Semester-IV

B.Sc (Honours) in Physics

C8T: Mathematical Physics III

Lecture on Matrices

By

Dr. K R Sahu

Dept. of Physics, Bhatter College

Lecture-I

A brief historical introduction to matrices and their applications

1. The ancient origin of matrices

Historically, magic squares were known to antiquity in China, India and Japan, and they were commonly used to serve as mathematical art and amusement. They were very largely found in these countries, engraved on stone, metal or painting for over 4000 years. In ancient time, magic squares have been commonly used for astrological and divinatory predictions about the future, especially in making predictions about longevity and prevention of diseases. Subsequently, it has been recognized that magic squares were something more than art and amusement and worth studying from a mathematical point of view. Indeed, a magic square seemed to be a device about which more has been written than any other form of mathematical art or amusement. Considerable evidence in the history of mathematics revealed that the existence of magic squares had perhaps served as the origin of the discovery of matrices in different areas of mathematics and mathematical physics.

The famous treatise *Shushu Jiyi* (*Memoir on Some Traditions of the Mathematical Art*) by Zhen Luan, a sixth-century Chinese mathematician, contained the 3 × 3 magic squares of nine numbers from 1 to 9 organized in 3 rows and 3 columns so that the sum of each row, each column and each of the two diagonals is 15. In modern notation, this magic square represents a 3×3 square matrix of $3^2 = 9$ elements. In the tenth century, during the Song Dynasty, magic squares were commonly used in devination, predicting what will happen in the future based on the purely numerical aspects of magic squares.

In 1514, among many symbolic elements in Albrecht Du¨rer’s (1471–1528) famous painting Melencolia I also contained the 4 × 4 magic square of $4^2 = 16$ integers from 1 to 16 arranged in 4 rows and 4 columns so that it has the property similar to a 3 × 3 magic square (1.1), that is, the sum of each row, each column and each of the two diagonals is 34. Interestingly, the two middle elements of the last row of (1.2) represent the date of the painting as 1514. In modern notation, this magic square (1.2) is nothing but a 4 4 square matrix of $4^2 = 16$ elements.

4	9	2
3	5	7
8	1	6

→ (1.1)

16	3	2	13
5	14	11	8
9	6	7	12
4	15	14	1

→ (1.2)

In spite of considerable study of magic squares for a long period of time, it was B. de Bessey Frenicle (1605–1675), a French mathematician interested in number theory and combinatories, who first described 880 different 4×4 magic squares in some detail. His work was first published posthumously in 1693. Based on the fourth-order magic square written originally by Frenicle himself as considerable research has recently been done on magic squares of order 4 and higher orders and their properties in the 1980s.

Many additional magic squares were also found in Chinese, Indian and Japanese ancient mathematics with more geometrical figures, including magic circles and magic hexagons. In general, if M_n is an $n \times n$ magic square that contains each of the entries 1, 2, 3, …, n^2 exactly one with the same sum of each row, each column and each of the two diagonals, then the common sum is called the *weight* (or *magic constant*) of M_n, denoted by $\mathrm{wt}(M_n)$ which is defined by

$$\begin{matrix} a & b & c & d \\ e & f & g & h \\ i & k & l & m \\ n & o & p & q \end{matrix} \longrightarrow \quad (1.3)$$

$$\mathrm{wt}(M_n) = \frac{n(n^2+1)}{2} = \frac{1}{n}\sum_{k=1}^{n^2} k. \longrightarrow \quad (1.4)$$

Thus, the weights of magic squares of order n=3, 4, 5, … are 15, 34, 65, … , respectively.

These extraordinary examples of magic squares illustrate the power and ability of recre-ational mathematics which can often lead to the most profound discoveries in mathematical Sciences.

2. Matrices and systems of linear equations

The above introduction to the magic squares led to the idea of matrices. A *matrix* is a rectangular array of real, complex or any other mathematical objects. The name matrix (the Latin word 'womb', originated from 'mater' – 'mother') was given by one of the founders of modern matrix algebra, the British mathematician, James Joseph Sylvester (1814–1887).

A matrix $A = (a_{ij})$ of m rows and n columns (m and n are positive integers) is called an $m \times n$ *matrix* defined by

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{22} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \longrightarrow (2.1)$$

where a_{ij} are called *elements* (or *entries*) in the ith row and jth column in A. If m n, the matrix A (a_{ij}) is called a *square matrix* of order n. A1×1 matrix is simply a number, for example, (-5) = -5. In particular, the elements a_{11}, a_{22}, … , a_{nn} are called the *diagonal elements* of A. If in a square matrix all elements except the diagonal elements are 0, the matrix is called a *diagonal matrix*. If all diagonal elements of a diagonal matrix is 1, then A is called an *identity matrix* which is usually denoted by I.

