Chapter 1
Mathematical Preliminaries

1.1. Introduction

It is assumed that the reader has had adequate exposure to basic concepts in Probability, Statistics, Calculus and Linear Algebra. This chapter provides a brief review of the results that will be needed in the remainder of this book. No detailed discussion of these topics will be attempted. For essential materials in these areas, the reader is, for instance, referred to Mathai and Haubold (2017a, 2017b). Some properties of vectors, matrices, determinants, Jacobians and wedge product of differentials to be utilized later on, are included in the present chapter. For the sake of completeness, we initially provide some elementary definitions. First, the concepts of vectors, matrices and determinants are introduced.

Consider the consumption profile of a family in terms of the quantities of certain food items consumed every week. The following table gives this family's consumption profile for three weeks:

Table 1.1: Consumption profile

	Rice	Lentils	Carrot	Beans
Week 1	2.00	0.50	1.00	2.00
Week 2	1.50	0.50	0.75	1.50
Week 3	2.00	0.50	0.50	1.25

All the numbers appearing in this table are in kilograms (kg). In Week 1 the family consumed 2 kg of rice, 0.5 kg of lentils, 1 kg of carrots and 2 kg of beans. Looking at the consumption over three weeks, we have an arrangement of 12 numbers into 3 rows and 4 columns. If this consumption profile is expressed in symbols, we have the following representation:

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} 2.00 & 0.50 & 1.00 & 2.00 \\ 1.50 & 0.50 & 0.75 & 1.50 \\ 2.00 & 0.50 & 0.50 & 1.25 \end{bmatrix}$$

where, for example, $a_{11} = 2.00, a_{13} = 1.00, a_{22} = 0.50, a_{23} = 0.75, a_{32} = 0.50, a_{34} = 1.25$.

Definition 1.1.1. **A matrix** An arrangement of mn items into m rows and n columns is called an m by n (written as $m \times n$) matrix.

Accordingly, the above consumption profile matrix is 3×4 (3 by 4), that is, it has 3 rows and 4 columns. The standard notation consists in enclosing the mn items within round ( ) or square [ ] brackets as in the above representation. The above 3×4 matrix is represented in different ways as A, (a_{ij}) and items enclosed by square brackets. The mn items in the $m \times n$ matrix are called *elements of the matrix*. Then, in the above matrix A, $a_{ij} =$ the i-th row, j-th column element or the *(i,j)*-th element. In the above illustration, $i = 1, 2, 3$ (3 rows) and $j = 1, 2, 3, 4$ (4 columns). A general $m \times n$ matrix A can be written as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}. \tag{1.1.1}$$

The elements are separated by spaces in order to avoid any confusion. Should there be any possibility of confusion, then the elements will be separated by commas. Note that the plural of "matrix" is "matrices". Observe that the position of each element in Table 1.1 has a meaning. The elements cannot be permuted as rearranged elements will give different matrices. In other words, *two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ are equal if and only if $a_{ij} = b_{ij}$ for all i and j, that is, they must be element-wise equal.*

In Table 1.1, the first row, which is also a 1×4 matrix, represents this family's first week's consumption. The fourth column represents the consumption of beans over the three weeks' period. Thus, each row and each column in an $m \times n$ matrix has a meaning and represents different aspects. In Eq. (1.1.1), all rows are $1 \times n$ matrices and all columns are $m \times 1$ matrices. *A $1 \times n$ matrix is called a row vector and an $m \times 1$ matrix is called a column vector*. For example, in Table 1.1, there are 3 row vectors and 4 column vectors. If the row vectors are denoted by R_1, R_2, R_3 and the column vectors by C_1, C_2, C_3, C_4, then we have

$R_1 = [2.00 \ 0.50 \ 1.00 \ 2.00], R_2 = [1.50 \ 0.50 \ 0.75 \ 1.50], R_3 = [2.00 \ 0.50 \ 0.50 \ 1.25]$

and

$$C_1 = \begin{bmatrix} 2.00 \\ 1.50 \\ 2.00 \end{bmatrix}, \; C_2 = \begin{bmatrix} 0.50 \\ 0.50 \\ 0.50 \end{bmatrix}, \; C_3 = \begin{bmatrix} 1.00 \\ 0.75 \\ 0.50 \end{bmatrix}, \; C_4 = \begin{bmatrix} 2.00 \\ 1.50 \\ 1.25 \end{bmatrix}.$$

If the total consumption in Week 1 and Week 2 is needed, it is obtained by adding the row vectors element-wise:

$$R_1 + R_2 = [2.00 + 1.50 \; 0.50 + 0.50 \; 1.00 + 0.75 \; 2.00 + 1.50] = [3.50 \; 1.00 \; 1.75 \; 3.50].$$

We will define the addition of two matrices in the same fashion as in the above illustration. For the addition to hold, both matrices must be of the same order $m \times n$. Let $A = (a_{ij})$ *and* $B = (b_{ij})$ *be two* $m \times n$ *matrices. Then the sum, denoted by* $A + B$, *is defined as*

$$A + B = (a_{ij} + b_{ij})$$

or equivalently as the matrix obtained by adding the corresponding elements. For example,

$$C_1 + C_3 = \begin{bmatrix} 2.00 \\ 1.50 \\ 2.00 \end{bmatrix} + \begin{bmatrix} 1.00 \\ 0.75 \\ 0.50 \end{bmatrix} = \begin{bmatrix} 3.00 \\ 2.25 \\ 2.50 \end{bmatrix}.$$

Repeating the addition, we have

$$C_1 + C_3 + C_4 = (C_1 + C_3) + \begin{bmatrix} 2.00 \\ 1.50 \\ 1.25 \end{bmatrix} = \begin{bmatrix} 3.00 \\ 2.25 \\ 2.50 \end{bmatrix} + \begin{bmatrix} 2.00 \\ 1.50 \\ 1.25 \end{bmatrix} = \begin{bmatrix} 5.00 \\ 3.75 \\ 3.75 \end{bmatrix}.$$

In general, if $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, $D = (d_{ij})$ are $m \times n$ matrices, then $A + B + C + D = (a_{ij} + b_{ij} + c_{ij} + d_{ij})$, that is, it is the matrix obtained by adding the corresponding elements.

Suppose that in Table 1.1, we wish to express the elements in terms of grams instead of kilograms; then, each and every element therein must be multiplied by 1000. Thus, if A is the matrix corresponding to Table 1.1 and B is the matrix in terms of grams, we have

$$A = \begin{bmatrix} 2.00 & 0.50 & 1.00 & 2.00 \\ 1.50 & 0.50 & 0.75 & 1.50 \\ 2.00 & 0.50 & 0.50 & 1.25 \end{bmatrix},$$

$$B = \begin{bmatrix} 1000 \times 2.00 & 1000 \times 0.50 & 1000 \times 1.00 & 1000 \times 2.00 \\ 1000 \times 1.50 & 1000 \times 0.50 & 1000 \times 0.75 & 1000 \times 1.50 \\ 1000 \times 2.00 & 1000 \times 0.50 & 1000 \times 0.50 & 1000 \times 1.25 \end{bmatrix}.$$

We may write this symbolically as $B = 1000 \times A = 1000\, A$. Note that 1000 is a 1×1 matrix or a scalar quantity. *Any 1×1 matrix is called a scalar quantity*. Then we may define *scalar multiplication* of a matrix A by the scalar quantity c as $c\, A$ or $A = (a_{ij}) \Rightarrow c\, A = (c\, a_{ij})$ or it is obtained by multiplying each and every element of A by the scalar quantity c. As a convention, c is written on the left of A as $c\, A$ and not as $A\, c$. Then, if $c = -1$, then $c\, A = (-1)A = -A$ and $A + (-1)A = A - A = O$ where the capital O denotes a matrix whose elements are all equal to zero. A general $m \times n$ matrix wherein every element is zero is referred to as a *null matrix* and it is written as O (not zero). We may also note that if A, B, C are $m \times n$ matrices, then $A + (B + C) = (A + B) + C$. Moreover, $A + O = O + A = A$. If $m = n$, in which case the number of rows is equal to the number of columns, the resulting matrix is referred to as *a square matrix* because it is a square arrangement of elements; otherwise the matrix is called *a rectangular matrix*. Some special cases of square matrices are the following: For an $n \times n$ matrix or *a square matrix of order n*, suppose that $a_{ij} = 0$ for all $i \neq j$ (that is, all non-diagonal elements are zeros; here "diagonal" means the diagonal going from top left to bottom right) and if there is at least one nonzero diagonal element, then such a matrix is called *a diagonal matrix* and it is usually written as $\mathrm{diag}(d_1, \ldots, d_n)$ where $d_1, \ldots, d_n$ are the diagonal elements. Here are some examples of 3×3 diagonal matrices:

$$D_1 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 7 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad D_3 = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}, \quad a \neq 0.$$

If in D_3, $a = 1$ so that all the diagonal elements are unities, the resulting matrix is called *an identity matrix* and a diagonal matrix whose diagonal elements are all equal to some number a that is not equal to 0 or 1, is referred to as *a scalar matrix*. A square non-null matrix $A = (a_{ij})$ that contains at least one nonzero element below its leading diagonal and whose elements above the leading diagonal are all equal to zero, that is, $a_{ij} = 0$ for all $i < j$, is called *a lower triangular matrix*. Some examples of 2×2 lower triangular matrices are the following:

$$T_1 = \begin{bmatrix} 5 & 0 \\ 2 & 1 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 3 & 0 \\ 1 & 0 \end{bmatrix}, \quad T_3 = \begin{bmatrix} 0 & 0 \\ -3 & 0 \end{bmatrix}.$$

If, in a square non-null matrix, all elements below the leading diagonal are zeros and there is at least one nonzero element above the leading diagonal, then such a square matrix is referred to as *an upper triangular matrix*. Here are some examples:

$$T_1 = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}, \ T_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}, \ T_3 = \begin{bmatrix} 0 & 0 & 7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Multiplication of Matrices Once again, consider Table 1.1. Suppose that by consuming 1 kg of rice, the family is getting 700 g (where g represents grams) of starch, 2 g protein and 1 g fat; that by eating 1 kg of lentils, the family is getting 200 g of starch, 100 g of protein and 100 g of fat; that by consuming 1 kg of carrots, the family is getting 100 g of starch, 200 g of protein and 150 g of fat; and that by eating 1 kg of beans, the family is getting 50 g of starch, 100 g of protein and 200 g of fat, respectively. Then the starch-protein-fat matrix, denoted by B, is the following where the rows correspond to rice, lentil, carrots and beans, respectively:

$$B = \begin{bmatrix} 700 & 2 & 1 \\ 200 & 100 & 100 \\ 100 & 200 & 150 \\ 50 & 100 & 200 \end{bmatrix}.$$

Let B_1, B_2, B_3 be the columns of B. Then, the first column B_1 of B represents the starch intake per kg of rice, lentil, carrots and beans respectively. Similarly, the second column B_2 represents the protein intake per kg and the third column B_3 represents the fat intake, that is,

$$B_1 = \begin{bmatrix} 700 \\ 200 \\ 100 \\ 50 \end{bmatrix}, \ B_2 = \begin{bmatrix} 2 \\ 100 \\ 200 \\ 100 \end{bmatrix}, \ B_3 = \begin{bmatrix} 1 \\ 100 \\ 150 \\ 200 \end{bmatrix}.$$

Let the rows of the matrix A in Table 1.1 be denoted by A_1, A_2 and A_3, respectively, so that

$A_1 = [2.00\ 0.50\ 1.00\ 2.00], A_2 = [1.50\ 0.50\ 0.75\ 1.50], A_3 = [2.00\ 0.50\ 0.50\ 1.25].$

Then, the total intake of starch by the family in Week 1 is available from

$$2.00 \times 700 + 0.50 \times 200 + 1.00 \times 100 + 2.00 \times 50 = 1700g.$$

This is the sum of the element-wise products of A_1 with B_1. We will denote this by $A_1 \cdot B_1$ (A_1 dot B_1). The total intake of protein by the family in Week 1 is determined as follows:

$$A_1.B_2 = 2.00 \times 2 + 0.50 \times 100 + 1.00 \times 200 + 2.00 \times 100 = 454 \text{ g}$$

and the total intake of fat in Week 1 is given by

$$A_1.B_3 = 2.00 \times 1 + 0.50 \times 100 + 1.00 \times 150 + 2.00 \times 200 = 602 \text{ g}.$$

Thus, the dot product of A_1 with B_1, B_2, B_3 provides the intake of starch, protein and fat in Week 1. Similarly, the dot product of A_2 with B_1, B_2, B_3 gives the intake of starch, protein and fat in Week 2. Thus, the configuration of starch, protein and fat intake over the three weeks is

$$AB = \begin{bmatrix} A_1 \cdot B_1 & A_1 \cdot B_2 & A_1 \cdot B_3 \\ A_2 \cdot B_1 & A_2 \cdot B_2 & A_2 \cdot B_3 \\ A_3 \cdot B_1 & A_3 \cdot B_2 & A_3 \cdot B_3 \end{bmatrix}.$$

A matrix having one column and m rows is an $m \times 1$ matrix that is referred to as *a column vector of m elements or a column vector of order m*. A matrix having one row and n column is a $1 \times n$ matrix called *a row vector of n components or a row vector of order n*. Let A be a row or column vector of order n, which consist of n elements or components. Let the elements comprising A be denoted by $a_1, \ldots, a_n$. Let B be a row or column vector of order n consisting of the elements $b_1, \ldots, b_n$. *Then, the dot product of A and B, denoted by $A \cdot B = B \cdot A$ is defined as $A \cdot B = a_1b_1 + a_2b_2 + \cdots + a_nb_n$ so that or the corresponding elements of A and B are multiplied and added up.* Let A be an $m \times n$ matrix whose m rows are written as $A_1, \ldots, A_m$. Let B be another $n \times r$ matrix whose r columns are written as $B_1, \ldots, B_r$. Note that the number of columns of A is equal to the number of rows of B, which in this case is n. When the number of columns of A is equal to the number of rows of B, the product AB is defined and equal to

$$AB = \begin{bmatrix} A_1 \cdot B_1 & A_1 \cdot B_2 & \ldots & A_1 \cdot B_r \\ A_2 \cdot B_1 & A_2 \cdot B_2 & \ldots & A_2 \cdot B_r \\ \vdots & \vdots & \ddots & \vdots \\ A_m \cdot B_1 & A_m \cdot B_2 & \ldots & A_m \cdot B_r \end{bmatrix} \text{ with } A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix}, \ B = [B_1 \ B_2 \ \cdots \ B_r],$$

the resulting matrix AB being of order $m \times r$. When AB is defined, BA need not be defined. However, if $r = m$, then BA is also defined, otherwise not. In other words, if $A = (a_{ij})$ is $m \times n$ and if $B = (b_{ij})$ is $n \times r$ and if $C = (c_{ij}) = AB$, then $c_{ij} = A_i \cdot B_j$ where A_i is the i-th row of A and B_j is the j-th column of B or $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$ for all i and j. For example,

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \end{bmatrix}, \ B = \begin{bmatrix} 2 & -2 \\ 3 & 2 \\ 1 & 0 \end{bmatrix} \Rightarrow$$

$$AB = \begin{bmatrix} (1)(2) + (-1)(3) + (0)(1) & (1)(-2) + (-1)(2) + (0)(0) \\ (2)(2) + (3)(3) + (5)(1) & 2(-2) + 3(2) + (5)(0) \end{bmatrix} = \begin{bmatrix} -1 & -4 \\ 18 & 2 \end{bmatrix}.$$

Note that, in this case, BA is defined and equal to

$$BA = \begin{bmatrix} 2 & -2 \\ 3 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (2)(1) + (-2)(2) & (2)(-1) + (-2)(3) & (2)(0) + (-2)(5) \\ (3)(1) + (2)(2) & (3)(-1) + (2)(3) & (3)(0) + (2)(5) \\ (1)(1) + (0)(2) & (1)(-1) + (0)(3) & (1)(0) + (0)(5) \end{bmatrix}$$

$$= \begin{bmatrix} -2 & -8 & -10 \\ 7 & 3 & 10 \\ 1 & -1 & 0 \end{bmatrix}.$$

As another example, let

$$A = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \ B = [2 \ 3 \ 5] \ \Rightarrow \ AB = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} [2 \ 3 \ 5] = \begin{bmatrix} 2 & 3 & 5 \\ -2 & -3 & -5 \\ 4 & 6 & 10 \end{bmatrix}$$

which is 3×3, whereas BA is 1×1:

$$BA = [2 \ 3 \ 5] \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = 9.$$

As yet another example, let

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

Note that here both A and B are lower triangular matrices. The products AB and BA are defined since both A and B are 3×3 matrices. For instance,

$$AB = \begin{bmatrix} (2)(1) + (0)(0) + (0)(1) & (2)(0) + (0)(2) + (0)(-1) & (2)(0) + (0)(0) + (0)(0) \\ (-1)(1) + (1)(0) + (0)(1) & (-1)(0) + (1)(2) + (0)(-1) & (-1)(0) + (1)(0) + (0)(0) \\ (1)(1) + (1)(0) + (1)(1) & (1)(0) + (1)(2) + (1)(-1) & (1)(0) + (1)(0) + (1)(0) \end{bmatrix}$$

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

Observe that since A and B are lower triangular, AB is also lower triangular. Here are some general properties of products of matrices: When the product of the matrices is defined, in which case we say that they are conformable for multiplication,

(1): *the product of two lower triangular matrices is lower triangular;*

(2): *the product of two upper triangular matrices is upper triangular;*

(3): *the product of two diagonal matrices is diagonal;*

(4): *if A is $m \times n$, $IA = A$ where $I = I_m$ is an identity matrix, and $A\,I_n = A$;*

(5): $OA = O$ whenever OA is defined, and $AO = O$ whenever AO is defined.

