2. Matrix Algebra and Random Vectors

2.1 Introduction

Multivariate data can be conveniently display as array of numbers. In general, a rectangular array of numbers with, for instance, n rows and p columns is called a *matrix* of dimension $n \times p$ The study of multivariate methods is greatly facilitated by the use of matrix algebra.

2.2 Some Basic of Matrix and Vector Algebra

Vectors

- Definition: An array **x** of n real number $x_1, x_2, \ldots, x_n$ is called a *vector*, and it is written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{or} \quad \mathbf{x}' = [x_1, x_2, \ldots, x_n]$$

where the prime denotes the operation of transposing a column to a row.

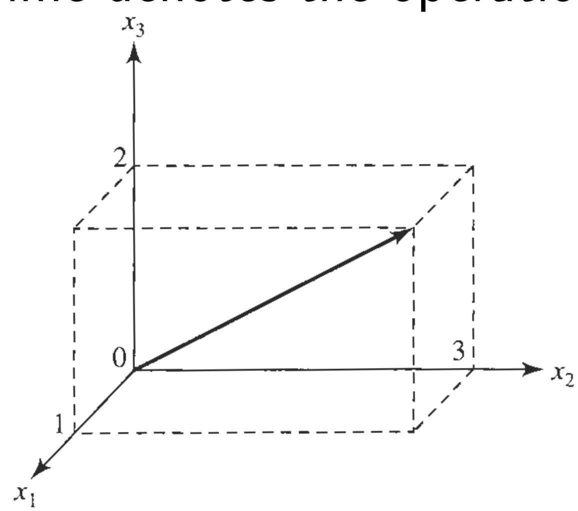

Figure 2.1 The vector $x' = [1, 3, 2]$.

- Multiplying vectors by a constant c:

$$c\mathbf{x} = \mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$$

- *Addition* of **x** and **y** is defined as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