Historically, the first use of matrix methods to solve simultaneous equations was found in the Chinese text on *The Nine Chapters on the Mathematical Art* written during the Han Dynasty (202 BC–200 AD). In 1683, Seki Kowa (1642–1708) in Japan first discovered the idea of determinants to solve two simultaneous quadratic equations and then developed many interesting properties of determinants. At about the same time, determinants were introduced in Europe, and in 1693, Gottfried Wilhelm Leibniz (1646–1716) initiated his study of a system of linear equations using determinants. Subsequently, in 1729, Colin Maclaurin (1698–1746) used the method of determinants to solve systems of simultaneous linear equations in two, three and four unknowns, and his work was then published in his posthumous *Treatise of Algebra* in 1748.

In 1750, Grabriel Cramer (1704–1752), a great Swiss mathematician, formulated a general rule for solving n algebraic equations in n unknowns $x_1, x_2, \ldots, x_n$ of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \quad\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \longrightarrow (2.2)$$

where a_{ij} are real or complex coefficients and b_i (I = 1, 2, 3,… , n, j = 1, 2,…, n) are non-homogeneous terms in the system (2.2).

where a_{ij} are real or complex coefficients and b_i (I = 1, 2, 3,… , n, *and* j = 1, 2,…, n) are non-homogeneous terms in the system (2.2). This system of n algebraic equations in n unknowns can be expressed in terms of the $n \times n$ square matrix $A = (a_{ij})$ given by (2.1) of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \longrightarrow \quad (2.3)$$

where $\mathbf{x}$ and $\mathbf{b}$ are $n \times 1$ *column matrices* (or *column vectors*) of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}. \longrightarrow \quad (2.4)$$

Cramer solved the system (2.3) in terms of determinants of the form

$$x_i = \frac{\det A_i}{\det A} = \frac{|A_i|}{|A|}, \longrightarrow \quad (2.5)$$

where $I = 1, 2, 3, \dots, n$ and $\det A = IAI$ is called the *determinant* which is defined by the unique scalar associated with the matrix A as

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \longrightarrow \quad (2.6)$$

which can be expressed as the sum of all $n!$ terms as

$$\det A = |A| = \sum (\pm 1)(a_{1i} a_{2j} a_{3k} \cdots), \longrightarrow \quad (2.7)$$

where the row suffixes appear in normal order, while the column suffixes $i, j, k, \ldots$ (n altogether) appear as some permutation of the normal order $1, 2, 3, \ldots, n$. In (2.5), $\det A_i = |A_i|$ is the determinant of order n obtained from $\det A = |A|$ by replacing its ith column by the column containing the non-homogeneous term $b_1, b_2, \ldots, b_n$.

In general, matrices and systems of linear equations arise in too many areas of mathe- matics, science, engineering, business and industry. Matrices have innumerable applications in solving problems in quantum mechanics, general theory of relativity, economics, sociology, cryptology (coding and decoding messages), seriation archaelogy, probability and statistics. There are also too many applications of systems of linear equations, including analysis of networks in civil and electrical engineering, transportation, communication, traffic flow and information theory.

Historically, determinants were discovered over two centuries before the discovery of matrices. In 1841, Arthur Cayley (1821–1895), a famous British mathematician, first introduced the notation of two vertical lines on either side of the array to denote the determinant which has now become a standard. The elegant formula (2.5) is universally known as *Cramer's Rule* which was published by Cramer in his treatise *Introduction to the Analysis of Algebraic Curves* in 1750. Although Cramer's Rule is exact and mathematically elegant, it is computationally inefficient for all but small systems of linear equations because it involves computation of determinants of large order. So, the computation of det A of order n from its definition (2.7) is a major problem of computing $n!$ terms. Indeed, the computation of very large determinants is almost a formidable task. In solving a system of n linear equations in n unknowns that has a unique solution, the Cramer Rule involves
$(1/3n^4 + 1/3\, n^3 + 2/3\, n^2 + 1/3\, n - 1)$ multiplications and $(1/3\, n^4 - 1/6\, n^3 - 1/3\, n^2 + 1/6\, n)$ additions so that the total number of arithmetic operations required, $T(n)$, is the sum of the above two expressions which is approximately equal to $T(n) = 2/3\, n^4$ for large n. Subsequently, other efficient iterative and numerical techniques, including the Gauss elimination and the Gauss–Jordan elimination, have replaced Cramer's Rule for solving linear systems of equations.

