22}\mathcal{B}_{22}^\top \end{pmatrix}.$$

An important particular case is the square matrix $\mathcal{A}(p \times p)$, partitioned in such a way that $\mathcal{A}_{11}$ and $\mathcal{A}_{22}$ are both square matrices (i.e., $n_j = p_j,\ j = 1, 2$). It can be verified that when $\mathcal{A}$ is non-singular ($\mathcal{A}\mathcal{A}^{-1} = \mathcal{I}_p$):

$$\mathcal{A}^{-1} = \begin{pmatrix} \mathcal{A}^{11} & \mathcal{A}^{12} \\ \mathcal{A}^{21} & \mathcal{A}^{22} \end{pmatrix} \tag{2.26}$$

where

$$\begin{cases} \mathcal{A}^{11} = (\mathcal{A}_{11} - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21})^{-1} \stackrel{\text{def}}{=} (\mathcal{A}_{11\cdot 2})^{-1} \\ \mathcal{A}^{12} = -(\mathcal{A}_{11\cdot 2})^{-1}\mathcal{A}_{12}\mathcal{A}_{22}^{-1} \\ \mathcal{A}^{21} = -\mathcal{A}_{22}^{-1}\mathcal{A}_{21}(\mathcal{A}_{11\cdot 2})^{-1} \\ \mathcal{A}^{22} = \mathcal{A}_{22}^{-1} + \mathcal{A}_{22}^{-1}\mathcal{A}_{21}(\mathcal{A}_{11\cdot 2})^{-1}\mathcal{A}_{12}\mathcal{A}_{22}^{-1}. \end{cases}$$

An alternative expression can be obtained by reversing the positions of $\mathcal{A}_{11}$ and $\mathcal{A}_{22}$ in the original matrix.

The following results will be useful if $\mathcal{A}_{11}$ is non-singular:

$$|\mathcal{A}| = |\mathcal{A}_{11}||\mathcal{A}_{22} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{A}_{12}| = |\mathcal{A}_{11}||\mathcal{A}_{22\cdot 1}|. \tag{2.27}$$

If $\mathcal{A}_{22}$ is non-singular, we have that:

$$|\mathcal{A}| = |\mathcal{A}_{22}||\mathcal{A}_{11} - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}| = |\mathcal{A}_{22}||\mathcal{A}_{11\cdot 2}|. \tag{2.28}$$

A useful formula is derived from the alternative expressions for the inverse and the determinant. For instance let

$$\mathcal{B} = \begin{pmatrix} 1 & b^\top \\ a & \mathcal{A} \end{pmatrix}$$

where a and b are $(p \times 1)$ vectors and $\mathcal{A}$ is non-singular. We then have:

$$|\mathcal{B}| = |\mathcal{A} - ab^\top| = |\mathcal{A}||1 - b^\top \mathcal{A}^{-1} a| \tag{2.29}$$

and equating the two expressions for $\mathcal{B}^{22}$, we obtain the following:

$$(\mathcal{A} - ab^\top)^{-1} = \mathcal{A}^{-1} + \frac{\mathcal{A}^{-1} a b^\top \mathcal{A}^{-1}}{1 - b^\top \mathcal{A}^{-1} a}. \tag{2.30}$$

Example 2.9 Let's consider the matrix

$$\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}.$$

We can use formula (2.26) to calculate the inverse of a partitioned matrix, i.e., $\mathcal{A}^{11} = -1$, $\mathcal{A}^{12} = \mathcal{A}^{21} = 1$, $\mathcal{A}^{22} = -1/2$. The inverse of $\mathcal{A}$ is

$$\mathcal{A}^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & -0.5 \end{pmatrix}.$$

It is also easy to calculate the determinant of $\mathcal{A}$:

$$|\mathcal{A}| = |1||2 - 4| = -2.$$

Let $\mathcal{A}(n \times p)$ and $\mathcal{B}(p \times n)$ be any two matrices and suppose that $n \geq p$. From (2.27) and (2.28) we can conclude that

$$\begin{vmatrix} -\lambda \mathcal{I}_n & -\mathcal{A} \\ \mathcal{B} & \mathcal{I}_p \end{vmatrix} = (-\lambda)^{n-p} |\mathcal{B}\mathcal{A} - \lambda \mathcal{I}_p| = |\mathcal{A}\mathcal{B} - \lambda \mathcal{I}_n|. \tag{2.31}$$

Since both determinants on the right-hand side of (2.31) are polynomials in λ, we find that the n eigenvalues of $\mathcal{A}\mathcal{B}$ yield the p eigenvalues of $\mathcal{B}\mathcal{A}$ plus the eigenvalue 0, $n - p$ times.