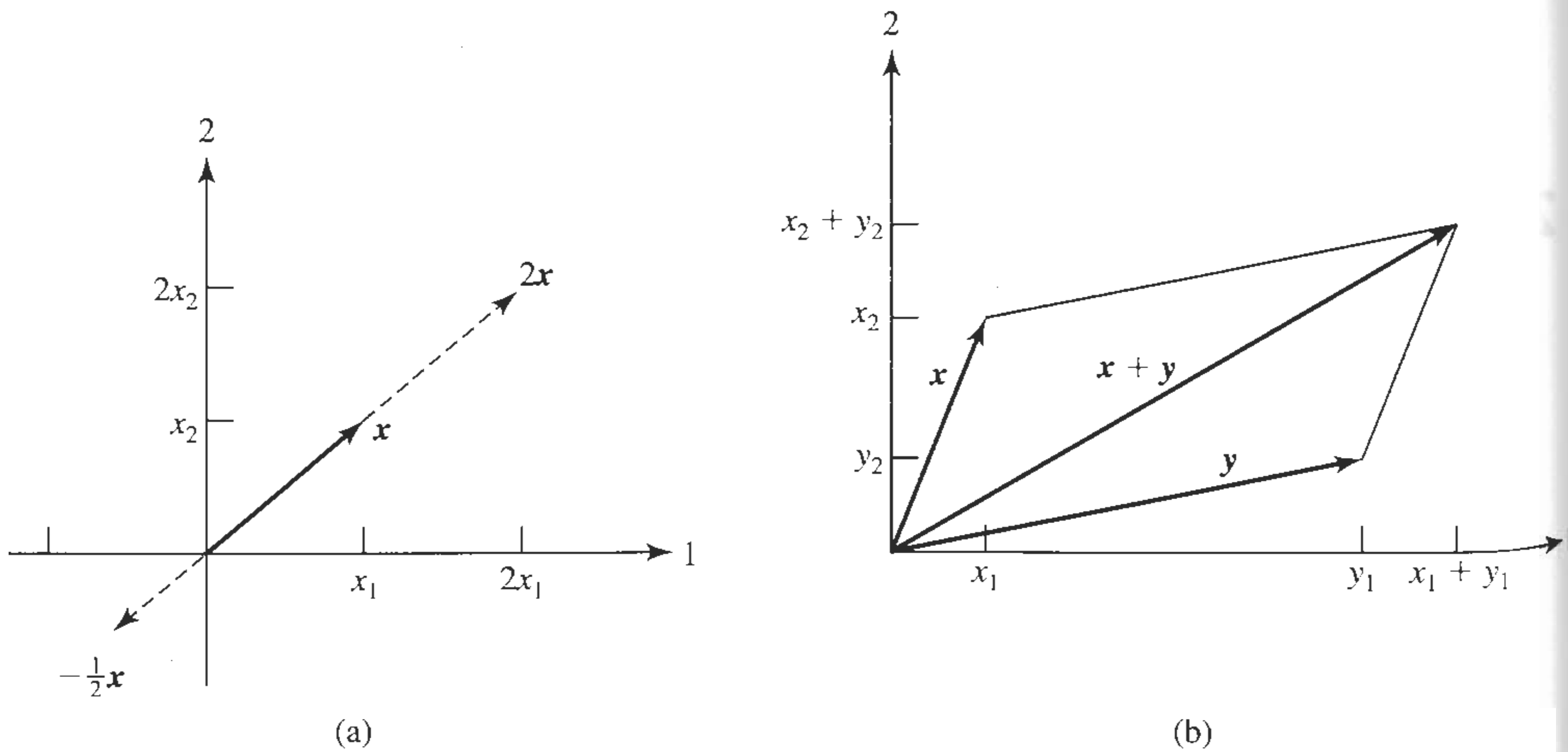

Figure 2.2 Scatter multiplication and vector addition

- Length of vectors, *unit vector*

  When $n = 2$, $\mathbf{x} = [x1, x2]'$, the length of $\mathbf{x}$, written $L_{\mathbf{x}}$ is defined to be

$$L_{\mathbf{x}} = \sqrt{x_1^2 + x_2^2}$$

  Geometrically, the length of a vector in two dimension can be viewed as the hypotenuse of a right triangle. The length of a vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]'$ and $c\mathbf{x} = [cx_1, cx_2, \ldots, cx_n]'$

$$L_{\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

$$L_{c\mathbf{x}} = \sqrt{c^2x_1^2 + c^2x_2^2 + \cdots + c^2x_n^2} = |c|\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = |c|L_{\mathbf{x}}$$

  Choosing $c = L_{\mathbf{x}}^{-1}$, we obtain the *unit vector* $L_{\mathbf{x}}^{-1}\mathbf{x}$, which has length 1 and lies in the direction of $\mathbf{x}$.

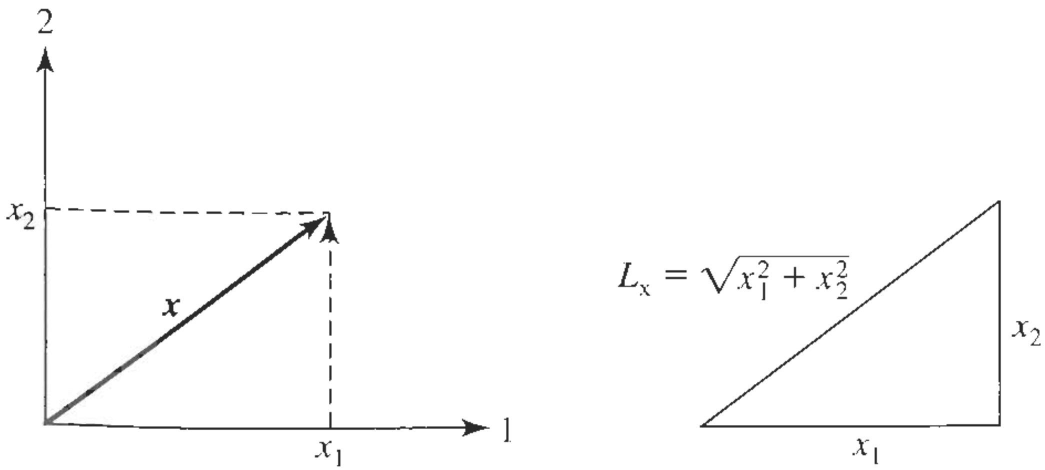

Figure 2.3
Length of $\mathbf{x} = \sqrt{x_1^2 + x_2^2}$.