Carl Friedrich Gauss (1777–1855) discovered the most famous algorithm for finding the general solution of a system of linear equations (2.3) by reducing the associated augmented matrix of the system to a triangular form so that the final solutions can be obtained by back substitution. This algorithm is universally known as the *Gauss elimination*. However, it was known to Chinese mathematicians in the third century BC, but it bears the name of Gauss because of his rediscovery of the method for finding solutions of a system of linear equations to describe the orbit of a planet asteroid. Subsequently, Wilhelm Jordan (1842–1899), a German mathematician, who made an important modification of the Gauss elimination algorithm, now known as the *Gauss–Jordan elimination method*, which simplifies the back substitution process. In fact, the Gauss–Jordan procedure involves elimination of the unknown x_k in the kth step of the procedure, not just in the kth equation, but also in all preceding equations of the system.

In solving a system of n equations with n unknowns, the Gauss elimination method requires a total operation, $T(n)$, of ($1/3n^3 + n^2 - 1/3\ n$) multiplications and ($1/3n^3 + 1/2\ n^2 - 5/6\ n$) additions. Thus, $T\ (n) \approx 2/3\ n^3$ for large n. On the other hand, the Gauss–Jordan elimination method requires the total number of operations, $T(n)$, which is the sum of ($1/2\ n^3 + 1/2\ n^2$) multiplications and ($1/3\ n^3 - 1/2\ n$) additions so that $T(n) \approx n3$ for large n. Thus, the Gauss elimination requires fewer operations of multiplication and addition, and hence it is 50% more efficient than the Gauss–Jordan elimination method. Hence, the reduction in the total number of arithmetic operations not only saves the computer time, but also increases the accuracy of the final solution as with less operations that are performed, the smaller the possible round-off errors. In general, the Gauss elimination is more accurate as well as more efficient than the Gauss–Jordan elimination method. Further, it is definitely more accurate and efficient than Cramer's Rule.

For example, the determinant of a 2 × 2 square matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \longrightarrow \quad (2.8)$$

is

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \longrightarrow \quad (2.9)$$

Similarly, the determinant of a 3 × 3 square matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \qquad (2.10)$$

is given by

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \longrightarrow \qquad (2.11)$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{32} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \longrightarrow \qquad (2.12)$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{32} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}). \longrightarrow \qquad (2.13)$$

Thus, det $A = |A|$ in (2.12) can be rewritten in a compact form as

$$|A| = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13}, \longrightarrow \qquad (2.14)$$

$$= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}, \longrightarrow \qquad (2.15)$$

where the determinant M_{ij} of order $n-1$ obtained from det $A = |a_{ij}|$ by deleting the row and the column containing the element a_{ij} is called the *minor* of the element a_{ij}, and the determinant A_{ij} of order $n-1$ of the element a_{ij} is called the *cofactor* of a_{ij} defined by

$$A_{ij} = (-1)^{i+j} M_{ij}. \tag{2.16}$$

Hence, $|A|$ in (2.14) and (2.15) has been expanded as the sum of elements of a row multiplied by their own minors or cofactors along the ith row:

$$|A| = \sum_{j=1}^{3} (-1)^{1+j} a_{1j} M_{1j} = \sum_{j=1}^{3} a_{1j} A_{1j}. \tag{2.17}$$

Similarly, in general, the determinant A of order n can be expanded as the sum of a row (or column) multiplied by their own cofactors of the forms

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{j=1}^{n} a_{ij} A_{ij}. \tag{2.18}$$

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij}. \tag{2.19}$$

Both expansions (2.18) and (2.19) are universally known as the *Laplace Expansion Theorem*, since Pierre Simon Laplace (1749–1827), a famous French mathematician, proved them in his work on the theory of determinants.

There is a simple consequence of (2.18)–(2.19) that follows from replacement of the ith row of $|A|$ by the kth row (or the jth column by the kth column) so that

$$0 = \sum_{j=1}^{n} a_{kj} A_{ij}, \quad k \neq i, \qquad 0 = \sum_{i=1}^{n} a_{ik} A_{ij}, \quad k \neq j, \tag{2.20}$$

where the sum of elements of a row (or column) is multiplied by the corresponding cofactor of another row (or column).

