Elements of Matrix Theory

M.T. Nair (I.I.T. Madras)

March 2007

1 Introduction – Notations and Preliminaries

Consider the following simultaneous equations - three equations in three unknowns:

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & b_2 \\ a_{31}x_1 & + & a_{31}x_2 & + & a_{33}x_3 & = & b_3 \end{array}$$

where a_{ij} for $i, j \in \{1, 2, 3\}$ and b_1, b_2, b_3 are known quantities, and x_1, x_2, x_3 are unknown quantities to be determined. Here, by 'quantities' we mean real or complex numbers, and are often called **scalars**. Using the notations

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

we may write the above simultaneous equations as

$$Ax = b.$$

Here we used the convention:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 \end{bmatrix}$$

and

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \iff u_1 = v_1, \quad u_2 = v_2, \quad u_3 = v_3.$$

Instead of the above simultaneous equations, three equations in three unknowns, we may consider m equations in n unknowns:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ \ldots & + & \ldots & + & \ldots & + & \ldots & = & \ldots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array} \tag{1.1}$$

By a **solution** of (1.1) we mean a n-tuple $(x_1, \dots, x_n)$ of scalars such that (1.1) is satisfied.

The system of equations (1.1) can be represented in the form $Ax = b$ with

$$A := \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad x := \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \quad b := \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}.$$

Again, we used the convention:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 & + & \dots & + & a_{1n}x_n \\ a_{21}x_1 & + & \dots & + & a_{2n}x_n \\ \dots & + & \dots & + & \dots \\ a_{m1}x_1 & + & \dots & + & a_{mn}x_n \end{bmatrix}. \tag{1.2}$$

The above *matrix* A contains m rows and n columns. Hence A is called an $m \times n$ **matrix**, (read as an m by n matrix), and A is said to be of **order** $m \times n$. If a_{ij} is the entry at the i^{th} row and j^{th} column of A, then we shall write

$$A = (a_{ij}) \quad \text{or} \quad A = (a_{ij})_{m \times n}.$$

An $n \times 1$ matrix is called a **column vector** or more precisely, an **column n-vector**, and a $1 \times n$ matrix is called a **row vector** or more precisely, a **row n-vector**. Thus, x and b above are column vectors.

As per the convention (1.2), if $A = (a_{ij})_{m \times n}$ is an $m \times n$ matrix and x is a column n-vector with its i^{th}-entry x_i, then $Ax =: y$ is a column m-vector with its i^{th}-entry y_i given by

$$y_i = \sum_{j=1}^{n} a_{ij}x_j := a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n, \quad i = 1, \dots, m.$$

We say that two $m \times n$ matrices $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{m \times n}$ are equal if their corresponding entries are equal, that is,

$$A = B \iff a_{ij} = b_{ij}, \quad i = 1, \dots, m; \ j = 1, \dots, n.$$

In view of the convention (1.2), an $m \times n$ matrix $A = (a_{ij})_{m \times n}$ can be considered as a *transformation* which maps column n-vectors onto coumn m-vectors. Thus, if $e_1, \dots, e_n$ are the **coordinate unit vectors** defined by

$$e_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix}, \quad e_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix}, \dots, e_n := \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 0 \\ 1 \end{bmatrix}$$

then, under the transformation induced by the matrix A, these column vectors $e_1, \dots, e_n$ are mapped onto the column vectors of A, namely,

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \cdots \\ a_{m1} \end{bmatrix}, \quad \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \cdots \\ a_{m2} \end{bmatrix}, \dots, \begin{bmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \cdots \\ a_{mn} \end{bmatrix},$$

respectively. Thus, the j^{th} column of A is given by Ae_j.

Here are a few special types of matrices:

• A matrix with all its entries zeroes is called a **zero matrix**, and it is denoted by O.

• An $n \times n$ matrix is called a **square matrix**. Abusing the terminology, we may say that an $n \times n$ square matrix is of **order** n. If A is a matrix of order $m \times n$ with $m \neq n$, then to emphasis this fact, we may say that A is a **rectangular matrix**.

A square matrix $A = (a_{ij})_{n \times n}$ is called

• a **diagonal matrix** if $a_{ij} = 0$ for $i \neq j$;

• an **upper triangular matrix** if $a_{ij} = 0$ for $i < j$;

• a **lower triangular matrix** if $a_{ij} = 0$ for $i > j$;

• an **identity matrix** or unit matrix if it is diagonal with $a_{ii} = 1$. Identity matrix is usually denoted by I, or by I_n if it is of order $n \times n$.

For example, $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is a zero 3×4 matrix, $\begin{bmatrix} 3 & 2 & 3 \\ 1 & 4 & 3 \\ -2 & 3 & 3 \end{bmatrix}$ is a square matrix of order 3, $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix}$ is a diagonal matrix, $\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix}$ is an upper triangular matrix, $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 3 & -3 \end{bmatrix}$ is a lower triangular matrix, and $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is an identity matrix.

2 Addition and Multiplication of Matrices

Let $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{m \times n}$. Then we define $A+B$ as a new matrix which corresponds to the transformation

$$x \mapsto Ax + Bx$$

for column n-vectors x. Since the j^{th} column of $A+B$ has to be $Ae_j + Be_j$, we have

$$A + B = (c_{ij})_{m \times n} \quad \text{with} \quad c_{ij} := a_{ij} + b_{ij}, \quad i = 1, \dots, m; \; j = 1, \dots, n.$$

For scalars α, we define $\alpha A := (\alpha\, a_{ij})_{m\times n}$, and

$$A - B := A + (-1)B.$$

We see that for $m \times n$ matrices A and B and $m \times n$ zero matrix O,

$$A + B = B + A, \quad A + O = A = O + A.$$

Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$. The we define a new matrix, denoted by AB and called **product of A and B**, by the requirement

$$ABx = A(Bx)$$

for all p-vectors x. Since every p-vector x can be written as $x = x_1e_1 + x_2e_2 + \ldots + x_pe_p$, where e_i is the i^{th} coordinate unit p-vector, it follows that

$$ABx = A(Bx) \quad \text{for all } p\text{-vectors } x \iff ABe_j = A(Be_j) \quad \text{for all} \quad j = 1, \ldots, p.$$

Hence, the j^{th} column vector of AB is given by

$$ABe_j = A(Be_j) = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \ldots \\ b_{nj} \end{bmatrix}.$$

Thus,

$$AB = (c_{ij})_{m\times p} \quad \text{with} \quad c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}, \quad i = 1, \ldots, m;\ j = 1, \ldots, p.$$

Note that in order to define product AB of matrices A and B, it is required that the number of columns of A and the number of columns of B to be equal. Thus for square matrices A and B of the same order, we can define AB and BA. But, AB need not be equal to BA (Give examples!).