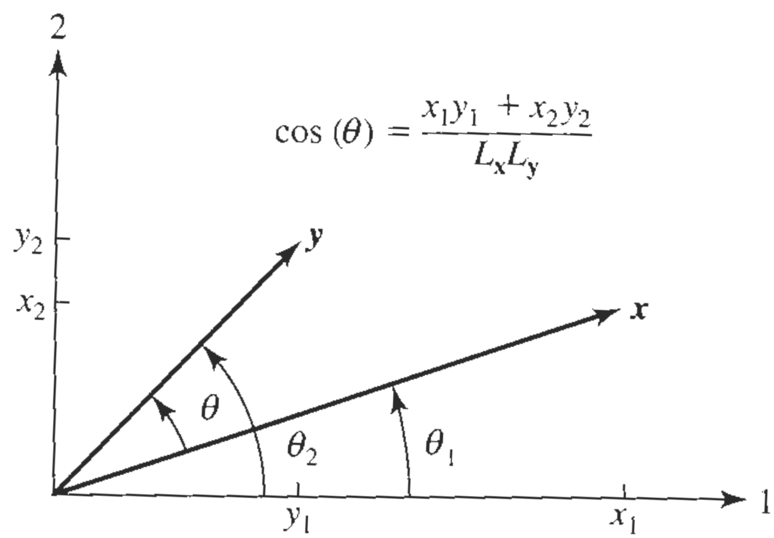

Figure 2.4 The angle θ between $\mathbf{x}' = [x_1, x_2]$ and $\mathbf{y}' = [y_1, y_2]$.

- *Angle, inner product. perpendicular*

  Consider two vectors $\mathbf{x}, \mathbf{y}$ in a plane and the angle θ between them, as in Figure 2.4. From the figure, θ can be represented as the difference the angle θ_1 and θ_2 formed by the two vectors and the first coordinate axis. Since, by the definition,

$$\cos(\theta_1) = \frac{x_1}{L_{\mathbf{x}}}, \cos(\theta_2) = \frac{y_1}{L_{\mathbf{y}}}$$

$$\sin(\theta_1) = \frac{x_2}{L_{\mathbf{x}}}, \sin(\theta_2) = \frac{y_2}{L_{\mathbf{y}}}$$

  and

$$\cos(\theta_2 - \theta_1) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2)$$

  the angle θ between the two vectors is specified by

$$\cos(\theta) = \cos(\theta_2 - \theta_1) = \frac{y_1}{L_{\mathbf{y}}} \cdot \frac{x_1}{L_{\mathbf{x}}} + \frac{y_2}{L_{\mathbf{y}}} \cdot \frac{x_2}{L_{\mathbf{x}}} = \frac{x_1 y_1 + x_2 y_2}{L_{\mathbf{x}} L_{\mathbf{y}}}.$$

- Definition of *inner product* of the two vectors **x** and **y**

$$\mathbf{x}'\mathbf{y} = x_1 y_1 + x_2 y_2.$$

With the definition of the inner product and $\cos(\theta)$,

$$L_{\mathbf{x}} = \sqrt{\mathbf{x}'\mathbf{x}}, \quad \cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}} L_{\mathbf{y}}} = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}.$$

Example 2.1.(Calculating lengths of vectors and the angle between them)
Given the vectors $\mathbf{x}' = [1\ 3\ 2]$ and $\mathbf{y}' = [-2\ 1\ \ -1]$, find $3\mathbf{x}$ and $\mathbf{x} + \mathbf{y}$. Next, determine the length of **x**, the length of $\mathbf{y}$, and the angle between **x** and $\mathbf{y}$. Also, check that the length of $3\mathbf{x}$ is three times the length of **x**

- A pair of vectors $\mathbf{x}$ and $\mathbf{y}$ of the same dimension is said to be *linearly dependent* if there exist constants c_1 and c_2, both not zero, such that $c_1\mathbf{x} + c_2\mathbf{y} = 0$. A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is said to be linearly dependent if there exist constants $c_1, c_2, \ldots, c_k$, not all zero, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_k\mathbf{x}_k = 0.$$

  Linear dependence implies that at least one vector in the set can be written as linear combination of the other vectors. Vector of the same dimension that are not linearly dependent are said to be *linearly independent*.