The relationship between the eigenvectors is described in the next theorem.

Theorem 2.6 *For $\mathcal{A}(n \times p)$ and $\mathcal{B}(p \times n)$, the non-zero eigenvalues of $\mathcal{A}\mathcal{B}$ and $\mathcal{B}\mathcal{A}$ are the same and have the same multiplicity. If x is an eigenvector of $\mathcal{A}\mathcal{B}$ for an eigenvalue $\lambda \neq 0$, then $y = \mathcal{B}x$ is an eigenvector of $\mathcal{B}\mathcal{A}$.*

Corollary 2.2 *For $\mathcal{A}(n \times p)$, $\mathcal{B}(q \times n)$, $a(p \times 1)$, and $b(q \times 1)$ we have*

$$\operatorname{rank}(\mathcal{A} a b^\top \mathcal{B}) \leq 1.$$

The non-zero eigenvalue, if it exists, equals $b^\top \mathcal{B}\mathcal{A}a$ (with eigenvector $\mathcal{A}a$).

Proof Theorem 2.6 asserts that the eigenvalues of $\mathcal{A}ab^\top\mathcal{B}$ are the same as those of $b^\top\mathcal{B}\mathcal{A}a$. Note that the matrix $b^\top\mathcal{B}\mathcal{A}a$ is a scalar and hence it is its own eigenvalue λ_1.

Applying $\mathcal{A}ab^\top\mathcal{B}$ to $\mathcal{A}a$ yields

$$(\mathcal{A}ab^\top\mathcal{B})(\mathcal{A}a) = (\mathcal{A}a)(b^\top\mathcal{B}\mathcal{A}a) = \lambda_1\mathcal{A}a. \qquad \square$$

2.6 Geometrical Aspects

Distance

Let $x, y \in \mathbb{R}^p$. A distance d is defined as a function

$$d : \mathbb{R}^{2p} \to \mathbb{R}_+ \quad \text{which fulfills} \quad \begin{cases} d(x,y) > 0 & \forall x \neq y \\ d(x,y) = 0 & \text{if and only if } x = y \\ d(x,y) \leq d(x,z) + d(z,y) & \forall x, y, z. \end{cases}$$

A *Euclidean distance* d between two points x and y is defined as

$$d^2(x,y) = (x-y)^T\mathcal{A}(x-y) \tag{2.32}$$

where $\mathcal{A}$ is a positive definite matrix ($\mathcal{A} > 0$). $\mathcal{A}$ is called a *metric*.

Example 2.10 A particular case is when $\mathcal{A} = \mathcal{I}_p$, i.e.,

$$d^2(x,y) = \sum_{i=1}^{p}(x_i - y_i)^2. \tag{2.33}$$

Figure 2.1 illustrates this definition for $p = 2$.

Note that the sets $E_d = \{x \in \mathbb{R}^p \mid (x-x_0)^\top(x-x_0) = d^2\}$, i.e., the spheres with radius d and centre x_0, are the Euclidean $\mathcal{I}_p$ *iso-distance* curves from the point x_0 (see Figure 2.2).

The more general distance (2.32) with a positive definite matrix $\mathcal{A}$ ($\mathcal{A} > 0$) leads to the iso-distance curves

$$E_d = \{x \in \mathbb{R}^p \mid (x-x_0)^\top\mathcal{A}(x-x_0) = d^2\}, \tag{2.34}$$

i.e., ellipsoids with centre x_0, matrix $\mathcal{A}$ and constant d (see Figure 2.3).

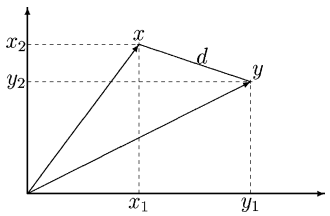


Fig. 2.1 Distance d

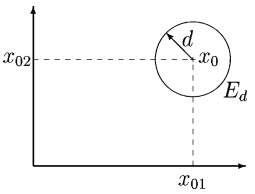


Fig. 2.2 Iso-distance sphere

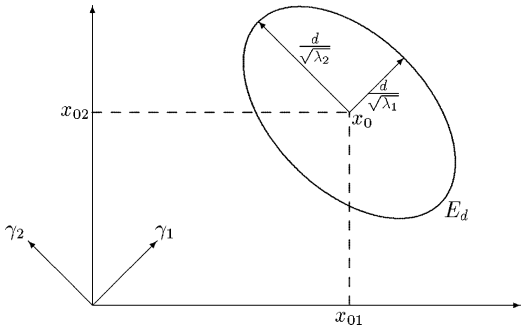


Fig. 2.3 Iso-distance ellipsoid

Let $\gamma_1, \gamma_2, \ldots, \gamma_p$ be the orthonormal eigenvectors of $\mathcal{A}$ corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. The resulting observations are given in the next theorem.