We see that the above definition of multiplication of matrices is in conformity with the convention of jextaposition Ax considered earlier, where A is an $m \times n$ matrix and x is a column n-vector, that is, x is an $n \times 1$ matrix.

For matrices A, B, C, P, Q, R of appropriate orders, we have

$$A(B + C) = AB + AC, \quad (P + Q)R = PR + QR.$$

In particular, if $A = (a_{ij})_{m\times n}$ and x and u are column n-vectors, then we have

$$A(x + u) = Ax + Au,$$

and for any scalar α,

$$A(\alpha x) = \alpha Ax.$$

Because of these two properties, the transformation induced by A is said to be a *linear transformation* on the space of column n-vectors.

2.1 Transpose of a matrix

Given an $m \times n$ matrix $A := (a_{ij})_{m\times n}$, the $n \times m$ matrix $(b_{ij})_{n\times m}$ with $b_{ij} := a_{ji}$ is called the **transpose** of A and it is denoted by A^T. Thus,

$$A := (a_{ij})_{m\times n} \implies A^T := (b_{ij})_{n\times m} \quad \text{with} \quad b_{ij} := a_{ji}.$$

That is,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \implies A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}.$$

For example,

$$A = \begin{bmatrix} 3 & 2 & 3 \\ 1 & 4 & 3 \\ -2 & 3 & 3 \end{bmatrix} \implies A^T = \begin{bmatrix} 3 & 1 & -2 \\ 2 & 4 & 3 \\ 3 & 3 & 3 \end{bmatrix}.$$

We may observe that for $m \times n$ matrices A and B,

$$(A+B)^T = A^T + B^T, \qquad (A^T)^T = A.$$

Also for an $m \times n$ matrix A and an $n \times p$ matrix B,

$$(AB)^T = B^T A^T.$$

Clearly, if u is a column vector, then u^T is a row vector, and vice-versa.

A square matrix A is called

- **symmetric** if $A^T = A$,
- **skew-symmetric** if $A^T = -A$.

Note that for any square matrix A, $A = B + C$ where $B = \frac{1}{2}(A + A^T)$ is symmetric and $C = \frac{1}{2}(A - A^T)$ is skew-symmetric. We may observe that diagonal entries of a skew symmetric matrix are zeroes.

3 Solution of System of Equations

In the following we use the notation V_n to denote the space of all column n-vectors, $\widetilde{V}_n$ to denote the space of all row n-vectors, and $\mathbb{K}$ for the set of all scalars.

Consider the system of equations (1.1). We know that (1.1) can be represented by a matrix equation

$$Ax = b, \tag{3.3}$$

where $A := (a_{ij})_{m\times n}$, $x \in V_n$ and $b \in V_m$.

We may observe that if $x \in V_n$ is a solution of (3.3) and $u \in V_n$ is a solution of

$$Au = 0, \tag{3.4}$$

then for any $\alpha \in \mathbb{K}$,

$$A(x + \alpha u) = Ax + \alpha Au = b + 0 = b.$$

That is, if x is a solution of the *non-homogeneous equation* (3.3), then for any solution u of the *homogeneous equation* (3.4) and for any scalar α, $x+\alpha u$ is a solution of the non-homogeneous equation (3.3). Also, if v is another solution of (3.3), then

$$A(v - x) = Av - Ax = b - b = 0$$

so that with $u := v - x$, we have $Au = 0$ and $v = x + u$. Thus, knowing a solution x of (3.3), any other solution v of (3.3) can be written as $v = x + u$ where u is a solution of (3.4). In other words, if x is a solution of (3.3), then the set of all solutions of (3.3) is given by $\{x + u : Au = 0\}$. The above observation leads to only one of the following alternatives:

(i) The system (3.3) does not have any solution.

(ii) The system (3.3) has a unique solution.

(iii) The system (3.3) has infinitely many solutions.

Note that that the set

$$N(A) := \{u \in V_n : Au = 0\}$$

has the property that

$$u, v \in N(A),\ \alpha \in \mathbb{K} \implies u + v \in N(A),\ \alpha u \in N(A).$$

A subset having the above property is called a subspace of V_n. More precisely, a subset S of V_n (respectively, $\widetilde{V}_n$) is called a **subspace** of V_n (respectively, $\widetilde{V}_n$) if

$$u, v \in S,\ \alpha \in \mathbb{K} \implies u + v \in S,\ \alpha u \in S.$$

Now, in order to solve the equation (3.3), it would be nice if we can transform the system (3.3) into another form, say

$$\widetilde{A}x = \tilde{b} \tag{3.5}$$

so that (3.5) can be solved rather easily, and the set of solutions of (3.3) and (3.5) are the same. First let us observe the following:

Consider two equations from (1.1), say

$$\sum_{j=1}^{n} a_{rj} x_j = b_r \quad \text{and} \quad \sum_{j=1}^{n} a_{sj} x_j = b_s. \tag{3.6}$$

Note that $(x_1, \ldots, x_n)$ is a solution of (3.6) if and only if it is a solution of

$$\sum_{j=1}^{n} a_{rj}x_j = b_r \quad \text{and} \quad \sum_{j=1}^{n} a_{sj}x_j + \alpha \sum_{j=1}^{n} a_{rj}x_j = b_s + \alpha b_r \tag{3.7}$$

for any $\alpha \in \mathbb{K}$. Now, we may choose α in such a way that the second equation in (3.7) is simpler than the second equation in (3.6). For instance, if $a_{r1} \neq 0$, then we may choose $\alpha := -a_{s1}/a_{r1}$ so that the first term in the second summand in (3.7) is zero. That is, (3.7) will take the form

$$\sum_{j=1}^{n} a_{rj}x_j = b_r \quad \text{and} \quad \sum_{j=2}^{n} (a_{sj} + \alpha\, a_{rj})x_j = b_s + \alpha b_r.$$

In case $a_{r1} = 0$, then we may interchange the r^{th} row and another row in which the first coefficient is nonzero, and do the above procedure. Of course, if $a_{r1} = 0$ for all $r \in \{1, \ldots, m\}$, then we do not have to take the variable x_1 at all. The above operations applied to the system (1.1) are called *elementary operations*. More precisely, the following operations on (1.1) are called **elementary operations**:

- Interchange of any two equations in (1.1),
- Multiplication of any equation in (1.1) by a nonzero scalar,
- Adding a scalar multiplication of an equation in (1.1) to another equation in (1.1).