Transpose of a Matrix A matrix whose rows are the corresponding columns of A or, equivalently, a matrix whose columns are the corresponding rows of A is called the transpose of A denoted as A' (A prime). For example,

$$A_1 = [1 \ 2 \ -1] \Rightarrow A_1' = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \ A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \Rightarrow A_2' = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix};$$

$$A_3 = \begin{bmatrix} 1 & 3 \\ 3 & 7 \end{bmatrix} \Rightarrow A_3' = \begin{bmatrix} 1 & 3 \\ 3 & 7 \end{bmatrix} = A_3; \ A_4 = \begin{bmatrix} 0 & 5 \\ -5 & 0 \end{bmatrix} \Rightarrow A_4' = \begin{bmatrix} 0 & -5 \\ 5 & 0 \end{bmatrix} = -A_4.$$

Observe that A_2 is lower triangular and A_2' is upper triangular, that $A_3' = A_3$, and that $A_4' = -A_4$. Note that if A is $m \times n$, then A' is $n \times m$. If A is 1×1 then A' is the same scalar (1×1) quantity. If A is a square matrix and $A' = A$, then A *is called a symmetric matrix*. If B is a square matrix and $B' = -B$, then B is called a *skew symmetric matrix*. Within a skew symmetric matrix $B = (b_{ij})$, a diagonal element must satisfy the equation $b_{jj}' = -b_{jj}$, which necessitates that $b_{jj} = 0$, whether B be real or complex. Here are some properties of the transpose: The transpose of a lower triangular matrix is upper triangular; the transpose of an upper triangular matrix is lower triangular; the transpose of a diagonal matrix is diagonal; the transpose of an $m \times n$ null matrix is an $n \times m$ null matrix;

$$(A')' = A; \ (AB)' = B'A'; \ (A_1 A_2 \cdots A_k)' = A_k' \cdots A_2' A_1'; \ (A + B)' = A' + B'$$

whenever AB, $A + B$, and $A_1 A_2 \cdots A_k$ are defined.

Trace of a Square Matrix The trace is defined only for square matrices. Let $A = (a_{ij})$ be an $n \times n$ matrix whose leading diagonal elements are $a_{11}, a_{22}, \ldots, a_{nn}$; then the trace of A, denoted by $\mathrm{tr}(A)$ is defined as $\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$, that is, the sum of the elements comprising the leading diagonal. The following properties can directly be deduced from the definition. Whenever AB and BA are defined, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ where AB need not be equal to BA. If A is $m \times n$ and B is $n \times m$, then AB is $m \times m$ whereas

BA is $n \times n$; however, the traces are equal, that is, $\text{tr}(AB) = \text{tr}(BA)$, which implies that $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$.

Length of a Vector Let V be a $n \times 1$ real column vector or a $1 \times n$ real row vector then V can be represented as a point in n-dimensional Euclidean space when the elements are real numbers. Consider a 2-dimensional vector (or 2-vector) with the elements $(1, 2)$. Then, this vector corresponds to the point depicted in Fig. 1.1.
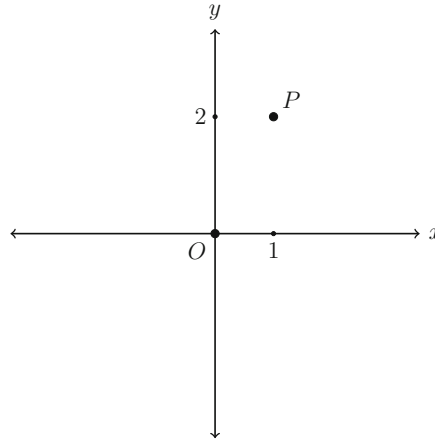
Figure 1.1 The point $P = (1, 2)$ in the plane

Let O be the origin and P be the point. Then, the length of the resulting vector is the Euclidean distance between O and P, that is, $+\sqrt{(1)^2 + (2)^2} = +\sqrt{5}$. Let $U = (u_1, \ldots, u_n)$ be a real n-vector, either written as a row or a column. *Then the length of U, denoted by $\|U\|$ is defined as follows*:

$$\|U\| = +\sqrt{u_1^2 + \cdots + u_n^2}$$

whenever the elements $u_1, \ldots, u_n$ are real. If $u_1, \ldots, u_n$ are complex numbers then $\|U\| = \sqrt{|u_1|^2 + \cdots + |u_n|^2}$ where $|u_j|$ denotes the absolute value or modulus of u_j. If $u_j = a_j + ib_j$, with $i = \sqrt{(-1)}$ and a_j, b_j real, then $|u_j| = +\sqrt{(a_j^2 + b_j^2)}$. If the length of a vector is unity, that vector is called a unit vector. For example, $e_1 = (1, 0, \ldots, 0), e_2 =$

$(0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ are all unit vectors. As well, $V_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $V_2 = (\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}})$ are unit vectors. If two n-vectors, U_1 and U_2, are such that $U_1 \cdot U_2 = 0$, that is, their dot product is zero, then the two vectors are said to be *orthogonal to each other*. For example, if $U_1 = (1, 1)$ and $U_2 = (1, -1)$, then $U_1 \cdot U_2 = 0$ and U_1 and U_2 are orthogonal to each other; similarly, if $U_1 = (1, 1, 1)$ and $U_2 = (1, -2, 1)$, then $U_1 \cdot U_2 = 0$ and U_1 and U_2 are orthogonal to each other. If $U_1, \ldots, U_k$ are k vectors, each of order n, all being either row vectors or column vectors, and if $U_i \cdot U_j = 0$ for all $i \neq j$, that is, all distinct vectors are orthogonal to each other, then we say that $U_1, \ldots, U_k$ forms *an orthogonal system of vectors*. In addition, if the length of each vector is unity, $\|U_j\| = 1, j = 1, \ldots, k$, then we say that $U_1, \ldots, U_k$ is *an orthonormal system of vectors*. If a matrix A is real and its rows and its columns form an orthonormal system, then A is called *an orthonormal matrix*. In this case, $AA' = I_n$ and $A'A = I_n$; accordingly, any square matrix A of real elements such that $AA' = I_n$ and $A'A = I_n$ *is referred to as an orthonormal matrix*. If only one equation holds, that is, B is a real matrix such that either $BB' = I, B'B \neq I$ or $B'B = I, BB' \neq I$, then B is called *a semiorthonormal matrix*. For example, consider the matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}; \quad \text{then } AA' = I_2, \ A'A = I_2,$$

and A is an orthonormal matrix. As well,

$$A = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \Rightarrow AA' = I_3, \ A'A = I_3,$$

and A here is orthonormal. However,

$$B = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \Rightarrow BB' = I_2, \ B'B \neq I$$

so that B is semiorthonormal. On deleting some rows from an orthonormal matrix, we obtain a semiorthonormal matrix such that $BB' = I$ and $B'B \neq I$. Similarly, if we delete some of the columns, we end up with a semiorthonormal matrix such that $B'B = I$ and $BB' \neq I$.

Linear Independence of Vectors Consider the vectors $U_1 = (1, 1, 1), U_2 = (1, -2, 1)$, $U_3 = (3, 0, 3)$. Then, we can easily see that $U_3 = 2U_1 + U_2 = 2(1, 1, 1) + (1, -2, 1) = (3, 0, 3)$ or $U_3 - 2U_1 - U_2 = O$ (a null vector). In this case, one of the vectors can be written

as a linear function of the others. Let $V_1 = (1, 1, 1)$, $V_2 = (1, 0, -1)$, $V_3 = (1, -2, 1)$. Can any one of these be written as a linear function of others? If that were possible, then there would exist a linear function of V_1, V_2, V_3 that is equal to is a null vector. Let us consider the equation $a_1 V_1 + a_2 V_2 + a_3 V_3 = (0, 0, 0)$ where a_1, a_2, a_3 are scalars where at least one of them is nonzero. Note that $a_1 = 0$, $a_2 = 0$ and $a_3 = 0$ will always satisfy the above equation. Thus, our question is whether $a_1 = 0$, $a_2 = 0$, $a_3 = 0$ is the only solution.

$$a_1 V_1 + a_2 V_2 + a_3 V_3 = O \Rightarrow a_1(1, 1, 1) + a_2(1, 0, -1) + a_3(1, -2, 1) = (0, 0, 0)$$
$$\Rightarrow a_1 + a_2 + a_3 = 0 \ (i); \quad a_1 - 2a_3 = 0 \ (ii); \quad a_1 - a_2 + a_3 = 0. \tag{iii}$$

From (ii), $a_1 = 2a_3$. Then, from (iii), $3a_3 - a_2 = 0 \Rightarrow a_2 = 3a_3$; then from (i), $2a_3 + 3a_3 + a_3 = 0$ or $6a_3 = 0$ or $a_3 = 0$. Thus, $a_2 = 0$, $a_1 = 0$ and there is no nonzero a_1 or a_2 or a_3 satisfying the equation and hence V_1, V_2, V_3 cannot be linearly dependent; so, they are linearly independent. Hence, we have the following definition: Let $U_1, \ldots, U_k$ be k vectors, each of order n, all being either row vectors or column vectors, so that addition and linear functions are defined. Let $a_1, \ldots, a_k$ be scalar quantities. Consider the equation

$$a_1 U_1 + a_2 U_2 + \cdots + a_k U_k = O \text{ (a null vector)}. \tag{iv}$$

If $a_1 = 0$, $a_2 = 0$, $\ldots$, $a_k = 0$ is the only solution to *(iv)*, then $U_1, \ldots, U_k$ *are linearly independent, otherwise they are linearly dependent*. If they are linearly dependent, then at least one of the vectors can be expressed as a linear function of others. The following properties can be established from the definition: Let $U_1, \ldots, U_k$ be n-vectors, $k \leq n$.

(1) If $U_1, \ldots, U_k$ are mutually orthogonal, then they are linearly independent, that is, if $U_i \cdot U_j = 0$, for all $i \neq j$, then $U_1, \ldots, U_k$ are linearly independent;

(2) There cannot be more than n mutually orthogonal n-vectors;

(3) There cannot be more than n linearly independent n-vectors.

Rank of a Matrix The maximum number of linearly independent row vectors of a $m \times n$ matrix is called the *row rank of the matrix*; the maximum number of linearly independent column vectors is called the *column rank of the matrix*. It can be shown that the row rank of any matrix is equal to its column rank, and this common rank is called the *rank of the matrix*. If r is the rank of a $m \times n$ matrix, then $r \leq m$ and $r \leq n$. If $m \leq n$ and the rank is m or if $n \leq m$ and the rank is n, then the matrix is called *a full rank matrix*. A square matrix of full rank is called *a nonsingular matrix*. When the rank of an $n \times n$ matrix is $r < n$, this matrix is referred to as *a singular matrix*. Singularity is defined only for square matrices. The following properties clearly hold:

(1) *A diagonal matrix with at least one zero diagonal element is singular or a diagonal matrix with all nonzero diagonal elements is nonsingular*;

(2) *A triangular matrix (upper or lower) with at least one zero diagonal element is singular or a triangular matrix with all diagonal elements nonzero is nonsingular*;

(3) *A square matrix containing at least one null row vector or at least one null column vector is singular*;

(4) *Linear independence or dependence in a collection of vectors of the same order and category (either all are row vectors or all are column vectors) is not altered by multiplying any of the vectors by a nonzero scalar*;

(5) *Linear independence or dependence in a collection of vectors of the same order and category is not altered by adding any vector of the set to any other vector in the same set*;

(6) *Linear independence or dependence in a collection of vectors of the same order and category is not altered by adding a linear combination of vectors from the same set to any other vector in the same set*;

(7) *If a collection of vectors of the same order and category is a linearly dependent system, then at least one of the vectors can be made null by the operations of scalar multiplication and addition*.

Note: We have defined "vectors" as an ordered set of items such as an ordered set of numbers. One can also give a general definition of a vector as an element in a set S which is closed under the operations of scalar multiplication and addition (these operations are to be defined on S), that is, letting S be a set of items, if $V_1 \in S$ and $V_2 \in S$, then $cV_1 \in S$ and $V_1 + V_2 \in S$ for all scalar c and for all V_1 and V_2, that is, if V_1 is an element in S, then cV_1 is also an element in S and if V_1 and V_2 are in S, then $V_1 + V_2$ is also in S, where operations cV_1 and $V_1 + V_2$ are to be properly defined. One can impose additional conditions on S. However, for our discussion, the notion of vectors as ordered set of items will be sufficient.

1.2. Determinants

Determinants are defined only for square matrices. They are certain scalar functions of the elements of the square matrix under consideration. We will motivate this particular function by means of an example that will also prove useful in other areas. Consider two 2-vectors, either both row vectors or both column vectors. Let

$U = OP$ and $V = OQ$ be the two vectors as shown in Fig. 1.2. If the vectors are separated by a nonzero angle θ then one can create the parallelogram $OPSQ$ with these two vectors as shown in Fig. 1.2.
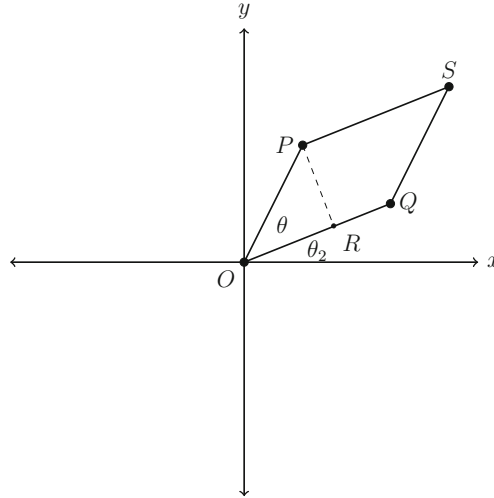
Figure 1.2 Parallelogram generated from two vectors

The area of the parallelogram is twice the area of the triangle OPQ. If the perpendicular from P to OQ is PR, then the area of the triangle is $\frac{1}{2}PR \times OQ$ or the area of the parallelogram $OPSQ$ is $PR \times \|V\|$ where PR is $OP \times \sin\theta = \|U\| \times \sin\theta$. Therefore the area is $\|U\| \, \|V\| \, \sin\theta$ or the area, denoted by v is

$$v = \|U\| \, \|V\| \sqrt{(1 - \cos^2\theta)}.$$

If θ_1 is the angle U makes with the x-axis and θ_2, the angle V makes with the x-axis, then if U and V are as depicted in Fig. 1.2, then $\theta = \theta_1 - \theta_2$. It follows that

$$\cos\theta = \cos(\theta_1 - \theta_2) = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 = \frac{U \cdot V}{\|U\| \, \|V\|},$$

as can be seen from Fig. 1.2. In this case,

$$v = \|U\| \, \|V\| \sqrt{1 - \left(\frac{(U \cdot V)}{\|U\| \, \|V\|}\right)^2} = \sqrt{(\|U\|)^2 \, (\|V\|)^2 - (U \cdot V)^2} \qquad (1.2.1)$$

and

$$v^2 = (\|U\|)^2(\|V\|)^2 - (U \cdot V)^2. \tag{1.2.2}$$

This can be written in a more convenient way. Letting $X = \begin{pmatrix} U \\ V \end{pmatrix}$,

$$XX' = \begin{pmatrix} U \\ V \end{pmatrix}(U'\ V') = \begin{bmatrix} UU' & UV' \\ VU' & VV' \end{bmatrix} = \begin{bmatrix} U \cdot U & U \cdot V \\ V \cdot U & V \cdot V \end{bmatrix}. \tag{1.2.3}$$

On comparing (1.2.2) and (1.2.3), we note that (1.2.2) is available from (1.2.3) by taking a scalar function of the following type. Consider a matrix

$$C = \begin{bmatrix} a & b \\ c & d \end{bmatrix}; \text{ then (1.2.2) is available by taking } ad - bc$$

where a, b, c, d are scalar quantities. A scalar function of this type is the determinant of the matrix C.

A general result can be deduced from the above procedure: If U and V are n-vectors and if θ is the angle between them, then

$$\cos \theta = \frac{U \cdot V}{\|U\| \, \|V\|}$$

or the dot product of U and V divided by the product of their lengths when $\theta \neq 0$, and the numerator is equal to the denominator when $\theta = 2n\pi, \ n = 0, 1, 2, \ldots$. We now provide a formal definition of the determinant of a square matrix.