- *projection* (or shadow) of a vector $\mathbf{x}$ on a vector $\mathbf{y}$ is

$$\text{Projection of } \mathbf{x} \text{ on } \mathbf{y} = \frac{(\mathbf{x}'\mathbf{y})}{\mathbf{y}'\mathbf{y}} \cdot \mathbf{y} = \frac{(\mathbf{x}'\mathbf{y})}{L_{\mathbf{y}}} \frac{1}{L_{\mathbf{y}}} \mathbf{y}$$

where the vector $L_{\mathbf{y}}^{-1}\mathbf{y}$ has unit length. The length of the projection is

$$\text{Length of projection} = \frac{|\mathbf{x}'\mathbf{y}|}{L_{\mathbf{y}}} = L_{\mathbf{x}} \left| \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}} L_{\mathbf{y}}} \right| = L_{\mathbf{x}} |\cos(\theta)|$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

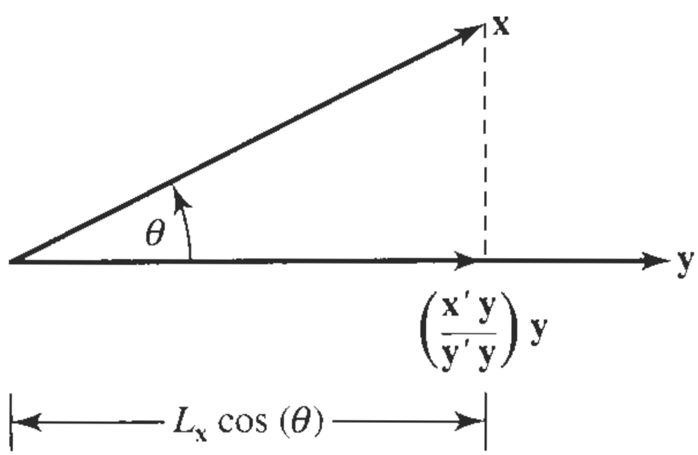

Figure 2.5 The projection of $\mathbf{x}$ on $\mathbf{y}$.

Example 2.2 (Identifying linearly independent vectors) Consider if the set of vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

is linearly dependent.

Matrices

A *matrix* is any rectangular array of real numbers. We denote an arbitrary array of n rows and p columns

$$\mathbf{A}_{\{n \times p\}} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p} \\ a_{21} & a_{22} & \ldots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

Example 2.3 (Transpose of a matrix) if

$$\mathbf{A}_{\{2 \times 3\}} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}$$

then

$$\mathbf{A}'_{\{3 \times 2\}} = \begin{bmatrix} 3 & 1 \\ -1 & 5 \\ 2 & 4 \end{bmatrix}$$

The *product* $c\mathbf{A}$ is the matrix that results from multiplying each elements of **A** by c. Thus

$$c\mathbf{A}_{\{n\times p\}} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1p} \\ ca_{21} & ca_{22} & \dots & ca_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{np} \end{bmatrix}$$

Example 2.4 (The sum of two matrices and multiplication of a matrix by a constant) If

$$\mathbf{A}_{\{2\times 3\}} = \begin{bmatrix} 0 & 3 & 1 \\ 1 & -1 & 1 \end{bmatrix} \quad \mathbf{B}_{\{2\times 3\}} = \begin{bmatrix} 1 & -2 & -3 \\ 2 & 5 & 1 \end{bmatrix}$$

then $4\mathbf{A}$ and $\mathbf{A}+\mathbf{B}$?

The *matrix product* $\mathbf{AB}$ is

$$A_{\{n\times k\}}B_{\{k\times p\}} \quad = \quad \text{the } (n\times p) \text{ matrix whose entry in the } i\text{th row and } j\text{th column is the inner product of the } i\text{th row of } \mathbf{A} \text{ and the } j\text{th column of } \mathbf{B}.$$

or

$$(i,j) \text{ entry of } \mathbf{AB} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} = \sum_{\ell=1}^{k} a_{i\ell}b_{\ell j}$$

Example 2.5 (Matrix multiplication) If

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -2 \\ 7 \\ 9 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix}$$

then $\mathbf{AB}$ and $\mathbf{CA}$?

Example 2.6 (Some typical products and their dimensions) Let

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 4 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 5 \\ 8 \\ -4 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 2 \\ 9 \end{bmatrix}$$

Then $\mathbf{Ab}, \mathbf{bc}', \mathbf{b}'\mathbf{c}$, and $\mathbf{d}'\mathbf{Ad}$?

- Square matrices will be of special importance in our development of statistical methods. A square matrix is said to be *symmetric* if $\mathbf{A} = \mathbf{A}'$ or $a_{ij} = a_{ji}$ for all i and j.