Theorem 2.7

(i) *The principal axes of E_d are in the direction of γ_i; $i = 1, \ldots, p$.*

(ii) *The half-lengths of the axes are $\sqrt{\frac{d^2}{\lambda_i}}$; $i = 1, \ldots, p$.*

(iii) *The rectangle surrounding the ellipsoid E_d is defined by the following inequalities*:

$$x_{0i} - \sqrt{d^2 a^{ii}} \leq x_i \leq x_{0i} + \sqrt{d^2 a^{ii}}, \quad i = 1, \ldots, p,$$

where a^{ii} is the (i, i) element of $\mathcal{A}^{-1}$. By the rectangle surrounding the ellipsoid E_d we mean the rectangle whose sides are parallel to the coordinate axis.

It is easy to find the coordinates of the tangency points between the ellipsoid and its surrounding rectangle parallel to the coordinate axes. Let us find the coordinates of the tangency point that are in the direction of the j-th coordinate axis (positive direction).

For ease of notation, we suppose the ellipsoid is centred around the origin ($x_0 = 0$). If not, the rectangle will be shifted by the value of x_0.

The coordinate of the tangency point is given by the solution to the following problem:

$$x = \arg \max_{x^\top \mathcal{A} x = d^2} e_j^\top x \tag{2.35}$$

where $e_j^\top$ is the j-th column of the identity matrix $\mathcal{I}_p$. The coordinate of the tangency point in the negative direction would correspond to the solution of the min problem: by symmetry, it is the opposite value of the former.

The solution is computed via the Lagrangian $L = e_j^\top x - \lambda(x^\top \mathcal{A} x - d^2)$ which by (2.23) leads to the following system of equations:

$$\frac{\partial L}{\partial x} = e_j - 2\lambda \mathcal{A} x = 0 \tag{2.36}$$

$$\frac{\partial L}{\partial \lambda} = x^T \mathcal{A} x - d^2 = 0. \tag{2.37}$$

This gives $x = \frac{1}{2\lambda}\mathcal{A}^{-1} e_j$, or componentwise

$$x_i = \frac{1}{2\lambda} a^{ij}, \quad i = 1, \ldots, p \tag{2.38}$$

where a^{ij} denotes the (i, j)-th element of $\mathcal{A}^{-1}$.

Premultiplying (2.36) by $x^\top$, we have from (2.37):

$$x_j = 2\lambda d^2.$$

Comparing this to the value obtained by (2.38), for $i = j$ we obtain $2\lambda = \sqrt{\frac{a^{jj}}{d^2}}$. We choose the positive value of the square root because we are maximising $e_j^\top x$. A minimum would correspond to the negative value. Finally, we have the coordinates of the tangency point between the ellipsoid and its surrounding rectangle in the positive direction of the j-th axis:

$$x_i = \sqrt{\frac{d^2}{a^{jj}}}\, a^{ij}, \quad i = 1, \ldots, p. \tag{2.39}$$

The particular case where $i = j$ provides statement (iii) in Theorem 2.7.

Remark: Usefulness of Theorem 2.7

Theorem 2.7 will prove to be particularly useful in many subsequent chapters. First, it provides a helpful tool for graphing an ellipse in two dimensions. Indeed, knowing the slope of the principal axes of the ellipse, their half-lengths and drawing the rectangle inscribing the ellipse, allows one to quickly draw a rough picture of the shape of the ellipse.

In Chapter 7, it is shown that the confidence region for the vector μ of a multivariate normal population is given by a particular ellipsoid whose parameters depend on sample characteristics. The rectangle inscribing the ellipsoid (which is much easier to obtain) will provide the simultaneous confidence intervals for all of the components in μ.