Definition 1.2.1.   The Determinant of a Square Matrix Let $A = (a_{ij})$ be a $n \times n$ matrix whose rows (columns) are denoted by $\alpha_1, \ldots, \alpha_n$. For example, if α_i is the i-th row vector, then

$$\alpha_i = (a_{i1}\ a_{i2}\ \ldots\ a_{in}).$$

The determinant of A will be denoted by $|A|$ or $\det(A)$ when A is real or complex and the absolute value of the determinant of A will be denoted by $|\det(A)|$ when A is in the complex domain. Then, $|A|$ will be a function of $\alpha_1, \ldots, \alpha_n$, written as

$$|A| = \det(A) = f(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_n),$$

which will be defined by the following four axioms (postulates or assumptions): (this definition also holds if the elements of the matrix are in the complex domain)

(1) $\qquad\qquad f(\alpha_1, \ldots, c\,\alpha_i, \ldots, \alpha_n) = cf(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n),$

which is equivalent to saying that if any row (column) is multiplied by a scalar quantity c (including zero), then the whole determinant is multiplied by c;

(2) $$f(\alpha_1, ...\alpha_i, \ldots, \alpha_i + \alpha_j, \ldots, \alpha_n) = f(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_n),$$

which is equivalent to saying that if any row (column) is added to any other row (column), then the value of the determinant remains the same;

(3) $$f(\alpha_1, \ldots, \gamma_i + \delta_i, \ldots, \alpha_n) = f(\alpha_1, \ldots, \gamma_i, \ldots, \alpha_n) + f(\alpha_1, \ldots, \delta_i, \ldots, \alpha_n),$$

which is equivalent to saying that if any row (column), say the i-th row (column) is written as a sum of two vectors, $\alpha_i = \gamma_i + \delta_i$ then the determinant becomes the sum of two determinants such that γ_i appears at the position of α_i in the first one and δ_i appears at the position of α_i in the second one;

(4) $$f(e_1, \ldots, e_n) = 1$$

where $e_1, \ldots, e_n$ are the basic unit vectors as previously defined; this axiom states that the determinant of an identity matrix is 1.

Let us consider some corollaries resulting from Axioms (1) to (4). On combining Axioms (1) and (2), we have that the value of a determinant remains unchanged if a linear function of any number of rows (columns) is added to any other row (column). As well, the following results are direct consequences of the axioms.

(i): *The determinant of a diagonal matrix is the product of the diagonal elements* [which can be established by repeated applications of Axiom (1)];

(ii): *If any diagonal element in a diagonal matrix is zero, then the determinant is zero, and thereby the corresponding matrix is singular; if none of the diagonal elements of a diagonal matrix is equal to zero, then the matrix is nonsingular.*

(iii): *If any row (column) of a matrix is null, then the determinant is zero or the matrix is singular* [Axiom (1)];

(iv): *If any row (column) is a linear function of other rows (columns), then the determinant is zero* [By Axioms (1) and (2), we can reduce that row (column) to a null vector]. Thus, the determinant of a singular matrix is zero or if the row (column) vectors form a linearly dependent system, then the determinant is zero.

By using Axioms (1) and (2), we can reduce a triangular matrix to a diagonal form when evaluating its determinant. For this purpose we shall use the following standard notation: "$c(i) + (j) \Rightarrow$" means "c times the i-th row is added to the j-th row which

results in the following:" Let us consider a simple example. Consider a triangular matrix and its determinant. Evaluate the determinant of the following matrix:

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

It is an upper triangular matrix. We take out -4 from the third row by using Axiom (1). Then,

$$|T| = -4 \begin{vmatrix} 2 & 1 & 5 \\ 0 & 3 & 4 \\ 0 & 0 & 1 \end{vmatrix}.$$

Now, add (-4) times the third row to the second row and (-5) times the third row to the first row. This in symbols is "$-4(3) + (2), -5(3) + (1) \Rightarrow$". The net result is that the elements 5 and 4 in the last column are eliminated without affecting the other elements, so that

$$|T| = -4 \begin{vmatrix} 2 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Now take out 3 from the second row and then use the second row to eliminate 1 in the first row. After taking out 3 from the second row, the operation is "$-1(2) + (1) \Rightarrow$". The result is the following:

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

Now, take out 2 from the first row, then by Axiom (4) the determinant of the resulting identity matrix is 1, and hence $|T|$ is nothing but the product of the diagonal elements. Thus, we have the following result:

(v): *The determinant of a triangular matrix (upper or lower) is the product of its diagonal elements; accordingly, if any diagonal element in a triangular matrix is zero, then the determinant is zero and the matrix is singular*. For a triangular matrix to be nonsingular, all its diagonal elements must be non-zeros.

    The following result follows directly from Axioms (1) and (2). The proof is given in symbols.

(vi): *If any two rows (columns) are interchanged (this means one transposition), then the resulting determinant is multiplied by -1 or every transposition brings in -1 outside that determinant as a multiple*. If an odd number of transpositions are done, then the whole

determinant is multiplied by -1, and for even number of transpositions, the multiplicative factor is $+1$ or no change in the determinant. An outline of the proof follows:

$$
\begin{aligned}
|A| &= f(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_n) \\
&= f(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_i + \alpha_j, \ldots, \alpha_n) \quad [\text{Axiom (2)}] \\
&= -f(\alpha_1, \ldots, \alpha_i, \ldots, -\alpha_i - \alpha_j, \ldots, \alpha_n) \quad [\text{Axiom (1)}] \\
&= -f(\alpha_1, \ldots, -\alpha_j, \ldots, -\alpha_i - \alpha_j, \ldots, \alpha_n) \quad [\text{Axiom (2)}] \\
&= f(\alpha_1, \ldots, \alpha_j, \ldots, -\alpha_i - \alpha_j, \ldots, \alpha_n) \quad [\text{Axiom (1)}] \\
&= f(\alpha_1, \ldots, \alpha_j, \ldots, -\alpha_i, \ldots, \alpha_n) \quad [\text{Axiom (2)}] \\
&= -f(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_i, \ldots, \alpha_n) \quad [\text{Axiom (1)}].
\end{aligned}
$$

Now, note that the i-th and j-th rows (columns) are interchanged and the result is that the determinant is multiplied by -1.

With the above six basic properties, we are in a position to evaluate most of the determinants.

Example 1.2.1.   Evaluate the determinant of the matrix

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

Solution 1.2.1.   Since, this is a triangular matrix, its determinant will be product of its diagonal elements. Proceeding step by step, take out 2 from the first row by using Axiom (1). Then $-1(1) + (2)$, $-2(1) + (3)$, $-3(1) + (3) \Rightarrow$. The result of these operations is the following:

$$
|A| = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 4 \end{vmatrix}.
$$

Now, take out 5 from the second row so that $1(2) + (3) \Rightarrow$, the result being the following:

$$
|A| = (2)(5) \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 4 \end{vmatrix}
$$

The diagonal element in the third row is 1 and there is nothing to be taken out. Now $-1(3) + (4) \Rightarrow$ and then, after having taken out 4 from the fourth row, the result is

$$|A| = (2)(5)(1)(4) \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

Now, by Axiom (4) the determinant of the remaining identity matrix is 1. Therefore, the final solution is $|A| = (2)(5)(1)(4) = 40$.

Example 1.2.2. Evaluate the determinant of the following matrix:

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

Solution 1.2.2. Since the first row, first column element is a convenient number 1 we start operating with the first row. Otherwise, we bring a convenient number to the $(1, 1)$-th position by interchanges of rows and columns (with each interchange the determinant is to be multiplied by (-1). Our aim will be to reduce the matrix to a triangular form so that the determinant is the product of the diagonal elements. By using the first row let us wipe out the elements in the first column. The operations are $-2(1) + (3)$, $-5(1) + (4) \Rightarrow$. Then

$$|A| = \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 2 & 1 & -1 & 0 \\ 5 & 2 & 1 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & -3 & -9 & -2 \\ 0 & -8 & -19 & -2 \end{vmatrix}.$$

Now, by using the second row we want to wipe out the elements below the diagonal in the second column. But the first number is 3. One element in the third row can be wiped out by simply adding $1(2) + (3) \Rightarrow$. This brings the following:

$$|A| = \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & -7 & -1 \\ 0 & -8 & -19 & -2 \end{vmatrix}.$$

If we take out 3 from the second row then it will bring in fractions. We will avoid fractions by multiplying the second row by 8 and the fourth row by 3. In order preserve the value,

we keep $\frac{1}{(8)(3)}$ outside. Then, we add the second row to the fourth row or $(2) + (4) \Rightarrow$. The result of these operations is the following:

$$|A| = \frac{1}{(8)(3)} \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 24 & 16 & 8 \\ 0 & 0 & -7 & -1 \\ 0 & -24 & -57 & -6 \end{vmatrix} = \frac{1}{(8)(3)} \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 24 & 16 & 8 \\ 0 & 0 & -7 & -1 \\ 0 & 0 & -41 & 2 \end{vmatrix}.$$

Now, multiply the third row by 41 and fourth row by 7 and then add $-1(3) + (4) \Rightarrow$. The result is the following:

$$|A| = \frac{1}{(8)(3)(7)(41)} \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 24 & 16 & 8 \\ 0 & 0 & -287 & -41 \\ 0 & 0 & 0 & 55 \end{vmatrix}.$$

Now, take the product of the diagonal elements. Then

$$|A| = \frac{(1)(24)(-287)(55)}{(8)(3)(7)(41)} = -55.$$

Observe that we did not have to repeat the 4×4 determinant each time. After wiping out the first column elements, we could have expressed the determinant as follows because only the elements in the second row and second column onward would then have mattered. That is,

$$|A| = (1) \begin{vmatrix} 3 & 2 & 1 \\ 0 & -7 & -1 \\ -8 & -19 & -2 \end{vmatrix}.$$

Similarly, after wiping out the second column elements, we could have written the resulting determinant as

$$|A| = \frac{(1)(24)}{(8)(3)} \begin{vmatrix} -7 & -1 \\ -41 & 2 \end{vmatrix},$$

and so on.

Example 1.2.3. Evaluate the determinant of a 2×2 general matrix.

Solution 1.2.3. A general 2×2 determinant can be opened up by using Axiom (3), that is,

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \text{[Axiom (3)]}$$

$$= a_{11} \begin{vmatrix} 1 & 0 \\ a_{21} & a_{22} \end{vmatrix} + a_{12} \begin{vmatrix} 0 & 1 \\ a_{21} & a_{22} \end{vmatrix} \quad \text{[Axiom (1)]}.$$

If any of a_{11} or a_{12} is zero, then the corresponding determinant is zero. In the second determinant on the right, interchange the second and first columns, which will bring a minus sign outside the determinant. That is,

$$|A| = a_{11} \begin{vmatrix} 1 & 0 \\ a_{21} & a_{22} \end{vmatrix} - a_{12} \begin{vmatrix} 1 & 0 \\ a_{22} & a_{21} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The last step is done by using the property that the determinant of a triangular matrix is the product of the diagonal elements. We can also evaluate the determinant by using a number of different procedures. Taking out a_{11} if $a_{11} \neq 0$,

$$|A| = a_{11} \begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} \\ a_{21} & a_{22} \end{vmatrix}.$$

Now, perform the operation $-a_{21}(1) + (2)$ or $-a_{21}$ times the first row is added to the second row. Then,

$$|A| = a_{11} \begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} \\ 0 & a_{22} - \frac{a_{12}a_{21}}{a_{11}} \end{vmatrix}.$$

Now, expanding by using a property of triangular matrices, we have

$$|A| = a_{11}(1)\left[a_{22} - \frac{a_{12}a_{21}}{a_{11}}\right] = a_{11}a_{22} - a_{12}a_{21}. \tag{1.2.4}$$

Consider a general 3×3 determinant evaluated by using Axiom (3) first.

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} 0 & 1 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} 0 & 0 & 1 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} 0 & 1 & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} 0 & 0 & 1 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}.$$

The first step consists in opening up the first row by making use of Axiom (3). Then, eliminate the elements in rows 2 and 3 within the column headed by 1. The next step is to bring the columns whose first element is 1 to the first column position by transpositions. The first matrix on the right-hand side is already in this format. One transposition is needed in the second matrix and two are required in the third matrix. After completing the transpositions, the next step consists in opening up each determinant along their second row and observing that the resulting matrices are lower triangular or can be made so after transposing their last two columns. The final result is then obtained. The last two steps are executed below:

$$|A| = a_{11} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{21} & a_{23} \\ 0 & a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{21} & a_{22} \\ 0 & a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}[a_{22}a_{33} - a_{23}a_{32}] - a_{12}[a_{21}a_{33} - a_{23}a_{31}] + a_{13}[a_{21}a_{32} - a_{22}a_{31}]$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

(1.2.5)

A few observations are in order. Once 1 is brought to the first row first column position in every matrix and the remaining elements in this first column are eliminated, one can delete the first row and first column and take the determinant of the remaining submatrix because only those elements will enter into the remaining operations involving opening up the second and successive rows by making use of Axiom (3). Hence, we could have written

$$|A| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

This step is also called *the cofactor expansion of the matrix*. In a general matrix $A = (a_{ij})$, *the cofactor of the element a_{ij} is equal to $(-1)^{i+j} M_{ij}$ where M_{ij} is the minor of a_{ij}. This minor is obtained by deleting the i-th row and j-the column and then taking the determinant of the remaining elements*. The second item to be noted from (1.2.5) is that, in the final expression for $|A|$, each term has one and only one element from each row and each column of A. Some elements have plus signs in front of them and others have minus signs. For each term, write the first subscript in the natural order 1, 2, 3 for the 3×3 case and in the general $n \times n$ case, write the first subscripts in the natural order $1, 2, \ldots, n$. Now, examine the second subscripts. Let the number of transpositions needed to bring the second subscripts into the natural order $1, 2, \ldots, n$ be ρ. Then, that term is multiplied by $(-1)^\rho$ so that an even number of transpositions produces a plus sign and an odd number of transpositions brings a minus sign, or equivalently if ρ is even, the coefficient is plus 1 and if ρ is odd, the coefficient is -1. This also enables us to open up a general determinant. This will be considered after pointing out one more property for

a 3×3 case. The final representation in the 3×3 case in (1.2.5) can also be written up by using the following mechanical procedure. Write all elements in the matrix A in the natural order. Then, augment this arrangement with the first two columns. This yields the following format:

$$
\begin{array}{ccccc}
a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\
a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\
a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \,.
\end{array}
$$

Now take the products of the elements along the diagonals going from the top left to the bottom right. These are the elements with the plus sign. Take the products of the elements in the second diagonals or the diagonals going from the bottom left to the top right. These are the elements with minus sign. As a result, $|A|$ is as follows:

$$
\begin{aligned}
|A| = {}& [a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}] \\
& - [a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}].
\end{aligned}
$$

This mechanical procedure applies only in the 3×3 case. The general expansion is the following:

$$
|A| = \begin{vmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \ldots & a_{nn}
\end{vmatrix} = \sum_{i_1} \cdots \sum_{i_n} (-1)^{\rho(i_1,\ldots,i_n)} a_{1i_1} a_{2i_2} \cdots a_{ni_n} \tag{1.2.6}
$$

where $\rho(i_1, \ldots, i_n)$ is the number of transpositions needed to bring the second subscripts $i_1, \ldots, i_n$ into the natural order $1, 2, \ldots, n$.

The cofactor expansion of a general matrix is obtained as follows: Suppose that we open up a $n \times n$ determinant A along the i-th row using Axiom (3). Then, after taking out $a_{i1}, a_{i2}, \ldots, a_{in}$, we obtain n determinants where, in the first determinant, 1 occupies the $(i, 1)$-th position, in the second one, 1 is at the $(i, 2)$-th position and so on so that, in the j-th determinant, 1 occupies the (i, j)-th position. Given that i-th row, we can now eliminate all the elements in the columns corresponding to the remaining 1. We now bring this 1 into the first row first column position by transpositions in each determinant. The number of transpositions needed to bring this 1 from the j-th position in the i-th row to the first position in the i-th row, is $j - 1$. Then, to bring that 1 to the first row first column position, another $i - 1$ transpositions are required, so that the total number of transpositions needed is $(i - 1) + (j - 1) = i + j - 2$. Hence, the multiplicative factor is $(-1)^{i+j-2} = (-1)^{i+j}$, and the expansion is as follows:

$$
\begin{aligned}
|A| &= (-1)^{i+1} a_{i1} M_{i1} + (-1)^{i+2} a_{i2} M_{i2} + \cdots + (-1)^{i+n} a_{in} M_{in} \\
&= a_{i1} C_{i1} + a_{i2} C_{i2} + \cdots + a_{in} C_{in}
\end{aligned} \tag{1.2.7}
$$

where $C_{ij} = (-1)^{i+j} M_{ij}$, C_{ij} is *the cofactor of a_{ij} and M_{ij} is the minor of a_{ij}*, the minor being obtained by taking the determinant of the remaining elements after deleting the i-th row and j-th column of A. Moreover, if we expand along a certain row (column) and the cofactors of some other row (column), then the result will be zero. That is,

$$0 = a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}, \text{ for all } i \neq j. \tag{1.2.8}$$

Inverse of a Matrix Regular inverses exist only for square matrices that are nonsingular. The standard notation for a regular inverse of a matrix A is A^{-1}. It is defined as $AA^{-1} = I_n$ and $A^{-1}A = I_n$. The following properties can be deduced from the definition. First, we note that $AA^{-1} = A^0 = I = A^{-1}A$. When A and B are $n \times n$ nonsingular matrices, then $(AB)^{-1} = B^{-1}A^{-1}$, which can be established by pre- or post-multiplying the right-hand side side by AB. Accordingly, with $A^m = A \times A \times \cdots \times A$, $A^{-m} = A^{-1} \times \cdots \times A^{-1} = (A^m)^{-1}$, $m = 1, 2, \ldots$, and when $A_1, \ldots, A_k$ are $n \times n$ nonsingular matrices, $(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_2^{-1} A_1^{-1}$. We can also obtain a formula for the inverse of a nonsingular matrix A in terms of cofactors. Assuming that A^{-1} exist and letting $\text{Cof}(A) = (C_{ij})$ be the matrix of cofactors of A, that is, if $A = (a_{ij})$ and if C_{ij} is the cofactor of a_{ij} then $\text{Cof}(A) = (C_{ij})$. It follows from (1.2.7) and (1.2.8) that

$$A^{-1} = \frac{1}{|A|}(\text{Cof}(A))' = \frac{1}{|A|} \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}', \tag{1.2.9}$$

that is, the transpose of the cofactor matrix divided by the determinant of A. What about $A^{\frac{1}{2}}$? For a scalar quantity a, we have the definition that if b exists such that $b \times b = a$, then b is a square root of a. Consider the following 2×2 matrices:

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I_2^2 = I_2, B_j^2 = I_2, j = 1, 2, 3, 4.$$

Thus, if we use the definition $B^2 = A$ and claim that B is the square root of A, there are several candidates for B; this means that, in general, the square root of a matrix cannot be uniquely determined. However, if we restrict ourselves to the class of positive definite matrices, then a square root can be uniquely defined. The definiteness of matrices will be considered later.