- *Identity matrix* $\mathbf{I}$ act like 1 in ordinary multiplication $(1 \cdot a = a \cdot 1 = a)$,

$$\mathbf{I}_{(k\times k)}\mathbf{A}_{(k\times k)} = \mathbf{A}_{(k\times k)}\mathbf{I}_{(k\times k)} = \mathbf{A}_{(k\times k)} \quad \text{for any} \quad \mathbf{A}_{(k\times k)}$$

- The fundamental scalar relation about the existence of an inverse number a^{-1} such that $a^{-1}a = aa^{-1} = 1$ if $a \neq 0$ has the following matrix algebra extension: If there exists a matrix $\mathbf{B}$ such that

$$\mathbf{B}\mathbf{A} = \mathbf{A}\mathbf{B} = \mathbf{I}$$

then $\mathbf{B}$ is called the *inverse* of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.

Example 2.7 (The existence of a matrix inverse) For

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$$

- Diagonal matrices

- Orthogonal matrices

$$\mathbf{Q}\mathbf{Q}' = \mathbf{Q}'\mathbf{Q} = \mathbf{I} \text{ or } \mathbf{Q}' = \mathbf{Q}^{-1}.$$

- Eigenvalue λ with corresponding eigenvector $\mathbf{x} \neq 0$ if

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

Ordinarily, $\mathbf{x}$ is normalized so that it has length unity; that is $\mathbf{x}'\mathbf{x} = 1$.

- Let **A** be a $k \times k$ square symmetric matrix. Then **A** has k pairs of eigenvalues and eigenvectors namely

$$\lambda_1 \ \mathbf{e}_1, \lambda_2 \ \mathbf{e}_2, \ldots, \lambda_k \ \mathbf{e}_k$$

  The eigenvectors can be chosen to satisfy $1 = \mathbf{e}_1'\mathbf{e}_1 = \cdots = \mathbf{e}_k'\mathbf{e}$ and be mutually perpendicular. The eigenvectors are unique unless two or more eigenvalues are equal.

Example 2.8 (Verifying eigenvalues and eigenvectors) Let

$$\mathbf{A} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}.$$

show that $\lambda_1 = 6$ and $\lambda_2 = -4$ is its eigenvalues and the corresponding eigenvectors are $\mathbf{e}_1 = [1/\sqrt{2}, -1/\sqrt{2}]'$ and $\mathbf{e}_2 = [1/\sqrt{2}, 1/\sqrt{2}]$.

2.3 Positive Definite Matrices

The study of variation and interrelationships in multivariate data is often based upon distances and the assumption that the data are multivariate normally distributed. Squared distance and the multivariate normal density can be expressed in terms of matrix products called ***quadratic forms***. Consequently, it should not be surprising that quadratic forms play central role in multivariate analysis. Quadratic forms that are always nonnegative and the associated positive definite matrices.

- ***spectral decomposition*** for symmetric matrices

$$\mathbf{A}_{(k\times k)} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_k \mathbf{e}_k \mathbf{e}_k'$$

  where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ are the associated normalized $k \times 1$ eigenvectors. $\mathbf{e}_i' \mathbf{e}_i = 1$ for $i = 1, 2, \ldots, k$ and $\mathbf{e}_i' \mathbf{e}_j = 0$ for $i \neq j$.

- Because $\mathbf{x}'\mathbf{A}\mathbf{x}$ has only square terms x_i^2 and product terms $x_i x_k$, it is called a quadratic form. When a $k \times k$ symmetric matrix $\mathbf{A}$ is such that

$$0 \le \mathbf{x}'\mathbf{A}\mathbf{x}$$

  for all $\mathbf{x}' = [x_1, x_2, \ldots, x_k]$, both the matrix $\mathbf{A}$ and the quadratic form are said to be ***nonnegative definite***. If the equality holds in the equation above only for the vector $\mathbf{x}' = [0, 0, \ldots, 0]$, then $\mathbf{A}$ or the quadratic form is said to be positive definite. In other words, $\mathbf{A}$ is positive definite if

$$0 < \mathbf{x}'\mathbf{A}\mathbf{x}$$

  for all vectors $\mathbf{x} \neq 0$.

- Using the spectral decomposition, we can easily show that a $k \times k$ matrix $\mathbf{A}$ is a positive definite matrix if and only if every eigenvalue of $\mathbf{A}$ is positive. $\mathbf{A}$ is a nonnegative definite matrix if and only if all of its eigenvalues are greater than or equal to zero.

Example 2.9 (The spectral decomposition of a matrix) Consider the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix},$$

find its spectral decomposition.