In addition it will be shown that the contour surfaces of the multivariate normal density are provided by ellipsoids whose parameters depend on the mean vector

and on the covariance matrix. We will see that the tangency points between the contour ellipsoids and the surrounding rectangle are determined by regressing one component on the $(p-1)$ other components. For instance, in the direction of the j-th axis, the tangency points are given by the intersections of the ellipsoid contours with the regression line of the vector of $(p-1)$ variables (all components except the j-th) on the j-th component.

Norm of a Vector

Consider a vector $x \in \mathbb{R}^p$. The norm or length of x (with respect to the metric $\mathcal{I}_p$) is defined as

$$\|x\| = d(0,x) = \sqrt{x^\top x}.$$

If $\|x\| = 1, x$ is called a *unit vector*. A more general norm can be defined with respect to the metric $\mathcal{A}$:

$$\|x\|_{\mathcal{A}} = \sqrt{x^\top \mathcal{A} x}.$$

Angle Between Two Vectors

Consider two vectors x and $y \in \mathbb{R}^p$. The angle θ between x and y is defined by the cosine of θ:

$$\cos\theta = \frac{x^\top y}{\|x\|\ \|y\|}, \tag{2.40}$$

see Figure 2.4. Indeed for $p=2$, $x = \binom{x_1}{x_2}$ and $y = \binom{y_1}{y_2}$, we have

$$\begin{aligned} \|x\|\cos\theta_1 &= x_1; \qquad \|y\|\cos\theta_2 = y_1 \\ \|x\|\sin\theta_1 &= x_2; \qquad \|y\|\sin\theta_2 = y_2, \end{aligned} \tag{2.41}$$

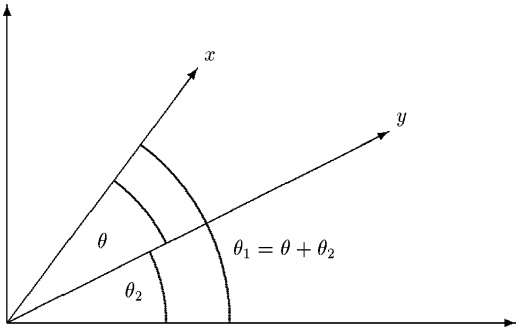
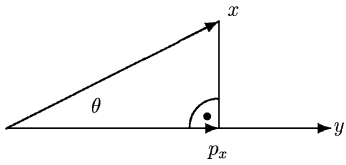


Fig. 2.4 Angle between vectors

Fig. 2.5 Projection



therefore,

$$\cos\theta = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 = \frac{x_1 y_1 + x_2 y_2}{\|x\|\ \|y\|} = \frac{x^\top y}{\|x\|\ \|y\|}\ .$$

Remark 2.1 If $x^\top y = 0$, then the angle θ is equal to $\frac{\pi}{2}$. From trigonometry, we know that the cosine of θ equals the length of the base of a triangle ($||p_x||$) divided by the length of the hypotenuse ($||x||$). Hence, we have

$$||p_x|| = ||x|| |\cos\theta| = \frac{|x^\top y|}{\|y\|}, \tag{2.42}$$

where p_x is the projection of x on y (which is defined below). It is the coordinate of x on the y vector, see Figure 2.5.

The angle can also be defined with respect to a general metric $\mathcal{A}$

$$\cos\theta = \frac{x^\top \mathcal{A} y}{\|x\|_\mathcal{A}\ \|y\|_\mathcal{A}}. \tag{2.43}$$

If $\cos\theta = 0$ then x is orthogonal to y with respect to the metric $\mathcal{A}$.

Example 2.11 Assume that there are two centred (i.e., zero mean) data vectors. The cosine of the angle between them is equal to their correlation (defined in (3.8)). Indeed for x and y with $\overline{x} = \overline{y} = 0$ we have

$$r_{XY} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} = \cos\theta$$

according to formula (2.40).

Rotations

When we consider a point $x \in \mathbb{R}^p$, we generally use a p-coordinate system to obtain its geometric representation, like in Figure 2.1 for instance. There will be situations in multivariate techniques where we will want to rotate this system of coordinates by the angle θ.

Consider for example the point P with coordinates $x = (x_1, x_2)^\top$ in $\mathbb{R}^2$ with respect to a given set of orthogonal axes. Let Γ be a (2×2) orthogonal matrix where

$$\Gamma = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{2.44}$$

If the axes are rotated about the origin through an angle θ in a clockwise direction, the new coordinates of P will be given by the vector y

$$y = \Gamma x, \tag{2.45}$$

and a rotation through the same angle in a anti-clockwise direction gives the new coordinates as

$$y = \Gamma^\top x. \tag{2.46}$$

More generally, premultiplying a vector x by an orthogonal matrix Γ geometrically corresponds to a rotation of the system of axes, so that the first new axis is determined by the first row of Γ. This geometric point of view will be exploited in Chapters 10 and 11.