1.2.1. Inverses by row operations or elementary operations

Basic elementary matrices are of two types. Let us call them the E-type and the F-type. An elementary matrix of the E-type is obtained by taking an identity matrix and multiplying any row (column) by a nonzero scalar. For example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ E_2 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

where E_1, E_2, E_3 are elementary matrices of the E-type obtained from the identity matrix I_3. If we pre-multiply an arbitrary matrix A with an elementary matrix of the E-type, then the same effect will be observed on the rows of the arbitrary matrix A. For example, consider a 3×3 matrix $A = (a_{ij})$. Then, for example,

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ -2a_{21} & -2a_{22} & -2a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Thus, the same effect applies to the rows, that is, the second row is multiplied by (-2). Observe that E-type elementary matrices are always nonsingular and so, their regular inverses exist. For instance,

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ E_1 E_1^{-1} = I_3 = E_1^{-1} E_1; \ E_2^{-1} = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ E_2 E_2^{-1} = I_3.$$

Observe that post-multiplication of an arbitrary matrix by an E-type elementary matrix will have the same effect on the columns of the arbitrary matrix. For example, AE_1 will have the same effect on the columns of A, that is, the second column of A is multiplied by -2; AE_2 will result in the first column of A being multiplied by 5, and so on. The F-type elementary matrix is created by adding any particular row of an identity matrix to another one of its rows. For example, consider a 3×3 identity matrix I_3 and let

$$F_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ F_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \ F_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

where F_1 is obtained by adding the first row to the second row of I_3; F_2 is obtained by adding the first row to the third row of I_3; and F_3 is obtained by adding the second row of I_3 to the third row. As well, F-type elementary matrices are nonsingular, and for instance,

$$F_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; F_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \; R_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

where $F_1 F_1^{-1} = I_3$, $F_2^{-1} F_2 = I_3$ and $F_3^{-1} F_3 = I_3$. If we pre-multiply an arbitrary matrix A by an F-type elementary matrix, then the same effect will be observed on the rows of A. For example,

$$F_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + a_{11} & a_{22} + a_{12} & a_{23} + a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Thus, the same effect applies to the rows, namely, the first row is added to the second row in A (as F_1 was obtained by adding the first row of I_3 to the second row of I_3). The reader may verify that $F_2 A$ has the effect of the first row being added to the third row and $F_3 A$ will have the effect of the second row being added to the third row. By combining E- and F-type elementary matrices, we end up with a G-type matrix wherein a multiple of any particular row of an identity matrix is added to another one of its rows. For example, letting

$$G_1 = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } G_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix},$$

it is seen that G_1 is obtained by adding 5 times the first row to the second row in I_3, and G_2 is obtained by adding -2 times the first row to the third row in I_3. Pre-multiplication of an arbitrary matrix A by G_1, that is, $G_1 A$, will have the effect that 5 times the first row of A will be added to its second row. Similarly, G_2 will have the effect that -2 times the first row of A will be added to its third row. Being product of E- and F-type elementary matrices, G-type matrices are also nonsingular. We also have the result that if A, B, C are $n \times n$ matrices and $B = C$, then $AB = AC$ as long as A is nonsingular. In general, if $A_1, \dots, A_k$ are $n \times n$ nonsingular matrices, we have

$$B = C \Rightarrow A_k A_{k-1} \cdots A_2 A_1 B = A_k A_{k-1} \cdots A_2 A_1 C;$$
$$\Rightarrow A_1 A_2 B = A_1 (A_2 B) = (A_1 A_2) B = (A_1 A_2) C = A_1 (A_2 C).$$

We will evaluate the inverse of a nonsingular square matrix by making use of elementary matrices. The procedure will also verify whether a regular inverse exists for a given matrix. If a regular inverse for a square matrix A exists, then $AA^{-1} = I$. We can pre- or post-multiply A by elementary matrices. For example,

$$
\begin{aligned}
AA^{-1} = I &\Rightarrow E_k F_r \cdots E_1 F_1 A A^{-1} = E_k F_r \cdots E_1 F_1 I \\
&\Rightarrow (E_k \cdots F_1 A) A^{-1} = (E_k \cdots F_1).
\end{aligned}
$$

Thus, if the operations $E_k \cdots F_1$ on A reduced A to an identity matrix, then A^{-1} is $E_k \cdots F_1$. If an inconsistency has occurred during the process, we can conclude that there is no inverse for A. Hence, our aim in performing our elementary operations on the left of A is to reduce it to an identity matrix, in which case the product of the elementary matrices on the right-hand side of the last equation will produce the inverse of A.

Example 1.2.4. Evaluate A^{-1} if it exists, where

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

Solution 1.2.4. If A^{-1} exists then $AA^{-1} = I$ which means

$$
\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & -1 & 1 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
$$

This is our starting equation. Only the configuration of the elements matters. The matrix notations and the symbol A^{-1} can be disregarded. Hence, we consider only the configuration of the numbers of the matrix A on the left and the numbers in the identity matrix on the right. Then we pre-multiply A and pre-multiply the identity matrix by only making use of elementary matrices. In the first set of steps, our aim consists in reducing every element in the first column of A to zeros, except the first one, by only using the first row. For each elementary operation on A, the same elementary operation is done on the identity matrix also. Now, utilizing the second row of the resulting A, we reduce all the elements in the second column of A to zeros except the second one and continue in this manner until all the elements in the last columns except the last one are reduced to zeros by making use of the last row, thus reducing A to an identity matrix, provided of course that A is nonsingular. In our example, the elements in the first column can be made equal to zeros

by applying the following operations. We will employ the following standard notation: $a(i) + (j) \Rightarrow$ meaning a times the i-th row is added to the j-th row, giving the result. Consider $(1) + (2)$; $-2(1) + (3)$; $-1(1) + (4) \Rightarrow$ (that is, the first row is added to the second row; and then -2 times the first row is added to the third row; then -1 times the first row is added to the fourth row), (for each elementary operation on A we do the same operation on the identity matrix also) the net result being

$$
\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & -2 & 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.
$$

Now, start with the second row of the resulting A and the resulting identity matrix and try to eliminate all the other elements in the second column of the resulting A. This can be achieved by performing the following operations: $(2) + (3)$; $-1(2) + (1) \Rightarrow$

$$
\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -2 & 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}
$$

Now, start with the third row and eliminate all other elements in the third column. This can be achieved by the following operations. Writing the row used in the operations (the third one in this case) within the first set of parentheses for each operation, we have $2(3) + (4)$; $-2(3) + (2)$; $(3) + (1) \Rightarrow$

$$
\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \Leftrightarrow \begin{pmatrix} -1 & 0 & 1 & 0 \\ 3 & -1 & -2 & 0 \\ -1 & 1 & 1 & 0 \\ -3 & 2 & 2 & 1 \end{pmatrix}
$$

Divide the 4th row by 2 and then perform the following operations: $\frac{1}{2}(4)$; $-1(4) + (3)$; $(4) + (2)$; $-1(4) + (1) \Rightarrow$

$$
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Leftrightarrow \begin{pmatrix} \frac{1}{2} & -1 & 0 & -\frac{1}{2} \\ \frac{3}{2} & 0 & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 1 & 1 & \frac{1}{2} \end{pmatrix}.
$$

Thus,

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & -1 & 0 & -\frac{1}{2} \\ \frac{3}{2} & 0 & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 1 & 1 & \frac{1}{2} \end{bmatrix}.$$

This result should be verified to ensure that it is free of computational errors. Since

$$AA^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -1 & 0 & -\frac{1}{2} \\ \frac{3}{2} & 0 & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 1 & 1 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

the result is indeed correct.

Example 1.2.5. Evaluate A^{-1} if it exists where

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

Solution 1.2.5. If A^{-1} exists, then $AA^{-1} = I_3$. Write

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 2 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Starting with the first row, eliminate all other elements in the first column with the following operations: $-1(1) + (2)$; $-2(1) + (3) \Rightarrow$

$$\begin{matrix} 1 & 1 & 1 & & 1 & 0 & 0 \\ 0 & -2 & 0 & \Leftrightarrow & -1 & 1 & 0 \\ 0 & -2 & 0 & & -2 & 0 & 1 \end{matrix}$$

The second and third rows on the left side being identical, the left-hand side matrix is singular, which means that A is singular. Thus, the inverse of A does not exist in this case.

1.3. Determinants of Partitioned Matrices

Consider a matrix A written in the following format:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ where } A_{11}, A_{12}, A_{21}, A_{22} \text{ are submatrices.}$$

For example,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \ A_{11} = [a_{11}], \ A_{12} = [a_{12} \ a_{13}], \ A_{21} = \begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix}$$

and $A_{22} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$. The above is a 2×2 partitioning or a partitioning into two sub-matrices by two sub-matrices. But a 2×2 partitioning is not unique. We may also consider

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \ A_{22} = [a_{33}], \ A_{12} = \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}, \ A_{21} = [a_{31} \ a_{32}],$$

which is another 2×2 partitioning of A. We can also have a 1×2 or 2×1 partitioning into sub-matrices. We may observe one interesting property. Consider a block diagonal matrix. Let

$$A = \begin{bmatrix} A_{11} & O \\ O & A_{22} \end{bmatrix} \Rightarrow |A| = \begin{vmatrix} A_{11} & O \\ O & A_{22} \end{vmatrix},$$

where A_{11} is $r \times r$, A_{22} is $s \times s$, $r + s = n$ and O indicates a null matrix. Observe that when we evaluate the determinant, all the operations on the first r rows will produce the determinant of A_{11} as a coefficient, without affecting A_{22}, leaving an $r \times r$ identity matrix in the place of A_{11}. Similarly, all the operations on the last s rows will produce the determinant of A_{22} as a coefficient, leaving an $s \times s$ identity matrix in place of A_{22}. In other words, for a diagonal block matrix whose diagonal blocks are A_{11} and A_{22},

$$|A| = |A_{11}| \times |A_{22}|. \tag{1.3.1}$$

Given a triangular block matrix, be it lower or upper triangular, then its determinant is also the product of the determinants of the diagonal blocks. For example, consider

$$A = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix} \Rightarrow |A| = \begin{vmatrix} A_{11} & A_{12} \\ O & A_{22} \end{vmatrix}.$$

By using A_{22}, we can eliminate A_{12} without affecting A_{11} and hence, we can reduce the matrix of the determinant to a diagonal block form without affecting the value of the determinant. Accordingly, the determinant of an upper or lower triangular block matrix whose diagonal blocks are A_{11} and A_{22}, is

$$|A| = |A_{11}| \, |A_{22}|. \tag{1.3.2}$$

Partitioning is done to accommodate further operations such as matrix multiplication. Let A and B be two matrices whose product AB is defined. Suppose that we consider a

2×2 partitioning of A and B into sub-matrices; if the multiplication is performed treating the sub-matrices as if they were scalar quantities, the following format is obtained:

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$
$$= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} .$$

If all the products of sub-matrices on the right-hand side are defined, then we say that A *and* B *are conformably partitioned for the product* AB. Let A be a $n \times n$ matrix whose determinant is defined. Let us consider the 2×2 partitioning

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ where } A_{11} \text{ is } r \times r, \ A_{22} \text{ is } s \times s, \ r + s = n.$$

Then, A_{12} is $r \times s$ and A_{21} is $s \times r$. In this case, the first row block is $[A_{11} \ A_{12}]$ and the second row block is $[A_{21} \ A_{22}]$. When evaluating a determinant, we can add linear functions of rows to any other row or linear functions of rows to other blocks of rows without affecting the value of the determinant. What sort of a linear function of the first row block could be added to the second row block so that a null matrix O appears in the position of A_{21}? It is $-A_{21}A_{11}^{-1}$ times the first row block. Then, we have

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} \\ O & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{vmatrix}.$$

This is a triangular block matrix and hence its determinant is the product of the determinants of the diagonal blocks. That is,

$$|A| = |A_{11}| \, |A_{22} - A_{21}A_{11}^{-1}A_{12}|, \ |A_{11}| \neq 0.$$

From symmetry, it follows that

$$|A| = |A_{22}| \, |A_{11} - A_{12}A_{22}^{-1}A_{21}|, \ |A_{22}| \neq 0. \tag{1.3.3}$$

Let us now examine the inverses of partitioned matrices. Let A and A^{-1} be conformably partitioned for the product AA^{-1}. Consider a 2×2 partitioning of both A and A^{-1}. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

where A_{11} and A^{11} are $r \times r$ and A_{22} and A^{22} are $s \times s$ with $r + s = n$, A is $n \times n$ and nonsingular. $AA^{-1} = I$ gives the following:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & I_s \end{bmatrix}.$$

That is,

$$A_{11}A^{11} + A_{12}A^{21} = I_r \tag{i}$$
$$A_{11}A^{12} + A_{12}A^{22} = O \tag{ii}$$
$$A_{21}A^{11} + A_{22}A^{21} = O \tag{iii}$$
$$A_{21}A^{12} + A_{22}A^{22} = I_s. \tag{iv}$$

From *(ii)*, $A^{12} = -A_{11}^{-1}A_{12}A^{22}$. Substituting in *(iv)*,

$$A_{21}[-A_{11}^{-1}A_{12}A^{22}] + A_{22}A^{22} = I_s \Rightarrow [A_{22} - A_{21}A_{11}^{-1}A_{12}]A^{22} = I_s.$$

That is,

$$A^{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, |A_{11}| \neq 0, \tag{1.3.4}$$

and, from symmetry, it follows that

$$A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, |A_{22}| \neq 0 \tag{1.3.5}$$
$$A_{11} = (A^{11} - A^{12}(A^{22})^{-1}A^{21})^{-1}, |A^{22}| \neq 0 \tag{1.3.6}$$
$$A_{22} = (A^{22} - A^{21}(A^{11})^{-1}A^{12})^{-1}, |A^{11}| \neq 0. \tag{1.3.7}$$

The rectangular components A_{12}, A_{21}, A^{12}, A^{21} can also be evaluated in terms of the sub-matrices by making use of Eqs. *(i)–(iv)*.

1.4. Eigenvalues and Eigenvectors

Let A be $n \times n$ matrix, X be an $n \times 1$ vector, and λ be a scalar quantity. Consider the equation

$$AX = \lambda X \Rightarrow (A - \lambda I)X = O.$$

Observe that $X = O$ is always a solution. If this equation has a non-null vector X as a solution, then the determinant of the coefficient matrix must be zero because this matrix must be singular. If the matrix $(A - \lambda I)$ were nonsingular, its inverse $(A - \lambda I)^{-1}$ would exist and then, on pre-multiplying $(A - I)X = O$ by $(A - \lambda I)^{-1}$, we would have $X = O$, which is inadmissible since $X \neq O$. That is,

$$|A - \lambda I| = 0, \ \lambda \text{ being a scalar quantity.} \tag{1.4.1}$$

Since the matrix A is $n \times n$, equation (1.4.1) has n roots, which will be denoted by $\lambda_1, \ldots, \lambda_n$. Then

$$|A - \lambda I| = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda), \ AX_j = \lambda_j X_j.$$

Then, $\lambda_1, \ldots, \lambda_n$ are called *the eigenvalues of A and $X_j \neq O$, an eigenvector corresponding to the eigenvalue λ_j.*

Example 1.4.1. Compute the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$.