In compact form, (2.18)–(2.20) can be written as

$$\sum_{j=1}^{n} a_{kj} A_{ij} = |A|\,\delta_{ik}, \qquad \sum_{i=1}^{n} a_{ik} A_{ij} = |A|\,\delta_{kj},$$

where δ_{ik} is the Kronecker delta function defined by

$$\delta_{ik} = \begin{Bmatrix} 0, & i \neq k \\ 1, & i = k \end{Bmatrix}.$$

3. Algebraic properties of matrices

(1) *Equality.* Two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ are *equal* if the corresponding elements are equal, that is, $a_{ij} = b_{ij}$ for each i and j.

(2) *Addition of matrices.* The sum of two $m \times n$ matrices, $A = (a_{ij})$ and $B = (b_{ij})$, is defined as the $m \times n$ matrix $C = (c_{ij})$ obtained by adding corresponding elements so that

$$C = A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}) = (c_{ij}). \tag{3.1}$$

It follows from this definition that matrix addition is commutative and associative so that

$$A + B = B + A, \qquad A + (B + C) = (A + B) + C. \tag{3.2}$$

(3) *Zero (or null) matrix.* An $m \times n$ matrix $A = (a_{ij})$ is called a *zero* (or *null*) *matrix* if all its elements $a_{ij} = 0$, and it is denoted by O. For every $m \times n$ matrix A, there exists an $m \times n$ zero matrix O such that $A + O = A = O + A$. This zero matrix is the *additive identity element* for the set of all $m \times n$ matrices.

(4) *Scalar multiplication.* The product of a matrix $A = (a_{ij})$ by a scalar (real or complex) α is defined as follows:

$$\alpha A = \alpha(a_{ij}) = (\alpha a_{ij}). \tag{3.3}$$

Thus, the following distributive law holds:

$$\alpha (A + B) = \alpha A + \beta B, \qquad (\alpha + \beta) A = \alpha A + \beta A, \tag{3.4}$$

for any two scalars α and β. In particular, the negative of A is denoted by $-A$ and defined by

$$-A = (-1) A. \tag{3.5}$$

(5) *Subtraction of matrices.* The difference $A - B$ of two $m \times n$ matrices, $A = (a_{ij})$ and $B = (b_{ij})$, is defined by

$$A - B = A + (-B) = (a_{ij}) + (-b_{ij}) = (a_{ij} - b_{ij}). \tag{3.6}$$

Thus, the difference (3.6) is similar to the sum (3.1).

(6) *Transpose of a matrix.* The *transpose* of a matrix is obtained by interchanging the rows by columns and is denoted by A^T. If $A = (a_{ij})$ is an $m \times n$ matrix given

by (2.1), then

$$A^T = (a_{ij})^T = (a_{ji}) = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}. \tag{3.7}$$

Thus, A^T is an $n \times m$ matrix with elements a_{ji}. If $A = (a_{ij})$, then $A^T = (a_{ji})$ and $(A^T)^T = (a_{ji})^T = (a_{ij}) = A$. Thus, the relation between A and A^T is symmetric, either matrix being a transpose of the other.

(7) *Row and column matrices (vectors)*. A *row matrix* $\mathbf{x}$ is a $1 \times n$ matrix so that $\mathbf{x} = (x_1\, x_2 \cdots x_n)$. A *column matrix*is an $n \times 1$ matrix so that the transpose of $\mathbf{x}$ is a column matrix

$$\mathbf{x}^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \tag{3.8}$$

Both row and column matrices can be regarded as row and column *vectors*. A matrix may be considered as the generalization of a vector. Thus, the transpose of a column matrix (vector) is a row matrix (vector) so that

$$\left(\mathbf{x}^T\right)^T = (x_1\, x_2 \cdots x_n) = \mathbf{x}. \tag{3.9}$$

It is convenient to use row vectors and column vectors in matrix algebra.