Column Space and Null Space of a Matrix

Define for $\mathcal{X}(n \times p)$

$$Im(\mathcal{X}) \stackrel{\text{def}}{=} C(\mathcal{X}) = \{x \in \mathbb{R}^n \mid \exists a \in \mathbb{R}^p \text{ so that } \mathcal{X}a = x\},$$

the space generated by the columns of $\mathcal{X}$ or the *column space* of $\mathcal{X}$. Note that $C(\mathcal{X}) \subseteq \mathbb{R}^n$ and $\dim\{C(\mathcal{X})\} = \text{rank}(\mathcal{X}) = r \leq \min(n, p)$.

$$Ker(\mathcal{X}) \stackrel{\text{def}}{=} N(\mathcal{X}) = \{y \in \mathbb{R}^p \mid \mathcal{X}y = 0\}$$

is the *null space* of $\mathcal{X}$. Note that $N(\mathcal{X}) \subseteq \mathbb{R}^p$ and that $\dim\{N(\mathcal{X})\} = p - r$.

Remark 2.2 $N(\mathcal{X}^\top)$ is the orthogonal complement of $C(\mathcal{X})$ in $\mathbb{R}^n$, i.e., given a vector $b \in \mathbb{R}^n$ it will hold that $x^\top b = 0$ for all $x \in C(\mathcal{X})$, if and only if $b \in N(\mathcal{X}^\top)$.

Example 2.12 Let $\mathcal{X} = \begin{pmatrix} 2 & 3 & 5 \\ 4 & 6 & 7 \\ 6 & 8 & 6 \\ 8 & 2 & 4 \end{pmatrix}$. It is easy to show (e.g. by calculating the determinant of $\mathcal{X}$) that $\text{rank}(\mathcal{X}) = 3$. Hence, the columns space of $\mathcal{X}$ is $C(\mathcal{X}) = \mathbb{R}^3$. The null space of $\mathcal{X}$ contains only the zero vector $(0, 0, 0)^\top$ and its dimension is equal to $\text{rank}(\mathcal{X}) - 3 = 0$.

For $\mathcal{X} = \begin{pmatrix} 2 & 3 & 1 \\ 4 & 6 & 2 \\ 6 & 8 & 3 \\ 8 & 2 & 4 \end{pmatrix}$, the third column is a multiple of the first one and the matrix $\mathcal{X}$ cannot be of full rank. Noticing that the first two columns of $\mathcal{X}$ are independent, we see that $\text{rank}(\mathcal{X}) = 2$. In this case, the dimension of the columns space is 2 and the dimension of the null space is 1.

Projection Matrix

A matrix $\mathcal{P}(n \times n)$ is called an (orthogonal) projection matrix in $\mathbb{R}^n$ if and only if $\mathcal{P} = \mathcal{P}^\top = \mathcal{P}^2$ ($\mathcal{P}$ is idempotent). Let $b \in \mathbb{R}^n$. Then $a = \mathcal{P}b$ is the projection of b on $C(\mathcal{P})$.

Projection on $C(\mathcal{X})$

Consider $\mathcal{X}(n \times p)$ and let

$$\mathcal{P} = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{X}^\top \tag{2.47}$$

and $\mathcal{Q} = \mathcal{I}_n - \mathcal{P}$. It's easy to check that $\mathcal{P}$ and $\mathcal{Q}$ are idempotent and that

$$\mathcal{P}\mathcal{X} = \mathcal{X} \quad \text{and} \quad \mathcal{Q}\mathcal{X} = 0. \tag{2.48}$$

Since the columns of $\mathcal{X}$ are projected onto themselves, the projection matrix $\mathcal{P}$ projects any vector $b \in \mathbb{R}^n$ onto $C(\mathcal{X})$. Similarly, the projection matrix $\mathcal{Q}$ projects any vector $b \in \mathbb{R}^n$ onto the orthogonal complement of $C(\mathcal{X})$.