Solution 1.4.1. Consider the equation

$$|A - \lambda I| = 0 \Rightarrow \left\| \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\| = 0 \Rightarrow$$

$$\begin{vmatrix} 1 - \lambda & 1 \\ 1 & 2 - \lambda \end{vmatrix} = 0 \Rightarrow (1 - \lambda)(2 - \lambda) - 1 = 0 \Rightarrow \lambda^2 - 3\lambda + 1 = 0 \Rightarrow$$

$$\lambda = \frac{3 \pm \sqrt{(9 - 4)}}{2} \Rightarrow \lambda_1 = \frac{3}{2} + \frac{\sqrt{5}}{2}, \ \lambda_2 = \frac{3}{2} - \frac{\sqrt{5}}{2}.$$

An eigenvector X_1 corresponding to $\lambda_1 = \frac{3}{2} + \frac{\sqrt{5}}{2}$ is given by $AX_1 = \lambda_1 X_1$ or $(A - \lambda_1 I)X_1 = O$. That is,

$$\begin{bmatrix} 1 - (\frac{3}{2} + \frac{\sqrt{5}}{2}) & 1 \\ 1 & 2 - (\frac{3}{2} + \frac{\sqrt{5}}{2}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow$$

$$\left(-\frac{1}{2} - \frac{\sqrt{5}}{2} \right) x_1 + x_2 = 0, \tag{i}$$

$$x_1 + \left(\frac{1}{2} - \frac{\sqrt{5}}{2} \right) x_2 = 0. \tag{ii}$$

Since $A - \lambda_1 I$ is singular, both *(i)* and *(ii)* must give the same solution. Letting $x_2 = 1$ in *(ii)*, $x_1 = -\frac{1}{2} + \frac{\sqrt{5}}{2}$. Thus, one solution X_1 is

$$X_1 = \begin{bmatrix} -\frac{1}{2} + \frac{\sqrt{5}}{2} \\ 1 \end{bmatrix}.$$

Any nonzero constant multiple of X_1 is also a solution to $(A - \lambda_1 I)X_1 = O$. An eigenvector X_2 corresponding to the eigenvalue λ_2 is given by $(A - \lambda_2 I)X_2 = O$. That is,

$$\left(-\frac{1}{2} + \frac{\sqrt{5}}{2} \right) x_1 + x_2 = 0, \tag{iii}$$

$$x_1 + \left(\frac{1}{2} + \frac{\sqrt{5}}{2} \right) x_2 = 0. \tag{iv}$$

Hence, one solution is

$$X_2 = \left[\begin{array}{c} -\frac{1}{2} - \frac{\sqrt{5}}{2} \\ 1 \end{array} \right].$$

Any nonzero constant multiple of X_2 is also an eigenvector corresponding to λ_2.

Even if all elements of a matrix A are real, its eigenvalues can be real, positive, negative, zero, irrational or complex. Complex and irrational roots appear in pairs. If $a + ib$, $i = \sqrt{(-1)}$, and a, b real, is an eigenvalue, then $a - ib$ is also an eigenvalue of the same matrix. The following properties can be deduced from the definition:

(1): *The eigenvalues of a diagonal matrix are its diagonal elements*;

(2): *The eigenvalues of a triangular (lower or upper) matrix are its diagonal elements*;

(3): *If any eigenvalue is zero, then the matrix is singular and its determinant is zero*;

(4): *If λ is an eigenvalue of A and if A is nonsingular, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}*;

(5): *If λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k, $k = 1, 2, \ldots$, their associated eigenvector being the same*;

(7): *The eigenvalues of an identity matrix are unities; however, the converse need not be true*;

(8): *The eigenvalues of a scalar matrix with diagonal elements $a, a, \ldots, a$ are a repeated n times when the order of A is n; however, the converse need not be true*;

(9): *The eigenvalues of an orthonormal matrix, $AA' = I$, $A'A = I$, are ± 1; however, the converse need not be true*;

(10): *The eigenvalues of an idempotent matrix, $A = A^2$, are ones and zeros; however, the converse need not be true. The only nonsingular idempotent matrix is the identity matrix*;

(11): *At least one of the eigenvalues of a nilpotent matrix of order r, that is, $A \neq O, \ldots, A^{r-1} \neq O, A^r = O$, is null*;

(12): *For an $n \times n$ matrix, both A and A' have the same eigenvalues*;

(13): *The eigenvalues of a symmetric matrix are real*;

(14): *The eigenvalues of a skew symmetric matrix can only be zeros and purely imaginary numbers*;

(15): *The determinant of A is the product of its eigenvalues: $|A| = \lambda_1 \lambda_2 \cdots \lambda_n$*;

(16): *The trace of a square matrix is equal to the sum of its eigenvalues*;

(17): *If $A = A'$ (symmetric), then the eigenvectors corresponding to distinct eigenvalues are orthogonal*;

(18): *If $A = A'$ and A is $n \times n$, then there exists a full set of n eigenvectors which are linearly independent, even if some eigenvalues are repeated.*

Result (16) requires some explanation. We have already derived the following two results:

$$|A| = \sum_{i_1} \cdots \sum_{i_n} (-1)^\rho a_{1i_1} a_{2i_2} \cdots a_{ni_n} \qquad (v)$$

and

$$|A - \lambda I| = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda). \qquad (vi)$$

Equation (vi) yields a polynomial of degree n in λ where λ is a variable. When $\lambda = 0$, we have $|A| = \lambda_1 \lambda_2 \cdots \lambda_n$, the product of the eigenvalues. The term containing $(-1)^n \lambda^n$, when writing $|A - \lambda I|$ in the format of equation (v), can only come from the term $(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda)$ (refer to the explicit form for the 3×3 case discussed earlier). Two factors containing λ will be missing in the next term with the highest power of λ. Hence, $(-1)^n \lambda^n$ and $(-1)^{n-1} \lambda^{n-1}$ can only come from the term $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$, as can be seen from the expansion in the 3×3 case discussed in detail earlier. From (v) the coefficient of $(-1)^{n-1} \lambda^{n-1}$ is $a_{11} + a_{22} + \cdots + a_{nn} = \text{tr}(A)$ and from (vi), the coefficient of $(-1)^{n-1} \lambda^{n-1}$ is $\lambda_1 + \cdots + \lambda_n$. Hence $\text{tr}(A) = \lambda_1 + \cdot + \lambda_n = $ sum of the eigenvalues of A. This does not mean that $\lambda_1 = a_{11}, \lambda_2 = a_{22}, \ldots, \lambda_n = a_{nn}$, only that the sums will be equal.

Matrices in the Complex Domain When the elements in $A = (a_{ij})$ can also be complex quantities, then a typical element in A will be of the form $a + ib$, $i = \sqrt{(-1)}$, and a, b real. The complex conjugate of A will be denoted by $\bar{A}$ and the conjugate transpose will be denoted by A^*. Then, for example,

$$A = \begin{bmatrix} 1+i & 2i & 3-i \\ 4i & 5 & 1+i \\ 2-i & i & 3+i \end{bmatrix} \Rightarrow \bar{A} = \begin{bmatrix} 1-i & -2i & 3+i \\ -4i & 5 & 1-i \\ 2+i & -i & 3-i \end{bmatrix}, A^* = \begin{bmatrix} 1-i & -4i & 2+i \\ -2i & 5 & -i \\ 3+i & 1-i & 3-i \end{bmatrix}.$$

Thus, we can also write $A^* = (\bar{A})' = (\bar{A}')$. When a matrix A is in the complex domain, we may write it as $A = A_1 + i A_2$ where A_1 and A_2 are real matrices. Then $\bar{A} = A_1 - i A_2$ and $A^* = A_1' - i A_2'$. In the above example,

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 5 & 1 \\ 2 & 0 & 3 \end{bmatrix} + i \begin{bmatrix} 1 & 2 & -1 \\ 4 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix} \Rightarrow$$

$$A_1 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 5 & 1 \\ 2 & 0 & 3 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 2 & -1 \\ 4 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix}.$$

A Hermitian Matrix If $A = A^*$, then A is called a Hermitian matrix. In the representation $A = A_1 + i A_2$, if $A = A^*$, then $A_1 = A_1'$ or A_1 is real symmetric, and $A_2 = -A_2'$ or A_2 is real skew symmetric. Note that when X is an $n \times 1$ vector, $X^* X$ is real. Let

$$X = \begin{bmatrix} 2 \\ 3+i \\ 2i \end{bmatrix} \Rightarrow \bar{X} = \begin{bmatrix} 2 \\ 3-i \\ -2i \end{bmatrix}, \quad X^* = [2 \ 3-i \ -2i],$$

$$X^*X = [2 \ 3-i \ -2i] \begin{bmatrix} 2 \\ 3+i \\ 2i \end{bmatrix} = (2^2 + 0^2) + (3^2 + 1^2) + (0^2 + (-2)^2) = 18.$$

Consider the eigenvalues of a Hermitian matrix A, which are the solutions of $|A - \lambda I| = 0$. As in the real case, λ may be real, positive, negative, zero, irrational or complex. Then, for $X \neq O$,

$$AX = \lambda X \Rightarrow \qquad\qquad\qquad (i)$$

$$X^*A^* = \bar{\lambda} X^* \qquad\qquad\qquad (ii)$$

by taking the conjugate transpose. Since λ is scalar, its conjugate transpose is $\bar{\lambda}$. Pre-multiply *(i)* by X^* and post-multiply *(ii)* by X. Then for $X \neq O$, we have

$$X^*AX = \lambda X^*X \qquad\qquad\qquad (iii)$$

$$X^*A^*X = \bar{\lambda} X^*X. \qquad\qquad\qquad (iv)$$

When A is Hermitian, $A = A^*$, and so, the left-hand sides of *(iii)* and *(iv)* are the same. On subtracting *(iv)* from *(iii)*, we have $0 = (\lambda - \bar{\lambda})X^*X$ where X^*X is real and positive, and hence $\lambda - \bar{\lambda} = 0$, which means that the imaginary part is zero or λ is real. If A is skew Hermitian, then we end up with $\lambda + \bar{\lambda} = 0 \Rightarrow \lambda$ is zero or purely imaginary. The above procedure also holds for matrices in the real domain. Thus, in addition to properties (13) and (14), we have the following properties:

(19) *The eigenvalues of a Hermitian matrix are real; however, the converse need not be true*;

(20) *The eigenvalues of a skew Hermitian matrix are zero or purely imaginary; however, the converse need not be true.*

1.5. Definiteness of Matrices, Quadratic and Hermitian Forms

Let X be an $n \times 1$ vector of real scalar variables $x_1, \ldots, x_n$ so that $X' = (x_1, \ldots, x_n)$. Let $A = (a_{ij})$ be a real $n \times n$ matrix. Then, all the elements of the *quadratic form*, $u = X'AX$, are of degree 2. One can always write A as an equivalent symmetric matrix when A is the matrix in a quadratic form. Hence, without any loss of generality, we may assume $A = A'$ (symmetric) when A appears in a quadratic form $u = X'AX$.

Definiteness of a quadratic form and definiteness of a matrix are only defined for $A = A'$ (symmetric) in the real domain and for $A = A^*$ (Hermitian) in the complex domain. Hence, the basic starting condition is that either $A = A'$ or $A = A^*$. If for *all* non-null X, that is, $X \neq O$, $X'AX > 0$, $A = A'$, *then A is said to be a positive definite matrix and $X'AX > 0$ is called a positive definite quadratic form*. If for *all* non-null X, $X^*AX > 0$, $A = A^*$, *then A is referred to as a Hermitian positive definite matrix and the corresponding Hermitian form $X^*AX > 0$, as Hermitian positive definite*. Similarly, if for all non-null X, $X'AX \geq 0$, $X^*AX \geq 0$, then A is positive semi-definite or Hermitian positive semi-definite. If for all non-null X, $X'AX < 0$, $X^*AX < 0$, then A is negative definite and if $X'AX \leq 0$, $X^*AX \leq 0$, then A is negative semi-definite. The standard notations being utilized are as follows:

$A > O$ (A and $X'AX > 0$ are *real positive definite*; (O is a capital o and not zero)
$A \geq O$ (A and $X'AX \geq 0$ are *positive semi-definite*)
$A < O$ (A and $X'AX < 0$ are *negative definite*)
$A \leq O$ (A and $X'AX \leq 0$ are *negative semi-definite*).

All other matrices, which do no belong to any of those four categories, are called *indefinite matrices*. That is, for example, A is such that for some X, $X'AX > 0$, and for some other X, $X'AX < 0$, then A is an indefinite matrix. The corresponding Hermitian cases are:

$$A > O, X^*AX > 0 \text{ (Hermitian positive definite)}$$
$$A \geq O, X^*AX \geq 0 \text{ (Hermitian positive semi-definite)}$$
$$A < O, X^*AX < 0 \text{ (Hermitian negative definite)}$$
$$A \leq O, X^*AX \leq 0 \text{ (Hermitian negative semi-definite).} \qquad (1.5.1)$$

In all other cases, the matrix A and the Hermitian form X^*AX are indefinite. Certain conditions for the definiteness of $A = A'$ or $A = A^*$ are the following:
(1) *All the eigenvalues of A are positive $\Leftrightarrow A > O$; all eigenvalues are greater than or equal to zero $\Leftrightarrow A \geq O$; all eigenvalues are negative $\Leftrightarrow A < O$; all eigenvalues are ≤ 0 $\Leftrightarrow A \leq O$; all other matrices $A = A'$ or $A = A^*$ for which some eigenvalues are positive and some others are negative are indefinite.*
(2) *$A = A'$ or $A = A^*$ and all the leading minors of A are positive (leading minors are determinants of the leading sub-matrices, the r-th leading sub-matrix being obtained by deleting all rows and columns from the $(r + 1)$-th onward), then $A > O$; if the leading minors are ≥ 0, then $A \geq O$; if all the odd order minors are negative and all the even order minors are positive, then $A < O$; if the odd order minors are ≤ 0 and the even order*

minors are ≥ 0, then $A \leq O$; all other matrices are indefinite. If $A \neq A'$ or $A \neq A^*$, then no definiteness can be defined in terms of the eigenvalues or leading minors. Let

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}.$$

Note that A is real symmetric as well as Hermitian. Since $X'AX = 2x_1^2 + 5x_2^2 > 0$ for all real x_1 and x_2 as long as both x_1 and x_2 are not both equal to zero, $A > O$ (positive definite). $X^*AX = 2|x_1|^2 + 5|x_2|^2 = 2[\sqrt{(x_{11}^2 + x_{12}^2)}]^2 + 5[\sqrt{(x_{21}^2 + x_{22}^2)}]^2 > 0$ for all $x_{11}, x_{12}, x_{21}, x_{22}$, as long as all are not simultaneously equal to zero, where $x_1 = x_{11} + ix_{12}, x_2 = x_{21} + ix_{22}$ with $x_{11}, x_{12}, x_{21}, x_{22}$ being real and $i = \sqrt{(-1)}$. Consider

$$B = \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 0 \\ 0 & -6 \end{bmatrix}.$$

Then, $B < O$ and C is indefinite. Consider the following symmetric matrices:

$$A_1 = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} -2 & 1 \\ 1 & -5 \end{bmatrix}.$$

The leading minors of A_1 are $5 > 0$, $\begin{vmatrix} 5 & 2 \\ 2 & 4 \end{vmatrix} = 16 > 0$. The leading minors of A_2 are $2 > 0$, $\begin{vmatrix} 2 & 2 \\ 2 & 1 \end{vmatrix} = -2 < 0$. The leading minors of A_3 are $-2 < 0$, $\begin{vmatrix} -2 & 1 \\ 1 & -5 \end{vmatrix} = 9 > 0$. Hence $A_1 > O$, $A_3 < O$ and A_2 is indefinite. The following results will be useful when reducing a quadratic form or Hermitian form to its canonical form.

(3) *For every real $A = A'$ (symmetric), there exists an orthonormal matrix Q, $QQ' = I$, $Q'Q = I$ such that $Q'AQ = \text{diag}(\lambda_1, \ldots, \lambda_n)$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the $n \times n$ matrix A and $\text{diag}(\ldots)$ denotes a diagonal matrix.* In this case, a real quadratic form will reduce to the following linear combination:

$$X'AX = Y'Q'AQY = Y'\text{diag}(\lambda_1, \ldots, \lambda_n)Y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2, \quad Y = Q'X. \quad (1.5.2)$$

(4): *For every Hermitian matrix $A = A^*$, there exists a unitary matrix U, $U^*U = I$, $UU^* = I$ such that*

$$X^*AX = Y^*\text{diag}(\lambda_1, \ldots, \lambda_n)Y = \lambda_1|y_1|^2 + \cdots + \lambda_n|y_n|^2, \quad Y = U^*X. \quad (1.5.3)$$

When $A > O$ (real positive definite or Hermitian positive definite) then all the λ_j's are real and positive. Then, $X'AX$ and X^*AX are strictly positive.