(8) *Matrix multiplication.* The *productAB* of two matrices is defined if the number of columns in A is the same as the number of rows of B. If $A = (a_{ij})$ is an $m \times n$ matrix and $B = (b_{ij})$ is an $n \times r$ matrix, then the product $C = AB$ is an $m \times r$ matrix, and the elements c_{ij} of $C = (c_{ij})$ are defined by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}. \tag{3.10}$$

In general, matrix multiplication is *non-commutative*. In order for both AB and BA to exist, it is necessary that A and B are square matrices of the same order. Even in that case two products may not be equal, so that, in general,

$$AB \neq BA. \tag{3.11}$$

For example,

$$A = \begin{pmatrix} 2 & 4 \\ 1 & -2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

$$AB = \begin{pmatrix} 2+4 & 0+4 \\ 1-2 & 0-2 \end{pmatrix} = \begin{pmatrix} 6 & 4 \\ -1 & -2 \end{pmatrix}$$

$$BA = \begin{pmatrix} 2+0 & 4+0 \\ 2+1 & 4-2 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 3 & 2 \end{pmatrix}.$$

Thus, $AB \neq BA$. Unlike the multiplication of numbers in ordinary algebra, the matrix multiplication is non-commutative.

If

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$

then $AB = O$, but neither A nor B is a zero matrix. In ordinary algebra, if $ab = 0$, then $a = 0$ or $b = 0$, that is, zero divisors do not exist. However, in matrix algebra, it is possible to have zero divisors. This is a striking contrast between ordinary algebra and matrix algebra.

It can be shown by direct calculation that matrix multiplication satisfies the associative and distributive laws:

$$(AB)C = A(BC), \qquad A(B+C) = AB + AC, \qquad (A+B)C = AC + BC. \tag{3.12}$$

(9) *Identity (unit) matrix*. An $n \times n$ square matrix $I = I_n = (\delta_{ij})$ is called an *identity* (or *unit*) *matrix* if δ_{ij} is the Kronecker delta symbol defined in (2.22). Thus, $IA = A = AI$ for any $n \times n$ matrix A. It is easy to check that if A is any $m \times n$ matrix, then $I_m A = A = A I_n$. We write $IA = A = IA$, since each matrix product is well defined for exactly one size of the identity matrix.

(10) *Inverse of a matrix*. If A is an $n \times n$ matrix, then *inverse* of A is an $n \times n$ matrix $X = A^{-1}$ which satisfies the property

$$AX = I = XA, \tag{3.13}$$

where $I = (\delta_{ij})$ is an $n \times n$ identity matrix defined by the Kronecker delta function, δ_{ij}, in (2.22) so that $IA = AI = A$, and I is the *multiplicative identity matrix* for the set of all $n \times n$ matrices.

Using this definition, it is easy to find the inverse of an 2×2 matrix, if it exists,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{3.14}$$

Suppose its inverse $X = A^{-1}$ is given by

$$X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}, \tag{3.15}$$

so that $AX = I = XA$, that is,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} ax + bz & ay + bw \\ cx + dz & cy + dw \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.16}$$

Thus, x, y, z, w satisfy the system of four linear equations

$$ax + bz = 1, \qquad ay + bw = 0,$$
$$cx + dz = 0, \qquad cy + dw = 1.$$

By simple elimination, the solutions are given by

$$x = \frac{d}{|A|}, \qquad y = -\frac{b}{|A|}, \qquad z = -\frac{c}{|A|}, \qquad w = \frac{a}{|A|}, \tag{3.17}$$

provided $|A| = ad - bc \neq 0$. Similarly, $XA = I$ leads to the same solutions (3.17). Thus, the inverse of A exists and is given by

$$X = A^{-1} = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{3.18}$$

provided $|A| \neq 0$, that is, A is a non-singular matrix. This is really a simple and elegant formula for the inverse of the 2×2 non-singular matrix A given by (3.14). However, there is *no* such simple formula for the inverse of larger square matrices. However, it is possible to use other effective methods to calculate the inverse of $n \times n$ matrices. The most famous algorithm used to find the inverse of a non-singular matrix is known as the *Gauss–Jordan elimination*. This algorithm was originally discovered by Gauss and then subsequently modified by Jordan.

The inversion of an $n \times n$ matrix A requires approximately n^3 addition and multiplication operations which means it involves a large amount of computational work.