Theorem 2.8 *Let $\mathcal{P}$ be the projection (2.47) and $\mathcal{Q}$ its orthogonal complement. Then*:

(i) $x = \mathcal{P}b$ *entails* $x \in C(\mathcal{X})$,
(ii) $y = \mathcal{Q}b$ *means that* $y^\top x = 0 \; \forall x \in C(\mathcal{X})$.

Proof (i) holds, since $x = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{X}^\top b = \mathcal{X}a$, where $a = (\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{X}^\top b \in \mathbb{R}^p$.

(ii) follows from $y = b - \mathcal{P}b$ and $x = \mathcal{X}a$. Hence $y^\top x = b^\top \mathcal{X}a - b^\top \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{X}a = 0$. □

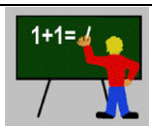
Remark 2.3 Let $x, y \in \mathbb{R}^n$ and consider $p_x \in \mathbb{R}^n$, the projection of x on y (see Figure 2.5). With $\mathcal{X} = y$ we have from (2.47)

$$p_x = y(y^\top y)^{-1} y^\top x = \frac{y^\top x}{\|y\|^2} \, y \tag{2.49}$$

and we can easily verify that

$$\|p_x\| = \sqrt{p_x^\top p_x} = \frac{|y^\top x|}{\|y\|}.$$

See again Remark 2.1.



Summary

- A distance between two p-dimensional points x and y is a quadratic form $(x-y)^\top \mathcal{A}(x-y)$ in the vectors of differences $(x-y)$. A distance defines the norm of a vector.
- Iso-distance curves of a point x_0 are all those points that have the same distance from x_0. Iso-distance curves are ellipsoids whose principal axes are determined by the direction of the eigenvectors of $\mathcal{A}$. The half-length of principal axes is proportional to the inverse of the roots of the eigenvalues of $\mathcal{A}$.
- The angle between two vectors x and y is given by $\cos\theta = \frac{x^\top \mathcal{A} y}{\|x\|_\mathcal{A}\ \|y\|_\mathcal{A}}$ w.r.t. the metric $\mathcal{A}$.
- For the Euclidean distance with $\mathcal{A} = \mathcal{I}$ the correlation between two centred data vectors x and y is given by the cosine of the angle between them, i.e., $\cos\theta = r_{XY}$.
- The projection $\mathcal{P} = \mathcal{X}(\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top$ is the projection onto the column space $C(\mathcal{X})$ of $\mathcal{X}$.
- The projection of $x \in \mathbb{R}^n$ on $y \in \mathbb{R}^n$ is given by $p_x = \frac{y^\top x}{\|y\|^2} y$.

2.7 Exercises

Exercise 2.1 Compute the determinant for a (3×3) matrix.

Exercise 2.2 Suppose that $|\mathcal{A}| = 0$. Is it possible that all eigenvalues of $\mathcal{A}$ are positive?

Exercise 2.3 Suppose that all eigenvalues of some (square) matrix $\mathcal{A}$ are different from zero. Does the inverse $\mathcal{A}^{-1}$ of $\mathcal{A}$ exist?

Exercise 2.4 Write a program that calculates the Jordan decomposition of the matrix

$$\mathcal{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Check Theorem 2.1 numerically.

Exercise 2.5 Prove (2.23), (2.24) and (2.25).

Exercise 2.6 Show that a projection matrix only has eigenvalues in $\{0, 1\}$.

Exercise 2.7 Draw some iso-distance ellipsoids for the metric $\mathcal{A} = \Sigma^{-1}$ of Example 3.13.

Exercise 2.8 Find a formula for $|\mathcal{A} + aa^\top|$ and for $(\mathcal{A} + aa^\top)^{-1}$. (Hint: use the inverse partitioned matrix with $\mathcal{B} = \left(\begin{smallmatrix} 1 & -a^\top \\ a & \mathcal{A} \end{smallmatrix}\right)$.)

Exercise 2.9 Prove the Binomial inverse theorem for two non-singular matrices $\mathcal{A}(p \times p)$ and $\mathcal{B}(p \times p)$: $(\mathcal{A}+\mathcal{B})^{-1} = \mathcal{A}^{-1} - \mathcal{A}^{-1}(\mathcal{A}^{-1} + \mathcal{B}^{-1})^{-1}\mathcal{A}^{-1}$. (Hint: use (2.26) with $\mathcal{C} = \left(\begin{smallmatrix} \mathcal{A} & I_p \\ -I_p & \mathcal{B}^{-1} \end{smallmatrix}\right)$.)