Example 2.10 (A positive definite matrix quadratic form) Show that the matrix for the following quadratic form is positive definite:

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2.$$

- the "distance " of the point $[x_1, x_2, \ldots, x_p]'$ to origin

$$\begin{aligned} (\text{distance})^2 &= a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{pp}^2 \\ &\quad +2(a_{12}x_1x_2 + a_{13}x_1x_3 + \ldots + a_{p-1,p}x_{p-1}x_p) \end{aligned}$$

- the square of the distance **x** to an arbitrary fixed point $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_p]$.

- A geometric interpretation based on the eigenvalues and eigenvectors of the matrix $\mathbf{A}$.

  For example, suppose $p = 2$, Then the points $\mathbf{x}' = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}^2 + 2a_{12}x_1x_2 = c^2$$

  By the spectral decomposition,

$$\mathbf{A} = \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2'$$

  so

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \lambda_2(\mathbf{x}'\mathbf{e}_2)^2$$

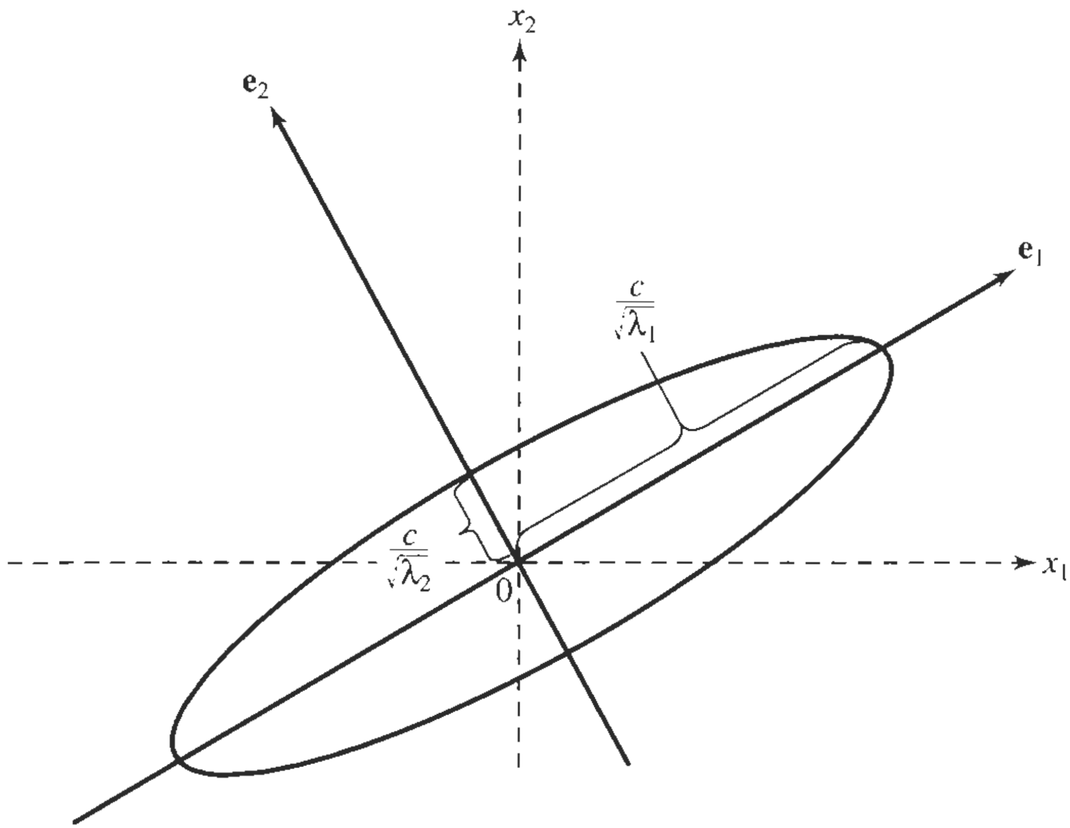

Figure 2.6 Points a constant distance c from the origin $(p = 2, 1 \leq \lambda_1 < \lambda_2)$.

2.4 A Square-Root Matrix

Let **A** be a $k\times$ positive definite matrix with spectral decomposition $\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i'$. Let the normalized eigenvectors be the columns of another matrix $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]$. Then

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i' = \mathbf{P}\Lambda\mathbf{P}'$$

where $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$ and Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \quad \text{with } \lambda_i > 0$$

Thus

$$\mathbf{A}^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\lambda_i}\mathbf{e}_i\mathbf{e}_i'$$

The ***square-root matrix***, of a positive definite matrix $\mathbf{A}$,

$$\mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\Lambda^{1/2}\mathbf{P}'$$

- symmetric: $\mathbf{A}^{1/2'} = \mathbf{A}^{1/2}$
- $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$
- $(\mathbf{A}^{1/2})^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\Lambda^{-1/2}\mathbf{P}'$
- $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$ and $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$, where $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1}$.