We first develop a method of solving the matrix equation

$$A\mathbf{x} = \mathbf{b}, \tag{3.19}$$

similar to the solution of the ordinary scalar equation $ax = b$, where a, b and x are real numbers. Thus, if $a \neq 0$, the unique solution is $x = a^{-1}b = \frac{b}{a}$. If $a = 0$, then the equation has an infinite number of solutions for x. We use the same procedure to solve the matrix equation (3.19) for a non-singular matrix A so that $|A| \neq 0$, and then we multiply (3.19) by A^{-1} to obtain the solution

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \qquad \text{or} \qquad \mathbf{x} = A^{-1}\mathbf{b}. \tag{3.20}$$

Moreover,

$$A\mathbf{x} = AA^{-1}\mathbf{b} = \mathbf{b}.$$

This proves that $\mathbf{x} = A^{-1}\mathbf{b}$ is the unique solution of the matrix equation (3.19). So, the natural question is how to find A^{-1} of an $n \times n$ matrix A. We can answer this question by introducing the *adjugate matrix* of $A = (a_{ij})$, denoted by adj A, and defined by

$$\operatorname{adj} A = (A_{ij})^T = (A_{ji}), \tag{3.21}$$

where A_{ij} are the cofactors of elements a_{ij} in the associated determinant $|A|$.

We next use the same example of the 2×2 matrix A given by (3.14) to find A^{-1} using the idea of its adjugate matrix. The cofactors of the associated determinant

$|A|$ are $A_{11} = d$, $A_{12} = -c$, $A_{21} = -b$ and $A_{22} = a$ so that

$$\text{adj } A = (A_{ij})^T = A_{ji} = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{3.22}$$

Thus,

$$A\,(\text{adj } A) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = \begin{pmatrix} |A| & 0 \\ 0 & |A| \end{pmatrix}$$
$$= |A| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = |A|\, I, \tag{3.23}$$

where

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \neq 0.$$

Similarly, it is easy to verify that

$$(\text{adj } A)\, A = |A|\, I. \tag{3.24}$$

Thus, it follows from (3.23) and (3.24) that

$$A\,\frac{(\text{adj } A)}{|A|} = I = \frac{(\text{adj } A)}{|A|} A. \tag{3.25}$$

This proves that A^{-1} exists and is given by

$$A^{-1} = \frac{1}{|A|}\,\text{adj } A = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{3.26}$$

This is the same formula (3.18) for A^{-1} obtained before from the definition (3.13).

The formula (3.25) is valid for an $n \times n$ non-singular matrix A which directly follows from the matrix multiplication and the Laplace expansion theorem. More explicitly, the product $A(\mathrm{adj}\, A)$ gives

$$A\,(\mathrm{adj}\, A) = (a_{ij})(A_{ij})^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

which is, by (2.20)–(2.21ab),

$$= \begin{pmatrix} |A| & 0 & \cdots & 0 \\ 0 & |A| & \cdots & 0 \\ \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & |A| \end{pmatrix} = |A|\, I. \qquad (3.27)$$

Similarly,

$$(\text{adj } A)\, A = |A|\, I = A\, (\text{adj } A)\,. \tag{3.28}$$

Thus, (3.27) and (3.28) lead to (3.25). Moreover, the inverse of the $n \times n$ non-singular matrix $A(|A| \neq 0)$ is given by

$$A^{-1} = \frac{1}{|A|}\, (\text{adj } A)\,. \tag{3.29}$$

The following properties can be proved for the inverse of the transpose matrix A^T and the product matrix AB:

$$(A^T)^{-1} = (A^{-1})^T, \tag{3.30}$$

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{3.31}$$

If A is a singular matrix, then $|A| = 0$ and hence formula (3.29) becomes invalid, and A^{-1} cannot be determined by (3.29). However, it follows from (3.28) that

$$A\, (\text{adj } A) = (\text{adj } A)\, A = I\, |A| = O\,. \tag{3.32}$$

In other words, the product of a singular matrix A and its adj A is the null matrix.

For a non-singular matrix A, (3.28) is still valid. We use $|AB| = |A||B|$ for any two square matrices A and B so that it follows from (3.28) that

$$|A|\, |\text{adj } A| = |I\, |A|| = |A|^n\,, \tag{3.33}$$

or

$$|\text{adj } A| = |A|^{n-1}\,. \tag{3.34}$$

In solving the linear system of equations (3.19), it often happens that small changes in the elements of the matrix A produce large changes in the solution of (3.19). In such a case, the matrix A is called *ill-conditioned*. On the other hand, if small changes of the elements of A produce only small changes in the solution of (3.19), then the matrix A is called *well-conditioned*.

(11) *Symmetric and skew-symmetric matrices.* A matrix $A = (a_{ij})$ is called *symmetric* if $A = A^T$, that is, $a_{ij} = a_{ji}$ for all i and j.