Instead of applying elementary operations on equations in (1.1) we can perform similar operations, called **elementary row operations**, on the rows of the matrix

$$B := \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix}.$$

The above matrix is obtained by *augmenting* the column vector

$$b := \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{bmatrix} \quad \text{to the matrix} \quad A := \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}.$$

Hence that matrix B given above is called the **augmented matrix** of the system (1.1), and it is denoted by $[A; b]$ or $[A|b]$.

Our idea is to apply elementary row operations on the augmented matrix $[A; b]$ so as to transform it to a simpler form, say $[\widetilde{A}; \tilde{b}]$. We would like to have $[\widetilde{A}; \tilde{b}]$ so simple that the corresponding system of equations can be solved by *back substitutions*.

4 Echelon Form of a Matrix

An $m \times n$ matrix $A = (a_{ij})_{m\times n}$ is said to be an **echelon matrix** if the following conditions are satisfied:

(i) If the r^{th} row is a zero vector, then for each $i \geq r$, the i^{th} row is also zero vector. In other words, if a row is nonzero, then all rows preceding to it are also nonzero.

(ii) First non-zero entry in each nonzero row is 1;

(iii) If $a_{rk} = 1$ is the first non-zero entry in the r^{th} row, then

$$a_{ij} = 0 \quad \text{for} \quad i > r; \&j \leq k.$$

For example,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

are echelon matrices.

We may observe that, any matirx can be transformed into an echelon matrix by applying elementary row operations. It can be shown that echelon form of a matrix is unique.

Now, consider the system of equations (1.1), and let $[A; b]$ be its augmented matrix. It is obvious that if $[\widetilde{A}; \tilde{b}]$ is the echelon form of $[A; b]$, then a solution $(x_1, \ldots, x_n)$ of (1.1) can be obtained by *back substitution*. For example, consider the following system of equations.

$$\begin{array}{rcrcrcrcr} 3x_1 & - & 9x_2 & + & x_3 & + & 4x_4 & = & 1 \\ 2x_1 & - & 13x_2 & + & 3x_3 & - & x_4 & = & 3 \\ -x_1 & + & 8x_2 & - & 2x_3 & + & 3x_4 & = & -2 \end{array}$$

The augmented matrix is

$$[A; b] = \begin{bmatrix} 3 & -9 & 1 & 4 & 1 \\ 2 & -13 & 3 & -1 & 3 \\ -1 & 8 & -2 & 3 & -2 \end{bmatrix}$$

Let us apply row operations to transform this into its echelon matrix:

$$\begin{bmatrix} 3 & -9 & 1 & 4 & 1 \\ 2 & -13 & 3 & -1 & 3 \\ -1 & 8 & -2 & 3 & -2 \end{bmatrix} (R_1 \to \frac{1}{3}R_1) \begin{bmatrix} 1 & -3 & 1/3 & 4/3 & 1/3 \\ 2 & -13 & 3 & -1 & 3 \\ -1 & 8 & -2 & 3 & -2 \end{bmatrix}$$

$$(R_2 \to R_2 - 2R_1)\&(R_3 \to R_3 + R_1) \begin{bmatrix} 1 & -3 & 1/3 & 4/3 & 1/3 \\ 0 & -7 & 7/3 & -11/3 & 7/3 \\ 0 & 5 & -5/3 & 13/3 & -5/3 \end{bmatrix}$$

$$(R_2 \to \frac{-1}{7}R_2)\&(R_3 \to \frac{-3}{5}R_3) \begin{bmatrix} 1 & -3 & 1/3 & 4/3 & 1/3 \\ 0 & 1 & -1/3 & 11/21 & -1/3 \\ 0 & 1 & -1/3 & 13/15 & -1/3 \end{bmatrix}$$

$$(R_3 \to R_3 - R_2) \begin{bmatrix} 1 & -3 & 1/3 & 4/3 & 1/3 \\ 0 & 1 & -1/3 & 11/21 & -1/3 \\ 0 & 0 & 0 & 12/35 & 0 \end{bmatrix}$$

$$(R_3 \to \frac{35}{12}R_3) \begin{bmatrix} 1 & -3 & 1/3 & 4/3 & 1/3 \\ 0 & 1 & -1/3 & 11/21 & -1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Thus the echelon form of $[A : b]$ is

$$[\widetilde{A}; \tilde{b}] = \begin{bmatrix} 1 & -3 & 1/3 & 4/3 & 1/3 \\ 0 & 1 & -1/3 & 11/21 & -1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Hence, it follows that

$$x_4 = 0, \quad x_2 = x_3/3 - 1/3, \quad x_1 = 2x_3/3 - 2/3.$$

Thus, the system has infinitely many solutions, and they are obtained by varying x_3 over $\mathbb{K}$. For $x_3 = 0$, we get the solution as $(-2/3, -1/3, 0, 0)$.

5 Linear Combination and Linear Dependence

In this section we shall consider vectors in V_n, that is, column n-vectors. But the derived results are valid for vectors in $\widetilde{V}_n$, that is, row n-vectors, as well.

Let $u, u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ be column (or row) n-vectors. We say that u is a **linear combination** of $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ if there exist scalars $\alpha_1, \ldots, \alpha_k$ such that

$$u = \alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots + \alpha_k u^{(k)}.$$

The set of all linear combinations of $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ is called the **span** of $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$, and it is denoted by $span(u^{(1)}, u^{(2)}, \ldots, u^{(k)})$. It can be easily seen that

- if $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ are vectors in V_n, then $span(u^{(1)}, u^{(2)}, \ldots, u^{(k)})$ is a subspace of V_n.

We may observe that

- zero vector is a linear combination of any (finite) number of vectors.