Let A and B be $n \times n$ matrices. If $AB = BA$, in which case we say that A and B commute, then both A and B can be simultaneously reduced to their canonical forms (diagonal forms with the diagonal elements being the eigenvalues) with the same orthonormal or unitary matrix P, $PP' = I$, $P'P = I$ if P is real and $PP^* = I$, $P^*P = I$ if complex, such that $P'AP = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $P'BP = \text{diag}(\mu_1, \ldots, \mu_n)$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A and $\mu_1, \ldots, \mu_n$ are the eigenvalues of B. In the complex case, $P^*AP = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $P^*BP = \text{diag}(\mu_1, \ldots, \mu_n)$. Observe that the eigenvalues of Hermitian matrices are real.

1.5.1. Singular value decomposition

For an $n \times n$ symmetric matrix $A = A'$, we have stated that there exists an $n \times n$ orthonormal matrix P, $PP' = I_n$, $P'P = I_n$ such that $P'AP = D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. If a square matrix A is not symmetric, there exists a nonsingular matrix Q such that $Q^{-1}AQ = D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ when the rank of A is n. If the rank is less than n, then we may be able to obtain a nonsingular Q such that the above representation holds; however, this is not always possible. If A is a $p \times q$ rectangular matrix for $p \neq q$ or if $p = q$ and $A \neq A'$, then can we find two orthonormal matrices U and V such that $A = U \begin{bmatrix} \Lambda & O \\ O & O \end{bmatrix} V'$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$, $UU' = I_p$, $U'U = I_p$, $VV' = I_q$, $V'V = I_q$ and k is the rank of A. This representation is equivalent to the following:

$$A = U \begin{bmatrix} \Lambda & O \\ O & O \end{bmatrix} V' = U_{(1)} \Lambda V'_{(1)} \qquad (i)$$

where $U_{(1)} = [U_1, \ldots, U_k]$, $V_{(1)} = [V_1, \ldots, V_k]$, U_j being the normalized eigenvector of AA' corresponding to the eigenvalue λ_j^2 and V_j, the normalized eigenvector of $A'A$ corresponding to the eigenvalue λ_j^2. The representation given in *(i)* is known as the *singular value decomposition of* A and $\lambda_1 > 0, \ldots, \lambda_k > 0$ are called the *singular values of* A. Then, we have

$$AA' = U \begin{bmatrix} \Lambda^2 & O \\ O & O \end{bmatrix} U' = U_{(1)} \Lambda^2 U'_{(1)}, \;\; A'A = V \begin{bmatrix} \Lambda^2 & O \\ O & O \end{bmatrix} V' = V_{(1)} \Lambda^2 V'_{(1)}. \qquad (ii)$$

Thus, the procedure is the following: If $p \leq q$, compute the nonzero eigenvalues of AA', otherwise compute the nonzero eigenvalues of $A'A$. Denote them by $\lambda_1^2, \ldots, \lambda_k^2$ where k is the rank of A. Construct the following normalized eigenvectors $U_1, \ldots, U_k$ from AA'. This

gives $U_{(1)} = [U_1, \ldots, U_k]$. Then, by using the same eigenvalues λ_j^2, $j = 1, \ldots, k$, determine the normalized eigenvectors, $V_1, \ldots, V_k$, from $A'A$, and let $V_{(1)} = [V_1, \ldots, V_k]$. Let us verify the above statements with the help of an example. Let

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Then,

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

The eigenvalues of AA' are $\lambda_1^2 = 3$ and $\lambda_2^2 = 2$. The corresponding normalized eigenvectors of AA' are U_1 and U_2, where

$$U_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ U_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \text{ so that } U_{(1)} = [U_1, U_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now, by using $\lambda_1^2 = 3$ and $\lambda_2^2 = 2$, compute the normalized eigenvectors from $A'A$. They are:

$$V_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \ V_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ so that } V_{(1)}' = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

Then $\Lambda = \text{diag}(\sqrt{3}, \sqrt{2})$. Also,

$$U_{(1)} \Lambda V_{(1)}' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = A.$$

This establishes the result. Observe that

$$AA' = [U_{(1)} \Lambda V_{(1)}'][U_{(1)} \Lambda V_{(1)}']' = U_{(1)} \Lambda^2 U_{(1)}'$$
$$A'A = [U_{(1)} \Lambda V_{(1)}']'[U_{(1)} \Lambda V_{(1)}'] = V_{(1)} \Lambda^2 V_{(1)}'$$
$$\Lambda^2 = \text{diag}(\lambda_1^2, \lambda_2^2).$$

1.6. Wedge Product of Differentials and Jacobians

If $y = f(x)$ is an explicit function of x, where x and y are real scalar variables, then we refer to x as the independent variable and to y as the dependent variable. In the present context, "independent" means that values for x are preassigned and the corresponding values of y are evaluated from the formula $y = f(x)$. The standard notations for small increment in x and the corresponding increment in y are Δx and Δy, respectively. By

convention, $\Delta x > 0$ and Δy can be positive, negative or zero depending upon the function f. If Δx goes to zero, then the limit is zero. However, if Δx goes to zero in the presence of the ratio $\frac{\Delta y}{\Delta x}$, then we have a different situation. Consider the identity

$$\Delta y \equiv \left(\frac{\Delta y}{\Delta x}\right)\Delta x \ \Rightarrow \ dy = A\,dx, \ A = \frac{dy}{dx}. \tag{1.6.1}$$

This identity can always be written due to our convention $\Delta x > 0$. Consider $\Delta x \to 0$. If $\frac{\Delta y}{\Delta x}$ attains a limit at some stage as $\Delta x \to 0$, let us denote it by $A = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$, then the value of Δx at that stage is the differential of x, namely dx, and the corresponding Δy is dy and A is the ratio of the differentials $A = \frac{dy}{dx}$. If $x_1, \dots, x_k$ are independent variables and if $y = f(x_1, \dots, x_k)$, then by convention $\Delta x_1 > 0, \dots, \Delta x_k > 0$. Thus, in light of (1.6.1), we have

$$dy = \frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_k}dx_k \tag{1.6.2}$$

where $\frac{\partial f}{\partial x_j}$ is the partial derivative of f with respect to x_j or the derivative of f with respect to x_j, keeping all other variables fixed.

Wedge Product of Differentials Let dx and dy be differentials of the real scalar variables x and y. Then the wedge product or skew symmetric product of dx and dy is denoted by $dx \wedge dy$ and is defined as

$$dx \wedge dy = -dy \wedge dx \ \Rightarrow \ dx \wedge dx = 0 \ \text{and} \ dy \wedge dy = 0. \tag{1.6.3}$$

This definition indicates that higher order wedge products involving the same differential are equal to zero. Letting

$$y_1 = f_1(x_1, x_2) \ \text{and} \ y_2 = f_2(x_1, x_2),$$

it follows from the basic definitions that

$$dy_1 = \frac{\partial f_1}{\partial x_1}dx_1 + \frac{\partial f_1}{\partial x_2}dx_2 \ \text{and} \ dy_2 = \frac{\partial f_2}{\partial x_1}dx_1 + \frac{\partial f_2}{\partial x_2}dx_2.$$

By taking the wedge product and using the properties specified in (1.6.3), that is, $dx_1 \wedge dx_1 = 0$, $dx_2 \wedge dx_2 = 0$, $dx_2 \wedge dx_1 = -dx_1 \wedge dx_2$, we have

$$\begin{aligned}
dy_1 \wedge dy_2 &= \left[\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_1}\right]dx_1 \wedge dx_2 \\
&= \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{vmatrix} dx_1 \wedge dx_2 \,.
\end{aligned}$$

In the general case we have the following corresponding result:

$$dy_1 \wedge \ldots \wedge dy_k = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_k} \end{vmatrix} dx_1 \wedge \ldots \wedge dx_k \tag{1.6.4}$$

$$dY = J \, dX \Rightarrow dX = \frac{1}{J} \, dY \text{ if } J \neq 0$$

where $dX = dx_1 \wedge \ldots \wedge dx_k$, $dY = dy_1 \wedge \ldots \wedge dy_k$ and $J = |(\frac{\partial f_i}{\partial x_j})| = $ the determinant of the matrix of partial derivatives where the (i, j)-th element is the partial derivative of f_i with respect to x_j. In dX and dY, the individual real scalar variables can be taken in any order to start with. However, for each interchange of variables, the result is to be multiplied by -1.

Example 1.6.1. Consider the transformation $x_1 = r \cos^2 \theta$, $x_2 = r \sin^2 \theta$, $0 \leq r < \infty$, $0 \leq \theta \leq \frac{\pi}{2}$, $x_1 \geq 0, x_2 \geq 0$. Determine the relationship between $dx_1 \wedge dx_2$ and $dr \wedge d\theta$.

Solution 1.6.1. Taking partial derivatives, we have

$$\frac{\partial x_1}{\partial r} = \cos^2 \theta, \ \frac{\partial x_1}{\partial \theta} = -2r \cos \theta \sin \theta,$$

$$\frac{\partial x_2}{\partial r} = \sin^2 \theta, \ \frac{\partial x_2}{\partial \theta} = 2r \cos \theta \sin \theta.$$

Then, the determinant of the matrix of partial derivatives is given by

$$\begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos^2 \theta & -2r \cos \theta \sin \theta \\ \sin^2 \theta & 2r \cos \theta \sin \theta \end{vmatrix} = 2r \cos \theta \sin \theta$$

since $\cos^2 \theta + \sin^2 \theta = 1$. Hence,

$$dx_1 \wedge dx_2 = 2r \cos \theta \sin \theta \, dr \wedge d\theta, \ J = 2r \cos \theta \sin \theta.$$

We may also establish this result by direct evaluation.

$$dx_1 = \frac{\partial x_1}{\partial r} dr + \frac{\partial x_1}{\partial \theta} d\theta = \cos^2 \theta \, dr - 2r \cos \theta \sin \theta \, d\theta,$$

$$dx_2 = \frac{\partial x_2}{\partial r} dr + \frac{\partial x_2}{\partial \theta} d\theta = \sin^2 \theta \, dr + 2r \cos \theta \sin \theta \, d\theta,$$

$$dx_1 \wedge dx_2 = \cos^2\theta \sin^2\theta \, dr \wedge dr + \cos^2\theta (2r\cos\theta\sin\theta) \, dr \wedge d\theta$$
$$- \sin^2\theta (2r\cos\theta\sin\theta) \, d\theta \wedge dr - (2r\cos\theta\sin\theta)^2 d\theta \wedge d\theta$$
$$= 2r\cos\theta\sin\theta \, [\cos^2\theta \, dr \wedge d\theta - \sin^2\theta \, d\theta \wedge dr], \ [dr \wedge dr = 0, \ d\theta \wedge d\theta = 0]$$
$$= 2r\cos\theta\sin\theta \, [\cos^2\theta + \sin^2\theta] \, dr \wedge d\theta, \ [d\theta \wedge dr = -dr \wedge d\theta]$$
$$= 2r\cos\theta\sin\theta \, dr \wedge d\theta.$$

Linear Transformation Consider the linear transformation $Y = AX$ where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & \ldots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \ldots & a_{pp} \end{bmatrix}.$$

Then, $\frac{\partial y_i}{\partial x_j} = a_{ij} \Rightarrow (\frac{\partial y_i}{\partial x_j}) = (a_{ij}) = A$. Then $dY = |A| \, dX$ or $J = |A|$. Hence, the following result:

Theorem 1.6.1. *Let X and Y be $p \times 1$ vectors of distinct real variables and $A = (a_{ij})$ be a constant nonsingular matrix. Then, the transformation $Y = AX$ is one to one and*

$$Y = AX, \ |A| \neq 0 \ \Rightarrow \ dY = |A| \, dX. \tag{1.6.5}$$

Let us consider the complex case. Let $\tilde{X} = X_1 + iX_2$ where a tilde indicates that the matrix is in the complex domain, X_1 and X_2 are real $p \times 1$ vectors if $\tilde{X}$ is $p \times 1$, and $i = \sqrt{(-1)}$. Then, the wedge product $d\tilde{X}$ is defined as $d\tilde{X} = dX_1 \wedge dX_2$. This is the general definition in the complex case whatever be the order of the matrix. If $\tilde{Z}$ is $m \times n$ and if $\tilde{Z} = Z_1 + iZ_2$ where Z_1 and Z_2 are $m \times n$ and real, then $d\tilde{Z} = dZ_1 \wedge dZ_2$. Letting the constant $p \times p$ matrix $A = A_1 + iA_2$ where A_1 and A_2 are real and $p \times p$, and letting $\tilde{Y} = Y_1 + iY_2$ be $p \times 1$ where Y_1 and Y_2 are real and $p \times 1$, we have

$$\tilde{Y} = A\tilde{X} \Rightarrow Y_1 + iY_2 = [A_1 + iA_2][X_1 + iX_2]$$
$$= [A_1 X_1 - A_2 X_2] + i[A_1 X_2 + A_2 X_1] \Rightarrow$$
$$Y_1 = A_1 X_1 - A_2 X_2, \ Y_2 = A_1 X_2 + A_2 X_1 \Rightarrow$$
$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}. \tag{i}$$

Now, applying Result 1.6.1 on *(i)*, it follows that

$$dY_1 \wedge dY_2 = \det \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} dX_1 \wedge dX_2. \tag{ii}$$

That is,

$$d\tilde{Y} = \det\begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} d\tilde{X} \implies d\tilde{Y} = J \, d\tilde{X} \qquad (iii)$$

where the Jacobian can be shown to be the absolute value of the determinant of A. If the determinant of A is denoted by $\det(A)$ and its absolute value, by $|\det(A)|$, and if $\det(A) = a + ib$ with a, b real and $i = \sqrt{(-1)}$ then the absolute value of the determinant is $+\sqrt{(a + ib)(a - ib)} = +\sqrt{(a^2 + b^2)} = +\sqrt{[\det(A)][\det(A^*)]} = +\sqrt{[\det(AA^*)]}$. It can be easily seen that the above Jacobian is given by

$$J = \det\begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} = \det\begin{bmatrix} A_1 & -iA_2 \\ -iA_2 & A_1 \end{bmatrix}$$

(multiplying the second row block by $-i$ and second column block by i)

$$= \det\begin{bmatrix} A_1 - iA_2 & A_1 - iA_2 \\ -iA_2 & A_1 \end{bmatrix} \text{ (adding the second row block to the first row block)}$$

$$= \det(A_1 - iA_2) \det\begin{bmatrix} I & I \\ -iA_2 & A_1 \end{bmatrix} = \det(A_1 - iA_2)\det(A_1 + iA_2)$$

(adding (-1) times the first p columns to the last p columns)

$$= [\det(A)][\det(A^*)] = [\det(AA^*)] = |\det(A)|^2.$$

Then, we have the following companion result of Theorem 1.6.1.

Theorem 1.6a.1. *Let $\tilde{X}$ and $\tilde{Y}$ be $p \times 1$ vectors in the complex domain, and let A be a $p \times p$ nonsingular constant matrix that may or may not be in the complex domain. If C is a constant $p \times 1$ vector, then*

$$\tilde{Y} = A\tilde{X} + C, \ \det(A) \neq 0 \implies d\tilde{Y} = |\det(A)|^2 d\tilde{X} = |\det(AA^*)| \, d\tilde{X}. \qquad (1.6a.1)$$

For the results that follow, the complex case can be handled in a similar way and hence, only the final results will be stated. For details, the reader may refer to Mathai (1997). A more general result is the following:

Theorem 1.6.2. *Let X and Y be real $m \times n$ matrices with distinct real variables as elements. Let A be a $m \times m$ nonsingular constant matrix and C be a $m \times n$ constant matrix. Then*

$$Y = AX + C, \ \det(A) \neq 0 \implies dY = |A|^n dX. \qquad (1.6.6)$$

The companion result is stated in the next theorem.

Theorem 1.6a.2. *Let $\tilde{X}$ and $\tilde{Y}$ be $m \times n$ matrices in the complex domain. Let A be a constant $m \times m$ nonsingular matrix that may or may not be in the complex domain, and C be a $m \times n$ constant matrix. Then*

$$\tilde{Y} = A\tilde{X} + C, \ \det(A) \neq 0 \Rightarrow d\tilde{Y} = |\det(AA^*)|^n d\tilde{X}. \qquad (1.6a.2)$$

For proving the Theorems 1.6.2 and 1.6a.2, consider the columns of Y and X. Then apply Theorems 1.6.1 and 1.6a.1 to establish the results. If X, $\tilde{X}$, Y, $\tilde{Y}$ are as defined in Theorems 1.6.2 and 1.6a.2 and if B is a $n \times n$ nonsingular constant matrix, then we have the following results:

Theorems 1.6.3 and 1.6a.3. *Let X, $\tilde{X}$, Y, $\tilde{Y}$ and C be $m \times n$ matrices with distinct elements as previously defined, C be a constant matrix and B be a $n \times n$ nonsingular constant matrix. Then*

$$Y = XB + C, \ \det(B) \neq 0 \Rightarrow dY = |B|^m dX \qquad (1.6.7)$$

and

$$\tilde{Y} = \tilde{X}B + C, \ \det(B) \neq 0 \Rightarrow d\tilde{Y} = |\det(BB^*)|^m d\tilde{X}. \qquad (1.6a.3)$$

For proving these results, consider the rows of Y, $\tilde{Y}$, X, $\tilde{X}$ and then apply Theorems 1.6.1, 1.6a.1 to establish the results. Combining Theorems 1.6.2 and 1.6.3, as well as Theorems 1.6a.2 and 1.6a.3, we have the following results:

Theorems 1.6.4 and 1.6a.4. *Let X, $\tilde{X}$, Y, $\tilde{Y}$ be $m \times n$ matrices as previously defined, and let A be $m \times m$ and B be $n \times n$ nonsingular constant matrices. Then*

$$Y = AXB, \ \det(A) \neq 0, \ \det(B) \neq 0 \Rightarrow dY = |A|^n |B|^m dX \qquad (1.6.8)$$

and

$$\tilde{Y} = A\tilde{X}B, \ \det(A) \neq 0, \ \det(B) \neq 0 \Rightarrow d\tilde{Y} = |\det(AA^*)|^n |\det(BB^*)|^m d\tilde{X}. \qquad (1.6a.4)$$

We now consider the case of linear transformations involving symmetric and Hermitian matrices.