Random Vectors and Matrices

A ***random vector*** is a vector whose elements are random variables. Similarly a ***random matrix*** is a matrix whose elements are random variables.

- The expected value of a random matrix

$$\mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{E}(X_{11}) & \mathrm{E}(X_{12}) & \cdots & \mathrm{E}(X_{1p}) \\ \mathrm{E}(X_{21}) & \mathrm{E}(X_{22}) & \cdots & \mathrm{E}(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(X_{n1}) & \mathrm{E}(X_{n2}) & \cdots & \mathrm{E}(X_{np}) \end{bmatrix}$$

- $\mathrm{E}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$

- $\mathrm{E}(\mathbf{A}\boldsymbol{X}\mathbf{B}) = \mathbf{A}\mathrm{E}(\boldsymbol{X})\mathbf{B}$

Example 2.11 (Computing expected values for discrete random variables) Suppose $p = 2$ and $n = 1$, and consider the random vector $\boldsymbol{X}' = [X_1, X_2]$. Let the discrete random variable X_1 have the following probability function

X_1	-1	0	1
$p_1(X_1)$	0.3	0.3	0.4

Similarly, let the discrete random varibale X_2 have the probability function

X_2	0	1
$p_2(X_2)$	0.8	0.2

Calculate $\mathrm{E}(\boldsymbol{X})$.

Mean Vectors and Covariance Matrices

Suppose $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ is a $p \times 1$ random vectors. Then each element of $\boldsymbol{X}$ is a random variables with its own marginal probability distribution.

- The marginal mean $\mu_i = \mathrm{E}(X_i), i = 1, 2, \ldots, p$.
- The marginal variance $\sigma_i^2 = \mathrm{E}(X_i - \mu_i)^2, i = 1, 2, \ldots, p$.
- The behavior of any pair of random variables, such as X_i and X_k, is described by their joint probability function, and a measure of the linear association between them is provided by the covariance

$$\sigma_{ik} = \mathrm{E}(X_i - \mu_i)(X_k - \mu_k)$$

- The means and covariances of $p \times 1$ random vector $\boldsymbol{X}$ can be set out as matrices named ***population variance-covariance*** (matrices).

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X}), \quad \Sigma = \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'.$$

- ***Statistical independent*** X_i and X_k if

$$P(X_i \leq x_i \text{ and } X_k \leq x_k) = P(X_i \leq x_i)P(X_k \leq x_k)$$

or

$$f_{ik}(x_i, x_k) = f_i(x_i)f_k(x_k).$$

- ***Mutually statistically independent*** of the p continuous random variables $X_1, X_2, \ldots, X_p$ if

$$f_{1,2,\ldots,p}(x_1, x_2, \ldots, x_p) = f_1(x_1)f_2(x_2)\cdots f_p(x_p)$$

- ***linear independent*** of X_i, X_k if

$$\text{Cov}(X_i, X_k) = 0$$

- ***Population correlation coefficient*** ρ_{ik}

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

The correlation coefficient measures the amount of *linear* association between the random variable X_i and X_k.

- ***The population correlation matrix*** $\boldsymbol{\rho}$

Example 2.12 (Computing the covariance matrix) Find the covariance matrix for the two random variables X_1 andX_2 introduced in Example 2.11 when their joint probability function $p_{12}(x_1, x_2)$ is represented by the entries in the body of the following table:

$x_1 \backslash x_2$	0	1	$p_1(x_1)$
-1	0.24	0.06	0.3
0	0.16	0.14	0.3
1	0.4	0.00	0.4
$p_2(x_2)$	0.8	0.2	1

Example 2.13 (Computing the correlation matrix from the covariance matrix) Suppose

$$\Sigma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

Obtain the population correlation matrix $\boldsymbol{\rho}$

Partitioning the Covariance Matrix

- Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \dots \\ X_{q+1} \\ \dots \\ X_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \cdots \\ \boldsymbol{X}^{(2)} \end{bmatrix} \text{ and then } \boldsymbol{\mu} = \mathrm{E}\boldsymbol{X} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \dots \\ \mu_{q+1} \\ \dots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \cdots \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}$$