A set $\{u^{(1)}, u^{(2)}, \ldots, u^{(k)}\}$ of vectors in V_n is said to be **linearly dependent** if at least one of them can be written as a linear combination of the rest; equivalently, *there exists* a nonzero k-tuple $(\alpha_1, \ldots, \alpha_k)$ of scalars such that

$$\alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots + \alpha_k u^{(k)} = 0.$$

The set $\{u^{(1)}, u^{(2)}, \ldots, u^{(k)}\}$ is said to be **linearly independent** if it is not linearly dependent, equivalently, *for any* nonzero k-tuple $(\alpha_1, \ldots, \alpha_k)$ of scalars,

$$\alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots + \alpha_k u^{(k)} \neq 0.$$

Thus, $\{u^{(1)}, u^{(2)}, \ldots, u^{(k)}\}$ is linearly independent if and only if for any k-tuple $(\alpha_1, \ldots, \alpha_k)$ of scalars,

$$\alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots + \alpha_k u^{(k)} = 0 \implies \alpha_1 = 0,\ \alpha_2 = 0,\ \ldots, \alpha_k = 0.$$

If $\{u^{(1)}, u^{(2)}, \ldots, u^{(k)}\}$ is a linearly dependent (respectively, independent) set of vectors, we may say that the vectors $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ **are linearly dependent** (respectively, independent) or we may say that the vectors $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ **form a linearly dependent** (respectively, independent) set.

For example, the set of column vectors as well as the set of row vectors of an identity matrix are linearly independent. Also, the vectors

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

are linearly independent, whereas

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

are linearly dependent (*Verify!*).

THEOREM 5.1. *Let $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ be linearly independent vectors in V_n. Suppose $v^{(1)}, v^{(2)}, \ldots, v^{(\ell)}$ are in V_n such that for each $i \in \{1, \ldots, k\}$, $u^{(i)}$ is a linear combination of $v^{(1)}, v^{(2)}, \ldots, v^{(\ell)}$, that is,*

$$span(u^{(1)}, u^{(2)}, \ldots, u^{(k)}) \subseteq span(v^{(1)}, v^{(2)}, \ldots, v^{(\ell)}).$$

Then $k \leq \ell$.

Proof. Since $u^{(1)} \in span(v^{(1)}, v^{(2)}, \ldots, v^{(\ell)})$, there exist scalars $\alpha_1^{(1)}, \alpha_2^{(1)} \ldots \alpha_\ell^{(1)}$ such that

$$u^{(1)} = \alpha_1^{(1)} v^{(1)} + \alpha_2^{(1)} v^{(2)} + \ldots \alpha_\ell^{(1)} v^{(\ell)}.$$

All $\alpha_i^{(1)}$ can not be zeroes. W.l.g. we may assume that $\alpha_1^{(1)} \neq 0$. Thus,

$$v^{(1)} \in span(u^{(1)}, v^{(2)}, \ldots, v^{(\ell)}).$$

Since $u^{(2)} \in span(v^{(1)}, v^{(2)}, \ldots, v^{(\ell)})$ it follows from the above that

$$u^{(2)} \in span(u^{(1)}, v^{(2)}, \ldots, v^{(\ell)}).$$

Hence there exist scalars $\alpha_1^{(2)}, \alpha_2^{(2)} \ldots \alpha_\ell^{(2)}$ such that

$$u^{(2)} = \alpha_1^{(2)} u^{(1)} + \alpha_2^{(2)} v^{(2)} + \ldots \alpha_\ell^{(2)} v^{(\ell)}.$$

Since $\{u^{(1)}, u^{(2)}\}$ is linearly independent, all $\alpha_i^{(2)}$ for $i = 2, \ldots, \ell$ can not be zeroes. W.l.g. we may assume that $\alpha_2^{(2)} \neq 0$. Thus,

$$v^{(2)} \in span(u^{(1)}, u^{(2)}, v^{(3)} \ldots, v^{(\ell)}).$$

Since $u^{(3)} \in span(v^{(1)}, v^{(2)}, \ldots, v^{(\ell)})$ it follows from the above that

$$u^{(3)} \in span(u^{(1)}, u^{(2)}, v^{(3)} \ldots, v^{(\ell)}).$$

Now, if $k > \ell$, then at the $(\ell + 1)^{th}$ step, we get

$$u^{(\ell+1)} \in span(u^{(1)}, u^{(2)}, \ldots, u^{(\ell)})$$

which is a contradiction. Hence, $k \leq \ell$. □

COROLLARY 5.2. *If $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ are linearly independent vectors in V_n, then $k \leq n$. In other words, any set consisting of more than n vectors in V_n are linearly dependent.*

Let S be a subspace of V_n, and $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ are linearly independent in S. If $S = span(u^{(1)}, u^{(2)}, \ldots, u^{(k)})$, then we say that $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ is a **basis** of S.

THEOREM 5.3. *Any two bases of a subspace of V_n have the same number of elements.*

Proof. Suppose E_1 and E_2 are bases of a subspace S of V_n. Since E_1 and E_2 are linearly independent sets and $span(E_1) = span(E_2) = S$ it follows from Theorem 5.1 that $E_1^\# \leq E_2^\#$ and $E_2^\# \leq E_1^\#$. Hence, $E_1^\# = E_2^\#$. □

The number of vectors in a basis of a subspace S of V_n is called the **dimension** of S, and it is denoted by $dim(S)$. Thus, if S_1 and S_2 are subspaces of V_n such that $S_1 \subseteq S_2$, then

$$dim(S_1) \leq dim(S_2).$$

6 Rank of a Matrix

Given a matrix A, the maximum number of linearly independent rows of A is called the **row-rank** of A, and the the maximum number of linearly independent columns of A is called the **column-rank** of A.

Let us denote the row rank of a matrix A by $row.rank(A)$ and column rank of A by and $col.rank(A)$. Thus, if $R_1, \ldots, R_m$ are the rows A and $C_1, \ldots, C_n$ are the columns of A, then

$$row.rank(A) = dim(span(R_1, \ldots, R_m)), \quad col.rank(A) = dim(span(C_1, \ldots, C_m)). \tag{6.8}$$

Note also that

$$row.rank(A) = col.rank(A^T), \qquad col.rank(A) = row.rank(A^T) \tag{6.9}$$

As an example, consider the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. Let $u^{(1)}, u^{(2)}, u^{(3)}$ be the columns of A. For scalars $\alpha_1, \alpha_2, \alpha_3$, if $\alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \alpha_3 u^{(3)} = 0$, then it follows that

$$\alpha_1 + \alpha_2 = 0, \quad \alpha_1 + \alpha_2 + \alpha_3 = 0, \quad \alpha_2 + \alpha_3 = 0.$$

From this we get $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$. Thus, $u^{(1)}, u^{(2)}, u^{(3)}$ are linearly independent. In particular, column-rank of A is 3. Similarly we can show that the rows of A are linearly independent.