Theorems 1.6.5 and 1.6a.5. *Let $X = X'$, $Y = Y'$ be real symmetric $p \times p$ matrices and let $\tilde{X} = \tilde{X}^*$, $\tilde{Y} = \tilde{Y}^*$ be $p \times p$ Hermitian matrices. If A is a $p \times p$ nonsingular constant matrix, then*

$$Y = AXA', \ Y = Y', \ X = X', \ \det(A) \neq 0 \Rightarrow dY = |A|^{p+1} dX \qquad (1.6.9)$$

and

$$\tilde{Y} = A\tilde{X}A^*, \ \det(A) \neq 0 \Rightarrow d\tilde{Y} = |\det(AA^*)|^p d\tilde{X} \qquad (1.6a.5)$$

for $\tilde{X} = \tilde{X}^*$ or $\tilde{X} = -\tilde{X}^*$.

The proof involves some properties of elementary matrices and elementary transformations. Elementary matrices were introduced in Sect. 1.2.1. There are two types of basic elementary matrices, the E and F types where the E type is obtained by multiplying any row (column) of an identity matrix by a nonzero scalar and the F type is obtained by adding any row to any other row of an identity matrix. A combination of E and F type matrices results in a G type matrix where a constant multiple of one row of an identity matrix is added to any other row. The G type is not a basic elementary matrix. By performing successive pre-multiplication with E, F and G type matrices, one can reduce a nonsingular matrix to a product of the basic elementary matrices of the E and F types, observing that the E and F type elementary matrices are nonsingular. This result is needed to establish Theorems 1.6.5 and 1.6a.5. Let $A = E_1 E_2 F_1 \cdots E_r F_s$ for some $E_1, \ldots, E_r$ and $F_1, \ldots, F_s$. Then

$$AXA' = E_1 E_2 F_1 \cdots E_r F_s X F_s' E_r' \cdots E_2' E_1'.$$

Let $Y_1 = F_s X F_s'$ in which case the connection between dX and dY_1 can be determined from F_s. Now, letting $Y_2 = E_r Y_1 E_r'$, the connection between dY_2 and dY_1 can be similarly determined from E_r. Continuing in this manner, we finally obtain the connection between dY and dX, which will give the Jacobian as $|A|^{p+1}$ for the real case. In the complex case, the procedure is parallel.

We now consider two basic nonlinear transformations. In the first case, X is a $p \times p$ nonsingular matrix going to its inverse, that is, $Y = X^{-1}$.

Theorems 1.6.6 and 1.6a.6. *Let X and $\tilde{X}$ be $p \times p$ real and complex nonsingular matrices, respectively. Let the regular inverses be denoted by $Y = X^{-1}$ and $\tilde{Y} = \tilde{X}^{-1}$, respectively. Then, ignoring the sign,*

$$Y = X^{-1} \Rightarrow dY = \begin{cases} |X|^{-2p} dX \text{ for a general } X \\ |X|^{-(p+1)} dX \text{ for } X = X' \\ |X|^{-(p-1)} dX \text{ for } X = -X' \end{cases} \tag{1.6.10}$$

and

$$\tilde{Y} = \tilde{X}^{-1} \Rightarrow d\tilde{Y} = \begin{cases} |\det(\tilde{X}\tilde{X}^*)|^{-2p} \text{ for a general } \tilde{X} \\ |\det(\tilde{X}\tilde{X}^*)|^{-p} \text{ for } \tilde{X} = \tilde{X}^* \text{ or } \tilde{X} = -\tilde{X}^*. \end{cases} \tag{1.6a.6}$$

The proof is based on the following observations: In the real case $XX^{-1} = I_p \Rightarrow (dX)X^{-1} + X(dX^{-1}) = O$ where (dX) represents the matrix of differentials in X. This means that

$$(dX^{-1}) = -X^{-1}(dX)X^{-1}.$$

The differentials are appearing only in the matrices of differentials. Hence this situation is equivalent to the general linear transformation considered in Theorems 1.6.4 and 1.6.5 where X and X^{-1} act as constants. The result is obtained upon taking the wedge product of differentials. The complex case is parallel.

The next results involve real positive definite matrices or Hermitian positive definite matrices that are expressible in terms of triangular matrices and the corresponding connection between the wedge product of differentials. Let X and $\tilde{X}$ be complex $p \times p$ real positive definite and Hermitian positive definite matrices, respectively. Let $T = (t_{ij})$ be a real lower triangular matrix with $t_{ij} = 0$, $i < j$, $t_{jj} > 0$, $j = 1, \ldots, p$, and the t_{ij}'s, $i \geq j$, be distinct real variables. Let $\tilde{T} = (\tilde{t}_{ij})$ be a lower triangular matrix with $\tilde{t}_{ij} = 0$, for $i < j$, the $\tilde{t}_{ij}$'s, $i > j$, be distinct complex variables, and $\tilde{t}_{jj}$, $j = 1, \ldots, p$, be positive real variables. Then, the transformations $X = TT'$ in the real case and $\tilde{X} = \tilde{T}\tilde{T}^*$ in the complex case can be shown to be one-to-one, which enables us to write dX in terms of dT and vice versa, uniquely, and $d\tilde{X}$ in terms of $d\tilde{T}$, uniquely. We first consider the real case. When $p = 2$,

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix}, \ x_{11} > 0, \ x_{22} > 0, \ x_{21} = x_{12}, \ x_{11}x_{22} - x_{12}^2 > 0$$

due to positive definiteness of X, and

$$X = TT' = \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{bmatrix} = \begin{bmatrix} t_{11}^2 & t_{21}t_{11} \\ t_{21}t_{11} & t_{21}^2 + t_{22}^2 \end{bmatrix} \Rightarrow$$

$$\frac{\partial x_{11}}{\partial t_{11}} = 2t_{11}, \ \frac{\partial x_{11}}{\partial t_{21}} = 0, \ \frac{\partial x_{11}}{\partial t_{22}} = 0$$

$$\frac{\partial x_{22}}{\partial t_{11}} = 0, \ \frac{\partial x_{22}}{\partial t_{21}} = 2t_{21}, \ \frac{\partial x_{22}}{\partial t_{22}} = 2t_{22}$$

$$\frac{\partial x_{12}}{\partial t_{11}} = t_{21}, \ \frac{\partial x_{12}}{\partial t_{21}} = t_{11}, \ \frac{\partial x_{12}}{\partial t_{22}} = 0.$$

Taking the x_{ij}'s in the order x_{11}, x_{12}, x_{22} and the t_{ij}'s in the order t_{11}, t_{21}, t_{22}, we form the following matrix of partial derivatives:

	t_{11}	t_{12}	t_{22}
x_{11}	$2t_{11}$	0	0
x_{21}	$*$	t_{11}	0
x_{22}	$*$	$*$	$2t_{22}$

where an asterisk indicates that an element may be present in that position; however, its value is irrelevant since the matrix is triangular and its determinant will simply be the

product of its diagonal elements. It can be observed from this pattern that for a general p, a diagonal element will be multiplied by 2 whenever x_{jj} is differentiated with respect to t_{jj}, $j = 1, \ldots, p$. Then t_{11} will appear p times, t_{22} will appear $p - 1$ times, and so on, and t_{pp} will appear once along the diagonal. Hence the product of the diagonal elements will be $2^p t_{11}^p t_{22}^{p-1} \cdots t_{pp} = 2^p \{\prod_{j=1}^p t_{jj}^{p+1-j}\}$. A parallel procedure will yield the Jacobian in the complex case. Hence, the following results:

Theorems 1.6.7 and 1.6a.7. *Let X, $\tilde{X}$, T and $\tilde{T}$ be $p \times p$ matrices where X is real positive definite, $\tilde{X}$ is Hermitian positive definite, and T and $\tilde{T}$ are lower triangular matrices whose diagonal elements are real and positive as described above. Then the transformations $X = TT'$ and $\tilde{X} = \tilde{T}\tilde{T}^*$ are one-to-one, and*

$$dX = 2^p \{\prod_{j=1}^p t_{jj}^{p+1-j}\} \, dT \qquad (1.6.11)$$

and

$$d\tilde{X} = 2^p \{\prod_{j=1}^p t_{jj}^{2(p-j)+1}\} \, d\tilde{T}. \qquad (1.6a.7)$$

Given these introductory materials, we will explore multivariate statistical analysis from the perspective of Special Functions. As far as possible, the material in this chapter is self-contained. A few more Jacobians will be required when tackling transformations involving rectangular matrices or eigenvalue problems. These will be discussed in the respective chapters later on.

Example 1.6.2. Evaluate the following integrals: (1): $\int_X e^{-X'AX} dX$ where $A > O$ (real positive definite) is 3×3 and X is a 3×1 vector of distinct real scalar variables; (2): $\int_X e^{-\text{tr}(AXBX')} dX$ where X is a 2×3 matrix of distinct real scalar variables, $A > O$ (real positive definite), is 2×2 and $B > O$ (real positive definite) is 3×3, A and B being constant matrices; (3): $\int_{X>O} e^{-\text{tr}(X)} dX$ where $X = X' > O$ is a 2×2 real positive definite matrix of distinct real scalar variables.

Solution 1.6.2.
(1) Let $X' = (x_1, x_2, x_3)$, $A > O$. Since $A > O$, we can uniquely define $A^{\frac{1}{2}} = (A^{\frac{1}{2}})'$. Then, write $X'AX = X'A^{\frac{1}{2}}A^{\frac{1}{2}}X = Y'Y$, $Y = A^{\frac{1}{2}}X$. It follows from Theorem 1.6.1 that $dX = |A|^{-\frac{1}{2}} dY$, and letting $Y' = (y_1, y_2, y_3)$, we have

$$\int_X e^{-X'AX} dX = |A|^{-\frac{1}{2}} \int_Y e^{-(Y'Y)} dY$$

$$= |A|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y_1^2 + y_2^2 + y_3^2)} dy_1 \wedge dy_2 \wedge dy_3.$$

Since

$$\int_{-\infty}^{\infty} e^{-y_j^2} dy_j = \sqrt{\pi}, \ j = 1, 2, 3,$$

$$\int_X e^{-X'AX} dX = |A|^{-\frac{1}{2}}(\sqrt{\pi})^3.$$

(2) Since A is a 2×2 positive definite matrix, there exists a 2×2 matrix $A^{\frac{1}{2}}$ that is symmetric and positive definite. Similarly, there exists a 3×3 matrix $B^{\frac{1}{2}}$ that is symmetric and positive definite. Let $Y = A^{\frac{1}{2}} X B^{\frac{1}{2}} \Rightarrow dY = |A|^{\frac{3}{2}} |B|^{\frac{2}{2}} dX$ or $dX = |A|^{-\frac{3}{2}} |B|^{-1} dY$ by Theorem 1.6.4. Moreover, given two matrices A_1 and A_2, $\text{tr}(A_1 A_2) = \text{tr}(A_2 A_1)$ even if $A_1 A_2 \neq A_2 A_1$, as long as the products are defined. By making use of this property, we may write

$$\text{tr}(AXBX') = \text{tr}(A^{\frac{1}{2}} A^{\frac{1}{2}} X B^{\frac{1}{2}} B^{\frac{1}{2}} X') = \text{tr}(A^{\frac{1}{2}} X B^{\frac{1}{2}} B^{\frac{1}{2}} X' A^{\frac{1}{2}})$$

$$= \text{tr}[(A^{\frac{1}{2}} X B^{\frac{1}{2}})(A^{\frac{1}{2}} X B^{\frac{1}{2}})'] = \text{tr}(YY')$$

where $Y = A^{\frac{1}{2}} X B^{\frac{1}{2}}$ and dY is given above. However, for any real matrix Y, whether square or rectangular, $\text{tr}(YY') = \text{tr}(Y'Y) =$ the sum of the squares of all the elements of Y. Thus, we have

$$\int_X e^{-\text{tr}(AXBX')} dX = |A|^{-\frac{3}{2}} |B|^{-1} \int_Y e^{-\text{tr}(YY')} dY.$$

Observe that since $\text{tr}(YY')$ is the sum of squares of 6 real scalar variables, the integral over Y reduces to a multiple integral involving six integrals where each variable is over the entire real line. Hence,

$$\int_Y e^{-\text{tr}(YY')} dY = \prod_{j=1}^{6} \int_{-\infty}^{\infty} e^{-y_j^2} dy_j = \prod_{j=1}^{6} (\sqrt{\pi}) = (\sqrt{\pi})^6.$$

Note that we have denoted the sum of the six y_{ij}^2 as $y_1^2 + \cdots + y_6^2$ for convenience. Thus,

$$\int_X e^{-\text{tr}(AXBX')} dX = |A|^{-\frac{3}{2}} |B|^{-1} (\sqrt{\pi})^6.$$

(3) In this case, X is a 2×2 real positive definite matrix. Let $X = TT'$ where T is lower triangular with positive diagonal elements. Then,

$$T = \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix}, \ t_{11} > 0, \ t_{22} > 0, \ TT' = \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{bmatrix} = \begin{bmatrix} t_{11}^2 & t_{11} t_{21} \\ t_{11} t_{21} & t_{21}^2 + t_{22}^2 \end{bmatrix},$$

and $\mathrm{tr}(TT') = t_{11}^2 + (t_{21}^2 + t_{22}^2)$, $t_{11} > 0$, $t_{22} > 0$, $-\infty < t_{21} < \infty$. From Theorem 1.6.7, the Jacobian is

$$\mathrm{d}X = 2^p\{\prod_{j=1}^{p} t_{jj}^{p+1-j}\}\mathrm{d}T = 2^2(t_{11}^2 t_{22})\,\mathrm{d}t_{11} \wedge \mathrm{d}t_{21} \wedge \mathrm{d}t_{22}.$$

Therefore

$$\int_{X>O} \mathrm{e}^{-\mathrm{tr}(X)}\mathrm{d}X = \int_T \mathrm{e}^{-\mathrm{tr}(TT')}[2^2(t_{11}^2 t_{22})]\mathrm{d}T$$

$$= \Big(\int_{-\infty}^{\infty} \mathrm{e}^{-t_{21}^2}\mathrm{d}t_{21}\Big)\Big(\int_0^{\infty} 2t_{11}^2 \mathrm{e}^{-t_{11}^2}\mathrm{d}t_{11}\Big)\Big(\int_0^{\infty} 2t_{22}\mathrm{e}^{-t_{22}^2}\mathrm{d}t_{22}\Big)$$

$$= [\sqrt{\pi}]\,[\Gamma(\frac{3}{2}))]\,[\Gamma(1)] = \frac{\pi}{2}.$$

Example 1.6.3. Let $A = A^* > O$ be a constant 2×2 Hermitian positive definite matrix. Let $\tilde{X}$ be a 2×1 vector in the complex domain and $\tilde{X}_2 > O$ be a 2×2 Hermitian positive definite matrix. Then, evaluate the following integrals: (1): $\int_{\tilde{X}} \mathrm{e}^{-(\tilde{X}^* A \tilde{X})}\mathrm{d}\tilde{X}$; (2): $\int_{\tilde{X}_2>O} \mathrm{e}^{-\mathrm{tr}(\tilde{X}_2)}\mathrm{d}\tilde{X}_2$.