- Define

$$\begin{aligned} & \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' \\ = \; & \mathrm{E}\begin{bmatrix} (\boldsymbol{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\boldsymbol{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & (\boldsymbol{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \\ (\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & (\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)}) \end{bmatrix} \\ = \; & \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \end{aligned}$$

- It is sometimes convenient to use Cov$(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$ note where

$$\text{Cov}(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \Sigma_{12} = \Sigma'_{21}$$

  is a matrix containing all of the covariance between a component of $\boldsymbol{X}^{(1)}$ and a component of $\boldsymbol{X}^{(2)}$.

The Mean Vector and Covariance Matrix for Linear Combinations of Random Variables

- The linear combination $\mathbf{c}'\boldsymbol{X} = c_1X_1 + \cdots + c_pX_p$ has

$$\text{mean} = \mathrm{E}(\mathbf{c}'\boldsymbol{X}) = \mathbf{c}'\boldsymbol{\mu}$$

$$\text{variance} = \mathrm{Var}(\mathbf{c}'\boldsymbol{X}) = \mathbf{c}'\Sigma\mathbf{c}$$

where $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\Sigma = \mathrm{Cov}(\boldsymbol{X})$.

- Let C be a matrix, then the linear combinations of $\mathbf{Z} = \mathbf{C}\boldsymbol{X}$ have

$$\boldsymbol{\mu}_{\mathbf{Z}} = \mathrm{E}(\mathbf{Z}) = \mathrm{E}(\mathbf{C}\boldsymbol{X}) = \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}}$$

$$\Sigma_{\mathbf{Z}} = \mathrm{Cov}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{C}\mathbf{X}) = \mathbf{C}\Sigma_{\mathbf{X}}\mathbf{C}'$$

- ***Sample Mean***

$$\bar{\mathbf{x}}' = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p]$$

- ***Sample Covariance Matrix***

$$\begin{aligned} S_n &= \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \cdots & s_{pp} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)^2 & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{jp}-\bar{x}_p) \\ \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{jp}-\bar{x}_p) & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{jp}-\bar{x}_p)^2 \end{bmatrix} \end{aligned}$$

2.7 Matrix Inequalities and Maximization

- ***Cauchy-Schwarz Inequality***

  Let $\mathbf{b}$ and $\mathbf{d}$ be any two $p \times 1$ vectors. Then

  $$(\mathbf{b}'\mathbf{d})^2 \leq (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d})$$

  with equality if and only if $\mathbf{b} = c\mathbf{d}$ or $\mathbf{d} = c\mathbf{b}$ for some constant c.

- ***Extended Cauchy-Schwarz Inequality***

  Let $\mathbf{b}$ and $\mathbf{d}$ be any two $p \times 1$ vectors, and $\mathbf{B}$ be a positive definite matrix. Then

  $$(\mathbf{b}'\mathbf{d})^2 \leq (\mathbf{b}'\mathbf{B}\mathbf{b})(\mathbf{d}'\mathbf{B}^{-1}\mathbf{d})$$

  with equality if and only if $\mathbf{b} = c\mathbf{B}^{-1}\mathbf{d}$ or $\mathbf{d} = c\mathbf{B}\mathbf{b}$ for some constant c.

- ***Maximization Lemma***

  Let $\mathbf{B}_{p\times p}$ be positive definite and $\mathbf{d}_{p\times 1}$ be a given vector. Then, for arbitrary nonzero vector $\mathbf{x}$,

$$\max_{\mathbf{x}\neq 0} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$$

  with the maximum attained when $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$ for any constant $c \neq 0$.

- ***Maximization of Quadratic Forms for Points on the Unit Sphere***

  Let $\mathbf{B}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then

$$\max_{x\neq 0} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1 \quad \text{(attained when } \mathbf{x} = \mathbf{e}_1\text{)}$$

$$\min_{x\neq 0} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_p \quad \text{(attained when } \mathbf{x} = \mathbf{e}_p\text{)}$$

  Moreover,

$$\max_{x\perp \mathbf{e}_1,\ldots \mathbf{e}_k} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{k+1} \quad \text{(attained when } \mathbf{x} = \mathbf{e}_{k+1}, k = 1, 2, \ldots, p-1\text{)}$$

  where the symbol $\perp$ is read "perpendicular to."