THEOREM 6.1. *Let A, B, C be matrices such that $A = BC$, then*

$$col.rank(A) \le col.rank(B), \qquad row.rank(A) \le row.rank(C).$$

Proof. Let A, B, C be of orders $m \times n$, $m \times p$ and $p \times n$ respectively. Let columns of A be $A_1, \ldots, A_n$, columns of B be $B_1, \ldots, B_p$, and columns of C be $C_1, \ldots, C_n$. Then, taking e_j as the t^{th} column n-vector, we have

$$A_j = Ae_j = BCe_j = BC_j, \quad j = 1, \ldots, n.$$

Let $C_j = [c_{1j}, c_{2j}, \ldots, c_{pj}]^T$. Then, since $B = [B_1, B_2, \ldots, B_p]$, it follows that

$$A_j = BC_j = c_{1j}B_1 + c_{2j}B_2 \ldots, c_{pj}B_p.$$

Hence,

$$span(A_1, A_2, \ldots, A_n) \subseteq span(B_1, B_2, \ldots, B_n).$$

Consequently, $col.rank(A) \le col.rank(B)$. Now, in view of the relations $A^T = C^T B^T$ and (6.9), we also have

$$row.rank(A) = col.rank(A^T) \le col.rank(C^T) \le row.rank(C).$$

This completes the proof. $\square$

THEOREM 6.2. *The row-rank and column-rank of a matrix are equal.*

Proof. Let A be an $m \times n$ matrix. Let $A_1, A_2, \ldots, A_n$ be the columns of A and $R_1, R_2, \ldots, R_m$ be the rows of A. Let r be the row rank of A so that the maximal linearly independent subset of $\{R_1, R_2, \ldots, R_m\}$ consist of r vectors, say $u^{(1)}, u^{(2)}, \ldots, u^{(r)}$. Then for each $i \in \{1, 2, \ldots, m\}$, there exist $\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ir}$ such that

$$R_i = \alpha_{i1} u^{(1)} + \alpha_{i2} u^{(2)} + \ldots + \alpha_{ir} u^{(r)}.$$

Thus, $A = BC$ with $B = (\alpha_{ij})_{m \times r}$ and $C = [u^{(1)}, u^{(2)}, \ldots, u^{(r)}]^T$. Hence, by Theorem 6.1,

$$col.rank(A) \leq col.rank(B) \leq r = row.rank(A).$$

Since $A^T = C^T B^T$, by the same procedure applied to A^T in place of A, we arrive at the relation $col.rank(A^T) \leq row.rank(A^T)$. Thus,

$$row.rank(A) = col.rank(A^T) \leq row.rank(A^T) = col.rank(A).$$

Thus, we have $col.rank(A) = row.rank(A)$. □

In view of the above theorem we define the **rank** of a matrix to be the maximum number of linearly independent columns of A, which is the same as the maximum number of linearly independent rows of A, and it is denoted by $rank(A)$.

COROLLARY 6.3. *Let A, B, C be matrices of appropriate orders such that $A = BC$. Then*

$$rank(A) \leq \min\{rank(B),\, rank(C)\}.$$

We observe that if $u^{(1)}, u^{(2)}, \ldots u^{(n)}$ are the columns of an $m \times n$ matrix A, and $u = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T$ is a column n-vector, then

$$Au = \alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots + \alpha_n u^{(n)}.$$

Hence, we can conclude the following:

- The columns of A are linearly independent if and only if zero vector is the only solution of the equation $Au = 0$.
- If $rank(A) < n$, then $Au = 0$ has a nonzero solution, consequently, if $Ax = b$ ha a solution then it has infinitely many solutions.
- If $rank(A) = n$, then $Au = 0$ has no nonzero solution, consequently, $Ax = b$ can have at most one solution.

6.1 Existence of Solutions

THEOREM 6.4. *Let A be an $m \times n$ matrix. Then, for a given $b \in V_m$, the equation $Ax = b$ has a solution if and only if*

$$rank(A) = rank([A; b]).$$

Proof. Suppose $rank(A) = r = rank([A; b])$, and let $u^{(1)}, u^{(2)}, \ldots u^{(r)}$ be linearly independent columns of A. Since $r = rank([A; b])$, it follows that $u^{(1)}, u^{(2)}, \ldots u^{(r)}, b$ are linearly dependent. Hence, there exist scalars $\alpha_1, \ldots, \alpha_r$ such that

$$b = \alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots \alpha_r u^{(r)}.$$

Now, let $A = [v^{(1)}, v^{(2)}, \ldots v^{(n)}]$. Since each column of A is a linear combination of $u^{(1)}, u^{(2)}, \ldots u^{(r)}$, it follows that there exist scalars $\beta_1, \ldots, \beta_n$ such that

$$b = \beta_1 v^{(1)} + \beta_2 v^{(2)} + \ldots \beta_r v^{(n)}.$$

In other words, $x = [\beta_1, \ldots, \beta_n]^T$ is a solution of $Ax = b$. We get the converse part of the theorem by retracing the above arguments backwards. $\square$

7 Inverse of a Matrix

Let A be a square matrix of order n. Suppose that for every $b \in V_n$, there exists $x \in V_n$ such that $Ax = b$. Then, in particular, for each $j \in \{1, \ldots, n\}$, there exists $x^{(j)} \in V_n$ such that $Ax^{(j)} = e_j$, where e_j is the j^{th} coordinate unit vector in V_n. Thus,

$$A[x^{(1)}, x^{(2)}, \ldots, x^{(n)}] = [e_1, e_2, \ldots, e_n],$$

that is the matrix $B := [x^{(1)}, x^{(2)}, \ldots, x^{(n)}]$ satisfies the equation

$$AB = I.$$

Conversely, if there exists a square matrix B of order n such that $AB = I$, then for every $b \in V_n$, the vector $x := Bb$ satisfies the equation

$$Ax = b.$$

Thus, we have proved the following theorem.

THEOREM 7.1. *For every $b \in V_n$, there exists $x \in V_n$ such that $Ax = b$ if and only if there exists a square matrix B of order n such that $AB = I$.*

Remark 7.2. *In the above, the matrix A can be of order $m \times n$ with $m \neq n$. In such cases, $b \in V_m$, B is of order $n \times m$ and I is the identity matrix of order m.*

The following results are worth noticing.

THEOREM 7.3. *Let A and B be square matrices of the same order. Then*

$$AB = I \iff BA = I.$$

Proof. Suppose $AB = I$. Let B_j is the j^{th} column of B, that is, $B_j = Be_j$ where e_j is the j^{th} coordinate unit vector in V_n. Then we have $AB_j = ABe_j = e_j$ for $j \in \{1, \ldots, n\}$. From this it follows that $\{B_1, B_2, \ldots, B_n\}$ is a linearly independent set. Consequently, $rank(B) = rank([B; b])$, so that for each j, there exists $u^{(j)} \in V_n$ such that $Bu^{(j)} = e_j$. Thus, with $C := [u^{(1)}, u^{(2)}, \ldots, u^{(n)}]$ satisfies $BC = I$. Hence, $ABC = A$ giving $C = A$. Thus, we get $BA = I$.