A matrix $A = (a_{ij})$ is called *skew-symmetric* or *anti-symmetric* if $A = -A^T$, that is, $a_{ij} = -a_{ji}$ for all i and j.

A matrix to be symmetric or skew-symmetric must necessarily be a square matrix. Moreover, the diagonal elements of a skew-symmetric matrix must be zero, since $a_{ii} = -a_{ii}$ which means $a_{ii} = 0$ for all i.

The matrix

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & -2 \\ 0 & -2 & 4 \end{pmatrix}$$

is a 3×3 symmetric matrix.

The matrix

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

is a 3×3 skew-symmetric matrix. However, there are matrices which are neither symmetric nor skew-symmetric. The matrix

$$C = \begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$$

is an example of neither symmetric nor skew-symmetric.

(12) *Conjugate, Hermitian and skew-Hermitian matrices.* A *real* (or *complex*) *matrix* is a matrix whose elements are real (or complex) numbers. The *conjugate* of a complex matrix $A = (a_{ij})$ is denoted by $\bar{A}$ and is obtained by replacing a_{ij} by its complex conjugate $\bar{a}_{ij}$ so that we can write

$$\overline{A} = (\overline{a_{ij}}) \qquad (3.35)$$

and hence $\overline{\overline{A}} = (\overline{\overline{a_{ij}}}) = (a_{ij})$, and the relation between A and $\overline{A}$ is symmetric, since either matrix is the conjugate of the other. Note that the matrix A is real if and only if $A = \overline{A}$.

The *transpose* of the conjugate matrix is denoted by A^*, and is defined by $A^* = (\overline{A})^T$. If $A = (a_{ij})$, then

$$A^T = (a_{ji}) \quad \text{and} \quad \overline{A} = (\overline{a_{ij}}) \qquad (3.36)$$

and

$$(\overline{A})^T = (\overline{a_{ji}}) = \overline{(A^T)}. \qquad (3.37)$$

This means that the transpose of the conjugate of a matrix is equal to the conjugate of the transpose matrix.

Furthermore, $\overline{AB} = \overline{A}\,\overline{B}$. Writing $AB = C = (c_{ij})$ and using the rule of matrix multiplication gives

$$c_{ij} = \sum_{k=1}^{n} a_{ik}\, b_{kj}.$$

Thus,

$$\overline{c_{ij}} = \overline{\sum_{k=1}^{n} a_{ik}\, b_{kj}} = \sum_{k=1}^{n} \overline{a_{ik}\, b_{kj}} = \sum_{k=1}^{n} \overline{a_{ik}}\,\overline{b_{kj}},$$

since the conjugate of the sum and the conjugate of the product of two complex numbers are equal to the sum of their conjugates and to the product of their conjugates, respectively. Consequently, $\overline{AB} = \overline{A}\,\overline{B}$, that is, the conjugate of the product of two matrices is equal to the product of their conjugates.

For example, if

$$A = \begin{pmatrix} 1 & 3+i \\ 4+3i & -5+2i \end{pmatrix}, \quad \text{then} \quad A^T = \begin{pmatrix} 1 & 4+3i \\ 3+i & -5+2i \end{pmatrix},$$

and

$$\overline{A} = \begin{pmatrix} 1 & 3-i \\ 4-3i & -5-2i \end{pmatrix},$$

and hence

$$A^* = \left(\overline{A}\right)^T = \begin{pmatrix} 1 & 4-3i \\ 3-i & -5-2i \end{pmatrix} = \overline{\left(A^T\right)}.$$

A (complex) matrix $A = (a_{ij})$ is called *Hermitian* if $A = A^* = \left(\overline{A}\right)^T$, that is, $a_{ij} = \overline{a_{ji}}$ for all i and j. Since $a_{ii} = \overline{a_{ii}}$, the diagonal elements a_{ii} of a Hermitian matrix are all real numbers. If A is a real symmetric matrix, then $a_{ij} = a_{ji}$ and $a_{ij} = \overline{a_{ji}}$ for all i and j, since $a_{ji} = \overline{a_{ji}}$. Thus, every real symmetric matrix A is a Hermitian matrix. Hermitian matrices are named after Charles Hermite (1822–1901), a renowned French mathematician who first discussed the theory of Hermitian matrices and their properties in 1854.

A matrix A is called *skew-Hermitian* if $A = -A^* = -\left(\overline{A}\right)^T$, that is, $a_{ij} = -\left(\overline{a_{ji}}\right)$ for all i and j. Every real skew-symmetric matrix is a skew-Hermitian matrix.