Solution 1.6.3.
(1): Since $A = A^* > O$, there exists a unique Hermitian positive definite square root $A^{\frac{1}{2}}$. Then,

$$\tilde{X}^* A \tilde{X} = \tilde{X}^* A^{\frac{1}{2}} A^{\frac{1}{2}} \tilde{X} = \tilde{Y}^* \tilde{Y},$$

$$\tilde{Y} = A^{\frac{1}{2}} \tilde{X} \Rightarrow \mathrm{d}\tilde{X} = |\det(A)|^{-1}\mathrm{d}\tilde{Y}$$

by Theorem 1.6a.1. But $\tilde{Y}^* \tilde{Y} = |\tilde{y}_1|^2 + |\tilde{y}_2|^2$ since $\tilde{Y}^* = (\tilde{y}_1^*, \tilde{y}_2^*)$. Since the $\tilde{y}_j$'s are scalar in this case, an asterisk means only the complex conjugate, the transpose being itself. However,

$$\int_{\tilde{y}_j} \mathrm{e}^{-|\tilde{y}_j|^2}\mathrm{d}\tilde{y}_j = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-(y_{j1}^2+y_{j2}^2)}\mathrm{d}y_{j1} \wedge \mathrm{d}y_{j2} = (\sqrt{\pi})^2 = \pi$$

where $\tilde{y}_j = y_{j1} + iy_{j2}$, $i = \sqrt{(-1)}$, y_{j1} and y_{j2} being real. Hence

$$\int_{\tilde{X}} \mathrm{e}^{-(\tilde{X}^* A \tilde{X})}\mathrm{d}\tilde{X} = |\det(A)|^{-1}\int_{\tilde{Y}} \mathrm{e}^{-(|\tilde{y}_1|^2+|\tilde{y}_2|^2)}\mathrm{d}\tilde{y}_1 \wedge \mathrm{d}\tilde{y}_2$$

$$= |\det(A)|^{-1}\Big(\prod_{j=1}^{2}\int_{\tilde{Y}_j} \mathrm{e}^{-|\tilde{y}_j|^2}\mathrm{d}\tilde{y}_j\Big) = |\det(A)|^{-1}\prod_{j=1}^{2}\pi$$

$$= |\det(A)|^{-1}\pi^2.$$

(2): Make the transformation $\tilde{X}_2 = \tilde{T}\tilde{T}^*$ where $\tilde{T}$ is lower triangular with its diagonal elements being real and positive. That is,

$$\tilde{T} = \begin{bmatrix} t_{11} & 0 \\ \tilde{t}_{21} & t_{22} \end{bmatrix}, \ \tilde{T}\tilde{T}^* = \begin{bmatrix} t_{11}^2 & t_{11}\tilde{t}_{21} \\ t_{11}\tilde{t}_{21} & |\tilde{t}_{21}|^2 + t_{22}^2 \end{bmatrix}$$

and the Jacobian is $d\tilde{X}_2 = 2^p\{\prod_{j=1}^{p} t_{jj}^{2(p-j)+1}\}d\tilde{T} = 2^2 t_{11}^3 t_{22}\, d\tilde{T}$ by Theorem 1.6a.7. Hence,

$$\int_{\tilde{X}_2 > O} e^{-\text{tr}(\tilde{X}_2)} d\tilde{X}_2 = \int_{\tilde{T}} 2^2 t_{11}^3 t_{22} e^{-(t_{11}^2 + t_{22}^2 + |\tilde{t}_{21}|^2)} dt_{11} \wedge dt_{22} \wedge d\tilde{t}_{21}.$$

But

$$2\int_{t_{11} > 0} t_{11}^3 e^{-t_{11}^2}\, dt_{11} = \int_{u=0}^{\infty} u\, e^{-u} du = 1,$$

$$2\int_{t_{22} > 0} t_{22} e^{-t_{22}^2} dt_{22} = \int_{v=0}^{\infty} e^{-v} dv = 1,$$

$$\int_{\tilde{t}_{21}} e^{-|\tilde{t}_{21}|^2} d\tilde{t}_{21} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(t_{211}^2 + t_{212}^2)} dt_{211} \wedge dt_{212}$$

$$= \left(\int_{-\infty}^{\infty} e^{-t_{211}^2} dt_{211}\right)\left(\int_{-\infty}^{\infty} e^{-t_{212}^2} dt_{212}\right)$$

$$= \sqrt{\pi}\sqrt{\pi} = \pi,$$

where $\tilde{t}_{21} = t_{211} + i t_{212}$, $i = \sqrt{(-1)}$ and t_{211}, t_{212} real. Thus

$$\int_{\tilde{X}_2 > O} e^{-\text{tr}(\tilde{X}_2)} d\tilde{X}_2 = \pi.$$

1.7. Differential Operators

Let

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \ \frac{\partial}{\partial X} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_p} \end{bmatrix}, \ \frac{\partial}{\partial X'} = \left[\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_p}\right]$$

where $x_1, \ldots, x_p$ are distinct real scalar variables, $\frac{\partial}{\partial X}$ is the partial differential operator and $\frac{\partial}{\partial X'}$ is the transpose operator. Then, $\frac{\partial}{\partial X} \frac{\partial}{\partial X'}$ is the configuration of all second order partial differential operators given by

$$\frac{\partial}{\partial X} \frac{\partial}{\partial X'} = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} & \frac{\partial^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2}{\partial x_1 \partial x_p} \\ \frac{\partial^2}{\partial x_2 \partial x_1} & \frac{\partial^2}{\partial x_2^2} & \cdots & \frac{\partial^2}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_p \partial x_1} & \frac{\partial^2}{\partial x_p \partial x_2} & \cdots & \frac{\partial^2}{\partial x_p^2} \end{bmatrix}.$$

Let $f(X)$ be a real-valued scalar function of X. Then, this operator, operating on f will be defined as

$$\frac{\partial}{\partial X} f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{bmatrix}.$$

For example, if $f = x_1^2 + x_1 x_2 + x_2^3$, then $\frac{\partial f}{\partial x_1} = 2x_1 + x_2$, $\frac{\partial f}{\partial x_2} = x_1 + 3x_2^2$, and

$$\frac{\partial f}{\partial X} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 3x_2^2 \end{bmatrix}.$$

Let $f = a_1 x_1 + a_2 x_2 + \cdots + a_p x_p = A'X = X'A$, $A' = (a_1, \ldots, a_p)$, $X' = (x_1, \ldots, x_p)$ where $a_1, \ldots, a_p$ are real constants and $x_1, \ldots, x_p$ are distinct real scalar variables. Then $\frac{\partial f}{\partial x_j} = a_j$ and we have the following result:

Theorem 1.7.1. *Let A, X and f be as defined above where $f = a_1 x_1 + \cdot + a_p x_p$ is a linear function of X, then*

$$\frac{\partial}{\partial X} f = A.$$

Letting $f = X'X = x_1^2 + \cdots + x_p^2$, $\frac{\partial f}{\partial x_j} = 2x_j$, and we have the following result.

Theorem 1.7.2. *Let X be a $p \times 1$ vector of real scalar variables so that $X'X = x_1^2 + \cdots + x_p^2$. Then*

$$\frac{\partial f}{\partial X} = 2X.$$

Now, let us consider a general quadratic form $f = X'AX$, $A = A'$, where X is a $p \times 1$ vector whose components are real scalar variables and A is a constant matrix. Then $\frac{\partial f}{\partial x_j} = (a_{j1}x_1 + \cdots + a_{jp}x_p) + (a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p)$ for $j = 1, \ldots, p$. Hence we have the following result:

Theorem 1.7.3. *Let $f = X'AX$ be a real quadratic form where X is a $p \times 1$ real vector whose components are distinct real scalar variables and A is a constant matrix. Then*

$$\frac{\partial f}{\partial X} = \begin{cases} (A + A')X \text{ for a general } A \\ 2AX \text{ when } A = A' \end{cases} .$$

1.7.1. Some basic applications of the vector differential operator

Let X be a $p \times 1$ vector with real scalar elements $x_1, \ldots, x_p$. Let $A = (a_{ij}) = A'$ be a constant matrix. Consider the problem of optimizing the real quadratic form $u = X'AX$. There is no unrestricted maximum or minimum. If $A = A' > O$ (positive definite), u can tend to $+\infty$ and similarly, if $A = A' < O$, u can go to $-\infty$. However, if we confine ourselves to the surface of a unit hypersphere or equivalently require that $X'X = 1$, then we can have a finite maximum and a finite minimum. Let $u_1 = X'AX - \lambda(X'X - 1)$ so that we have added zero to u and hence u_1 is the same as u, where λ is an arbitrary constant or a Lagrangian multiplier. Then, differentiating u_1 with respect to $x_1, \ldots, x_p$, equating the resulting expressions to zero, and thereafter solving for critical points, is equivalent to solving the equation $\frac{\partial u_1}{\partial X} = O$ (null) and solving this single equation. That is,

$$\frac{\partial u_1}{\partial X} = O \Rightarrow 2AX - 2\lambda X = O \Rightarrow (A - \lambda I)X = O. \qquad (i)$$

For *(i)* to have a non-null solution for X, the coefficient matrix $A - \lambda I$ has to be singular or its determinant must be zero. That is, $|A - \lambda I| = 0$ and $AX = \lambda X$ or λ is an eigenvalue of A and X is the corresponding eigenvector. But

$$AX = \lambda X \Rightarrow X'AX = \lambda X'X = \lambda \text{ since } X'X = 1. \qquad (ii)$$

Hence the maximum value of $X'AX$ corresponds to the largest eigenvalue of A and the minimum value of $X'AX$, to the smallest eigenvalue of A. Observe that when $A = A'$ the eigenvalues are real. Hence we have the following result:

Theorem 1.7.4. *Let $u = X'AX$, $A = A'$, X be a $p \times 1$ vector of real scalar variables as its elements. Letting $X'X = 1$, then*

$$\max_{X'X=1} [X'AX] = \lambda_1 = \text{ the largest eigenvalue of } A$$

$$\min_{X'X=1} [X'AX] = \lambda_p = \text{ the smallest eigenvalue of } A.$$

Principal Component Analysis where it is assumed that $A > O$ relies on this result. This will be elaborated upon in later chapters. Now, we consider the optimization of $u = X'AX$, $A = A'$ subject to the condition $X'BX = 1$, $B = B'$. Take λ as the Lagrangian multiplier and consider $u_1 = X'AX - \lambda(X'BX - 1)$. Then

$$\frac{\partial u_1}{\partial X} = O \Rightarrow AX = \lambda BX \Rightarrow |A - \lambda B| = 0. \qquad (iii)$$

Note that $X'AX = \lambda X'BX = \lambda$ from *(i)*. Hence, the maximum of $X'AX$ is the largest value of λ satisfying (i) and the minimum of $X'AX$ is the smallest value of λ satisfying (i). Note that when B is nonsingular, $|A - \lambda B| = 0 \Rightarrow |AB^{-1} - \lambda I| = 0$ or λ is an eigenvalue of AB^{-1}. Thus, this case can also be treated as an eigenvalue problem. Hence, the following result:

Theorem 1.7.5. *Let $u = X'AX$, $A = A'$ where the elements of X are distinct real scalar variables. Consider the problem of optimizing $X'AX$ subject to the condition $X'BX = 1$, $B = B'$, where A and B are constant matrices, then*

$$\max_{X'BX=1} [X'AX] = \lambda_1 = \text{ largest eigenvalue of } AB^{-1}, \ |B| \neq 0$$

$$= \text{ the largest root of } |A - \lambda B| = 0;$$

$$\min_{X'BX=1} [X'AX] = \lambda_p = \text{ smallest eigenvalue of } AB^{-1}, \ |B| \neq 0$$

$$= \text{ the smallest root of } |A - \lambda B| = 0.$$

Now, consider the optimization of a real quadratic form subject to a linear constraint. Let $u = X'AX$, $A = A'$ be a quadratic form where X is $p \times 1$. Let $B'X = X'B = 1$ be a constraint where $B' = (b_1, \ldots, b_p)$, $X' = (x_1, \ldots, x_p)$ with $b_1, \ldots, b_p$ being real constants and $x_1, \ldots, x_p$ being real distinct scalar variables. Take 2λ as the Lagrangian multiplier and consider $u_1 = X'AX - 2\lambda(X'B - 1)$. The critical points are available from the following equation

$$\frac{\partial}{\partial X} u_1 = O \Rightarrow 2AX - 2\lambda B = O \Rightarrow X = \lambda A^{-1} B, \text{ for } |A| \neq 0$$

$$\Rightarrow B'X = \lambda B'A^{-1}B \Rightarrow \lambda = \frac{1}{B'A^{-1}B}.$$

In this problem, observe that the quadratic form is unbounded even under the restriction $B'X = 1$ and hence there is no maximum. The only critical point corresponds to a minimum. From $AX = \lambda B$, we have $X'AX = \lambda X'B = \lambda$. Hence the minimum value is $\lambda = [B'A^{-1}B]^{-1}$ where it is assumed that A is nonsingular. Thus following result:

Theorem 1.7.6. *Let* $u = X'AX$, $A = A'$, $|A| \neq 0$. *Let* $B'X = 1$ *where* $B' = (b_1, \ldots, b_p)$ *be a constant vector and* X *is* $p \times 1$ *vector of real distinct scalar variables. Then, the minimum of the quadratic form* u, *under the restriction* $B'X = 1$ *where* B *is a constant vector, is given by*

$$\min_{B'X=1}[X'AX] = \frac{1}{B'A^{-1}B}.$$

Such problems arise for instance in regression analysis and model building situations. We could have eliminated one of the variables with the linear constraint; however, the optimization would still involve all other variables, and thus not much simplification would be achieved by eliminating one variable. Hence, operating with the vector differential operator is the most convenient procedure in this case.

We will now consider the mathematical part of a general problem in prediction analysis where some variables are predicted from another set of variables. This topic is related to *Canonical Correlation Analysis*. We will consider the optimization part of the problem in this section. The problem consists in optimizing a bilinear form subject to quadratic constraints. Let X be a $p \times 1$ vector of real scalar variables $x_1, \ldots, x_p$, and Y be a $q \times 1$ vector of real scalar variables $y_1, \ldots, y_q$, where q need not be equal to p. Consider the bilinear form $u = X'AY$ where A is a $p \times q$ rectangular constant matrix. We would like to optimize this bilinear form subject to the quadratic constraints $X'BX = 1$, $Y'CY = 1$, $B = B'$ and $C = C'$ where B and C are constant matrices. In Canonical Correlation Analysis, B and C are constant real positive definite matrices. Take λ_1 and λ_2 as Lagrangian multipliers and let $u_1 = X'AY - \frac{\lambda_1}{2}(X'BX - 1) - \frac{\lambda_2}{2}(Y'CY - 1)$. Then

$$\frac{\partial}{\partial X}u_1 = O \Rightarrow AY - \lambda_1 BX = O \Rightarrow AY = \lambda_1 BX$$

$$\Rightarrow X'AY = \lambda_1 X'BX = \lambda_1; \tag{i}$$

$$\frac{\partial}{\partial Y}u_1 = O \Rightarrow A'X - \lambda_2 CY = O \Rightarrow A'X = \lambda_2 CY$$

$$\Rightarrow Y'A'X = \lambda_2 Y'CY = \lambda_2. \tag{ii}$$

It follows from *(i)* and *(ii)* that $\lambda_1 = \lambda_2 = \lambda$, say. Observe that $X'AY$ is 1×1 so that $X'AY = Y'A'X$. After substituting λ to λ_1 and λ_2, we can combine equations *(i)* and *(ii)* in a single matrix equation as follows:

$$\begin{bmatrix} -\lambda B & A \\ A' & -\lambda C \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = O \Rightarrow$$

$$\begin{vmatrix} -\lambda B & A \\ A' & -\lambda C \end{vmatrix} = 0. \tag{iii}$$

Opening up the determinant by making use of a result on partitioned matrices from Sect. 1.3, we have

$$|-\lambda B|\,|-\lambda C - A'(-\lambda B)^{-1}A| = 0, \ |B| \neq 0 \Rightarrow$$
$$|A'B^{-1}A - \lambda^2 C| = 0. \tag{iv}$$

Then $\nu = \lambda^2$ is a root obtained from Eq. *(iv)*. We can also obtain a parallel result by opening up the determinant in *(iii)* as

$$|-\lambda C|\,|-\lambda B - A(-\lambda C)^{-1}A'| = 0 \Rightarrow |AC^{-1}A' - \lambda^2 B| = 0, \ |C| \neq 0. \tag{v}$$

Hence we have the following result.

Theorem 1.7.7. *Let X and Y be respectively $p \times 1$ and $q \times 1$ real vectors whose components are distinct scalar variables. Consider the bilinear form $X'AY$ and the quadratic forms $X'BX$ and $Y'CY$ where $B = B'$, $C = C'$, and B and C are nonsingular constant matrices. Then,*

$$\max_{X'BX=1,Y'CY=1}[X'AY] = |\lambda_1|$$
$$\min_{X'BX=1,Y'CY=1}[X'AY] = |\lambda_p|$$

where λ_1^2 is the largest root resulting from equation (iv) or (v) and λ_p^2 is the smallest root resulting from equation (iv) or (v).

Observe that if $p < q$, we may utilize equation *(v)* to solve for λ^2 and if $q < p$, then we may use equation *(iv)* to solve for λ^2, and both will lead to the same solution. In the above derivation, we assumed that B and C are nonsingular. In Canonical Correlation Analysis, both B and C are real positive definite matrices corresponding to the variances $X'BX$ and $Y'CY$ of the linear forms and then, $X'AY$ corresponds to covariance between these linear forms.

Note 1.7.1. We have confined ourselves to results in the real domain in this subsection since only real cases are discussed in connection with the applications that are considered in later chapters, such as Principal Component Analysis and Canonical Correlation Analysis. The corresponding complex cases do not appear to have practical applications.

Accordingly, optimizations of Hermitian forms will not be discussed. However, parallel results to Theorems 1.7.1–1.7.7 could similarly be worked out in the complex domain.

References

A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A.M. Mathai and Hans J. Haubold (2017a): *Linear Algebra: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai and Hans J. Haubold (2017b): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.