Exchanging the roles of A and B in the above paragraph, we see that $BA = I$ implies $AB = I$. □

THEOREM 7.4. *Let A be a square matrix of order n. If B and C are square matrices of order n such that $AB = I$ and $AC = I$, then $B = C$.*

Proof. Suppose $AB = I = AC$. Then we have $B = BAC$. But, by Theorem 7.3, $BA = I$. Hence, we get $B = C$. □

Exercise 7.5. *Give examples of rectangular matrices A and B of order $m \times n$ and $n \times m$ respectively such that $AB = I_m$ but $BA \neq I_n$.*

We say that a square matrix A of order n is **invertible** if there exists a a square matrix B of order n such that $AB = I$. By Theorems 7.3 and 7.4, we know that if A is invertible, then there exists a unique B such that which also satisfies $BA = I$. This B is called the **inverse** of A and it is denoted by A^{-1}.

Once we have the inverse of a matrix A, then the system $Ax = b$ is uniquely solvable with solution x given by $x := A^{-1}b$.

THEOREM 7.6. *Suppose A and B are square matrices of order n. Then AB is invertible if and only if both A and B are invertible, and in that case*

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

Proof. Suppose AB is invertible. Then there exists C such that $(AB)C = I = C(AB)$. Thus, $A(BC) = I = (CA)B$. Hence, by Theorem 7.3, both A and B are invertible, and hence we have

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$$

so that $(AB)^{-1} = B^{-1}A^{-1}$.

Conversely, if both A and B are invertible, then the equality $(AB)(B^{-1}A^{-1}) = I$ shows that AB is invertible. □

Exercise 7.7. *Suppose A and B are matrices of order $m \times n$ and $n \times m$ respectively. If AB is invertible, then show that the equation $Ax = b$ is solvable for every $b \in V_m$.*

7.1 Inverse Using Elementary Row Operations

Now, we discuss briefly how elementary row operations can be used to obtain the inverse of an invertible matrix.

Recall that the set of solutions of a system of equations represented by matrix equation $Ax = b$ is same that of another system $\widetilde{A}x = \tilde{b}$ if $[\widetilde{A}; \tilde{b}]$ is obtained from $[A; b]$ by elementary row operations. Suppose we are able to perform row operations in such a way that $[\widetilde{A}; \tilde{b}] = [I; \tilde{b}]$. Then it is obvious that $x := \tilde{b}$ is a solution of $Ax = b$. Now, if we are able to reduce $[A; I] := [A; e_1, e_2, \ldots, e_n]$ into the form $[I; \tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n]$ by elementary row operations, then we would get $A\tilde{e}_1 = e_1$, $A\tilde{e}_2 = e_2$, $\ldots$, $A\tilde{e}_n = e_n$. Thus,

$$A[\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n] = [e_1, e_2, \ldots, e_n] = I,$$

and hence $B := [\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n]$ is the inverse of A.

Let us check whether $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ has an inverse. Applying row operations on $[A; I]$ we get the following:

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{array}{c} (R_2 \to R_2 + R_1) \\ (R_3 \to R_3 + R_2) \end{array} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 & 1 \end{bmatrix}$$

$$(R_1 \to R_1 - R_2) \begin{bmatrix} 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 & 1 \end{bmatrix} (R_3 \to 2R_3 - R_2) \begin{bmatrix} 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & 0 & 2 & -1 & 1 & 2 \end{bmatrix}$$

$$(R_2 \to R_2 - R_3) \begin{bmatrix} 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 2 & 0 & 2 & 0 & -2 \\ 0 & 0 & 2 & -1 & 1 & 2 \end{bmatrix} \begin{array}{l} (R_1 \to 2R_1 + R_3) \\ (R_2 \to \frac{1}{2}R_2) \\ (R_3 \to \frac{1}{2}R_3) \end{array} \begin{bmatrix} 2 & 0 & 0 & -1 & -1 & 2 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1/2 & 1/2 & 1 \end{bmatrix}$$

$$(R_1 \to \frac{1}{2}R_1) \begin{bmatrix} 1 & 0 & 0 & -1/2 & -1/2 & 1 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1/2 & 1/2 & 1 \end{bmatrix}$$

Thus inverse of A exists and it is given by

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

8 Determinants and Inverse

Consider a system of two equations in two unknowns,

$$\begin{aligned} ax + by &= \alpha && (8.10)\\ cx + dy &= \beta. && (8.11) \end{aligned}$$

On multiplying (8.10) by d and (8.11) by b, we get

$$\begin{aligned} adx + bdy &= d\alpha && (8.12)\\ bcx + bdy &= b\beta. && (8.13) \end{aligned}$$

Subtracting (8.13) from (8.12) we obtain

$$(ad - bc)x = d\alpha - b\beta. \tag{8.14}$$

Similarly (8.10) and (8.11) lead to

$$acx + bcy = c\alpha, \quad acx + ady = b\beta$$

so that we get

$$(ad - bc)y = b\beta - c\alpha. \tag{8.15}$$

Equations (8.14) and (8.15) show that the system of equations (8.10) and (8.11) is solvable for any vector on the right hand sides if and only if $ad - bc \neq 0$. The quantity $ad - bc$ is called the determinant of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Formally, the **determinant** of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is defined as the number

$$det(A) := ad - bc.$$

Now, knowing the determinants of square matrices of order $n - 1$, the **determinant** of a square matrix $A := (a_{ij})_{n\times n}$ of order n is defined as the number

$$det(A) := a_{11}M_{11} - a_{12}M_{12} + \ldots + (-1)^{1+n}a_{1n}M_{1n}$$

where M_{ij} denotes the determinant of the square matrix of order $n-1$ obtained by deleting the i^{th} row and j^{th} column of A. Thus,

$$det(A) := \sum_{j=1}^{n} (-1)^{1+j} a_{1j} M_{1j}.$$

Determinant of

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad \text{is also denoted by} \quad \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

For example,

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

Thus, if $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{vmatrix} &= 1 \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} - 2 \begin{vmatrix} -1 & 1 \\ 1 & 1 \end{vmatrix} + 1 \begin{vmatrix} -1 & 0 \\ 1 & 1 \end{vmatrix} \\ &= 1(-1) - 2(-2) + 1(-1) = 2. \end{aligned}$$

The following can be shown.
(i) For each $i \in \{1, \dots, n\}$, $det(A) := \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}$.
(ii) For each $j \in \{1, \dots, n\}$, $det(A) := \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}$.
An important relation that $det(A)$ satisfies is

$$A.adj(A) = det(A)I,$$

where $adj(A)$, called the **adjugate** of the matrix A, is the matrix (A_{ij}) with $A_{ij} := (-1)^{i+j} M_{ij}$. The above relation shows that

- a square matrix A is invertible if and only if $det(A) \neq 0$.