It follows from the definition that the diagonal elements of a skew-Hermitian matrix are either purely imaginary or zero because $\overline{a_{ii}} = -a_{ii}$, or a_{ii} is purely imaginary or zero.

(13) *Orthogonal and unitary matrices.* A square matrix is called *orthogonal* if $A A^T = I$. This definition implies that $A^T A = I$. Hence, $A^T = A^{-1}$, that is, a matrix is orthogonal if and only if its transpose is equal to its inverse. Further, if A is orthogonal, then its inverse is also orthogonal which follows from the fact that

$$(A^{-1})(A^{-1})^T = (A^T)(A^T)^T = A^T A = I. \tag{3.38}$$

If A is an orthogonal matrix, then $|AA^T| = |I| = 1$ and hence $|A||A^T| = 1$. Since $|A| = |A^T|$, $|A|^2 = 1$ or $|A| = \pm 1$. If $|A| = 1$, then A is called a *proper orthogonal matrix*. If $|A| = -1$, then the matrix A is called an *improper orthogonal matrix*. The rotation of the coordinate axes through an angle θ is represented by a rotation matrix

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{3.39}$$

This matrix R is orthogonal, since

$$R R^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Also, $|R| = 1$; hence, R is a proper orthogonal matrix. The linear transformation $\mathbf{x}' = R\,\mathbf{x}$ represents the rotation of the Cartesian coordinate system xy in the plane through an angle θ about the origin where $\mathbf{x} = (x\ y)^T$ and $\mathbf{x}' = (x'\ y')^T$.

A complex matrix A is called *unitary* if and only if $AA^* = A^*A = I$, that is, if $\left(\overline{A}\right)^T = A^{-1}$. A real unitary matrix is an orthogonal matrix and the determinant of a unitary matrix is ± 1. For example, the matrix A

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \tag{3.40}$$

is unitary and its determinant is -1. Furthermore, the product of two unitary matrices is unitary, since

$$(AB)^{-1} = B^{-1}A^{-1} = \left(\overline{B}\right)^T \left(\overline{A}\right)^T = \left(\overline{AB}\right)^T = \left(\overline{A}\,\overline{B}\right)^T. \tag{3.41}$$

(14) *Triangular and diagonal matrices.* An $n \times n$ matrix $A = (a_{ij})$ is called a *triangular matrix* if (i) $a_{ij} = 0$ for $i < j$, or (ii) $a_{ij} = 0$ for $i > j$. More precisely, in case (i), A is called the *lower triangular matrix*, that is, all elements above the main diagonal are zero. In case (ii), A is called the *upper diagonal matrix*, that is, all elements below the main diagonal are zero. Elements on the main diagonal can be zero or non-zero.

An $n \times n$ matrix $D = (d_{ij})$ is called a *diagonal matrix* if $d_{ij} = 0$ for $i \neq j$, that is, if all elements above and below the main diagonal are zero. So, any diagonal matrix is both upper and lower triangular. A diagonal matrix $D = (d_{ii})$ is called a *scalar matrix* if all d_{ii} are equal to scalar k. Thus, if D is a scalar matrix and A is an $n \times n$ matrix, then $AD = DA = kA$, that is, D commutes with any matrix A and the multiplication by D has the same effect as the multiplication by the scalar k. In particular, the scalar matrix D whose all elements on the main diagonal is 1 is the *identity* (or *unit*) *matrix* so that $AI = IA = A$.

(15) *Rank of a matrix.* The maximum number of linearly independent row vectors of an $n \times n$ matrix $A = (a_{ij})$ is called the *rank* of A, and is denoted by rank A. Similarly, the rank of A is equal to the maximum number of linearly independent column vectors of A. Hence, rank$A$ = rank(A^T). If A is a null matrix, then its rank is zero. For example, the ranks of the following matrices

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 1 & 1 & 5 \end{pmatrix}$$

are 2, 3 and 2, respectively.

If $A = (a_{ij})$ is an $n \times n$ non-singular matrix, then its rank is n, and if A is a singular matrix, then rank A is less than n.

The concept of the rank of a matrix was first defined by a German mathematician, Georg Frobenius (1849–1917) who made major contributions to matrix algebra, differential equations, group theory and group representations.

The reader is referred to Olver and Shakiban [2] and Poole [3] for detailed proofs of algebraic properties of matrices.

Next Lecture
Lecture-II

Matrices in graph theory and electrical networks