Exercise 8.1. *For a square matrix A,*
(i) $det(A) = 0$ if and only if columns of A are linearly dependent,
(ii) $det(A) = 0$ if and only if rows of A are linearly dependent.
–Why?

9 Eigenvalues and Eigenvectors

In this section we assume that the set $\mathbb{K}$ of scalars is the set of all complex numbers.

Let A be a square matrix of order n. Then a scalar λ is called an **eigenvalue** of A if there exists a nonzero vector $x \in V_n$ such that

$$Ax = \lambda x,$$

and in that case such a nonzero vector x is called an **eigenvector** of A corresponding to the eigenvalue λ. Thus, a scalar λ is an eigenvalue of A if and only if the subspace

$$N(A - \lambda I) := \{x \in V_n : (A - \lambda I)x = \{0\}$$

is not the zero space. If λ is an eigenvalue of A, then the subspace $N(A - \lambda I)$ is called the **eigen space** of A corresponding to the eigenvalue λ. Clearly, all nonzero vectors in $N(A-\lambda I)$ are eigenvectors of A corresponding to the eigenvalue λ. Thus, linear combination of eigenvectors of A corresponding to an eigenvalue λ are eigenvectors of A corresponding to the same eigenvalue λ.

Note that, for a scalar λ, the equation $Ax = \lambda x$ has a nonzero solution if and only if the homogeneous equation

$$(A - \lambda I)x = 0$$

has a nonzero solution, if and only if $A - \lambda I$ is not invertible, if and only if

$$det(A - \lambda I) = 0.$$

Thus, eigenvalues of A are those λ for which $det(A - \lambda I) = 0$.

It can be seen that if A is a square matrix of order n, then

$$p_A(\lambda) := det(A - \lambda I)$$

is a polynomial of degree n, called the **characteristic polynomial** of A. It is a well-known fact (but, you must not have seen its proof!) that

Every polynomial with real or complex coefficients has at least one zero.

The above result is known as **fundamental theorem of algebra**. Hence, for a polynomial of degree $n \geq 1$, say $p(z) := a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n$, there exist complex numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$p(z) = a_n(z - \lambda_1)(z - \lambda_2)\ldots(z - \lambda_n).$$

In particular, the characteristic $p_A(z) := det(A - zI)$ of an $n \times n$ square matrix A can also be written in the above form, and in that case, the numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

If λ is a zero of $det(A - zI) = 0$ of order m, that is, $det(A - zI) = (z - \lambda)^m q(z)$ with $q(\lambda) \neq 0$, then we say that λ is an eigenvalue of **algebraic multiplicity** m. If algebraic multiplicity of an eigenvalue λ is 1, then λ is said to be a **simple eigenvalue**. The dimension of the subspace $N(A - \lambda I)$ is called the **geometric multiplicity** of λ.

Example 9.1. *If A is a diagonal matrix or an upper triangular matrix or lower triangular matrix, then the diagonal entries $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. In case A is a diagonal matrix, then for each $i \in \{1, \ldots, n\}$, e_i is an eigenvector corresponding to the eigenvalue λ_i.*

Example 9.2. *Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Then 1 is the only eigenvalue of A. The vectors $u := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $v := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ are linearly independent eigenvectors of A corresponding to the eigenvalue 1.*

Example 9.3. *Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. In this case the characteristic polynomial is given by*

$$p_A(\lambda) = (1 - \lambda)^2 - 1 = \lambda^2 - 2\lambda.$$

Hence, 0 and 2 are the eigenvalues of A. It is seen that $u := [1, -1]^T$ and $v := [1, 1]^T$ are eigenvectors corresponding to the eigenvalues 0 and 2 respectively.

Example 9.4. *Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. In this case the characteristic polynomial is given by*

$$p_A(\lambda) = \lambda^2 + 1.$$

Hence, i and $-i$ are the eigenvalues of A. It is seen that $u : [1, i]^T$ and $v := [1, -i]^T$ are eigenvectors corresponding to the eigenvalues i and $-i$ respectively.

Example 9.5. *Let $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$. In this case the characteristic polynomial is given by*

$$p_A(\lambda) = (1 - \lambda)^2 - 1.$$

Hence, 0 and 2 are the eigenvalues of A. It is seen that $u := [1, i]^T$ and $v := [1, -i]^T$ are eigenvectors corresponding to the eigenvalues 0 and 2 respectively.

THEOREM 9.6. *Suppose $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ respectively. Then $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ are linearly independent.*

Proof. Let $\alpha_1, \dots, \alpha_k$ be scalars such that

$$\alpha_1 u^{(1)} + \alpha_2 u^{(2)} + \ldots + \alpha_k u^{(k)} = 0. \tag{9.16}$$

Then applying A to (9.16),

$$\alpha_1 \lambda_1 u^{(1)} + \alpha_2 \lambda_2 u^{(2)} + \ldots + \alpha_k \lambda_k u^{(k)} = 0. \tag{9.17}$$

and multiplying (9.16) by λ_1,

$$\alpha_1 \lambda_1 u^{(1)} + \alpha_2 \lambda_1 u^{(2)} + \ldots + \alpha_k \lambda_1 u^{(k)} = 0. \tag{9.18}$$

Now, subtracting (9.18) from (9.17), we have

$$\alpha_2(\lambda_2 - \lambda_1) u^{(2)} + \alpha_3(\lambda_3 - \lambda_1) u^{(3)} + \ldots + \alpha_k(\lambda_k - \lambda_1) u^{(k)} = 0. \tag{9.19}$$

Next, let us apply A to (9.19) first, and then multiply (9.19) by λ_2 to get

$$\alpha_2(\lambda_2 - \lambda_1)\lambda_2 u^{(2)} + \alpha_3(\lambda_3 - \lambda_1)\lambda_3 u^{(3)} + \ldots + \alpha_k(\lambda_k - \lambda_1)\lambda_k u^{(k)} = 0, \tag{9.20}$$

$$\alpha_2(\lambda_2 - \lambda_1)\lambda_2 u^{(2)} + \alpha_3(\lambda_3 - \lambda_1)\lambda_2 u^{(3)} + \ldots + \alpha_k(\lambda_k - \lambda_1)\lambda_2 u^{(k)} = 0. \tag{9.21}$$

Subtracting (9.21) from (9.20), we have

$$\alpha_3(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2) u^{(3)} + \ldots + \alpha_k(\lambda_k - \lambda_1)(\lambda_k - \lambda_2) u^{(k)} = 0.$$

Continuing this, at the $(k-1)^{th}$ step, we arrive at

$$\alpha_k(\lambda_k - \lambda_1)(\lambda_k - \lambda_2)(\lambda_k - \lambda_3)\ldots(\lambda_k - \lambda_{k-1}) u^{(k)} = 0.$$

Since $\lambda_k - \lambda_j \neq 0$ for $j = 1, 2, \ldots, k-1$, it follows that $\alpha_k = 0$. Hence, from the previous step, we get $\alpha_{k-1} = 0$. Tracing backwards, we get $\alpha_1 = 0$ for every $j = 1, 2, \ldots, k$. Thus, $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ are linearly independent. □

Suppose A has linearly independent eigenvectors $u^{(1)}, u^{(2)}, \ldots, u^{(n)}$ and suppose that $\lambda_1, \ldots, \lambda_n$ are the eigenvalues associated with these eigenvectors, that is, $Au^{(j)} = \lambda_j u^{(j)}$ for $j = 1, \ldots, n$. Then we see that

$$A[u^{(1)}, u^{(2)}, \ldots, u^{(n)}] = [\lambda_1 u^{(1)}, \lambda_2 u^{(2)}, \ldots, \lambda_n u^{(n)}] = [u^{(1)}, u^{(2)}, \ldots, u^{(n)}]D$$

where D is the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$.

Since columns of $U := [u^{(1)}, u^{(2)}, \ldots, u^{(n)}]$ are linearly independent, it follows that U is invertible. Thus, we have proved the following theorem.

THEOREM 9.7. *Suppose A is an $n \times n$ matrix having n linearly independent eigenvectors $u^{(1)}, u^{(2)}, \ldots, u^{(n)}$. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues associated with these eigenvectors, that is, $Au^{(j)} = \lambda_j u^{(j)}$ for $j = 1, \ldots, n$, then with $U := [u^{(1)}, u^{(2)}, \ldots, u^{(n)}]$ we have*

$$U^{-1}AU = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

In view of Theorems 9.6 and 9.7, we can state the following.

THEOREM 9.8. *Suppose A is an $n \times n$ matrix having n distinct eigenvalues. Then there exists an invertible matrix U such that $U^{-1}AU$ is a diagonal matrix.*

PROPOSITION 9.9. *If A is a square matrix with real entries and if λ is a real eigenvalue of A, then there is an eigenvector with real entries corresponding to the eigenvalue λ.*

Proof. Suppose λ is a real eigenvalue of A with a corresponding eigenvector u. If u has complex entries, then we can write u as $u = v + iw$ where v and w are vectors with real entries. Then we have

$$Av + iAw = A(v + iw) = \lambda(v + iw) = \lambda v + i\lambda w.$$

Hence, it follows that $Av = \lambda v$ and $Aw = \lambda w$. Since $u \neq 0$, at least one of v and w is a nonzero vector. Thus, corresponding to the eigenvalue λ, A has an eigenvector with real entries. □

9.1 Hermitian and Unitary Matrices

Let $A = (a_{ij})_{n\times n}$. Then the **conjugate transpose** of A is the matrix $A^* := (b_{ij})$ with $b_{ij} := \bar{a}_{ij}$. Here, $\bar{\alpha}$ is the complex conjugate of the scalar α. Thus,

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \Longrightarrow A^* := \begin{bmatrix} \bar{a}_{11} & \bar{a}_{21} & \ldots & \bar{a}_{m1} \\ \bar{a}_{12} & \bar{a}_{22} & \ldots & \bar{a}_{m2} \\ \ldots & \ldots & \ldots & \ldots \\ \bar{a}_{1n} & \bar{a}_{2n} & \ldots & \bar{a}_{mn} \end{bmatrix}.$$

In particular, if $u := [u_1, u_2, \ldots, u_n]^T \in V_n$, then

$$u^* = [\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_n].$$

We observe that for matrices A and B of appropriate orders and scalar α,

$$(A + B)^* = A^* + B^*, \quad (\alpha A)^* = \bar{\alpha}A^*, \quad (AB)^* = B^*A^*, \quad (A^*)^* = A.$$

Note that if A is a real matrix, that is, if the entries of A are real numbers, then $A^* = A^T$.

We observe that, for vectors u and v in V_n, u^*v is a scalar and vu^* is an $n \times n$ matrix of rank at most 1. Also, if $u := [u_1, u_2, \ldots, u_n]^T \in V_n$, then

$$u^*u = \bar{u}_1 u_1 + \bar{u}_2 u_2 + \ldots + \bar{u}_n u_n = \sqrt{|u_1|^2 + |u_2|^2 + \ldots + |u_n|^2},$$

a non-negative real number, and it is denoted by $\|u\|$, called the **Euclidean norm** of u.

A vector $u \in V_n$ is said to be a **unit vector** if $\|u\| = 1$.

Vectors u and v in V_n are said to be **orthogonal** to each other, if $u^*v = 0$. A subset S of V_n is said to be an **orthogonal set** if every pair of (distinct) vectors in S are orthogonal to each other. An orthogonal set S is said to be an **orthonormal set** if every vector in S is a unit vector.

A square matrix A is said to be a
(i) **hermitian matrix** if $A^* = A$.
(ii) **unitary matrix** if $A^*A = I = AA^*$.
(iii) A unitary matrix with real entries is called an **orthogonal matrix**.

Note that a hermitian matrix with real entries is a **symmetric matrix**.
Clearly, if A is a unitary matrix, then A is invertible and $A^{-1} = A^*$. If A is a hermitian matrix, then for every $x \in V_n$, we have

$$(x^*Ax)^* = x^*Ax$$

so that x^*Ax is a real number.

If A is a unitary matrix, then we see that columns of A form an orthonormal set. Indeed, if A is with columns $A_1, A_2, \ldots, A_n$, that is, $A = [A_1\, A_2\, \ldots\, A_n]$ with $A_j \in V_n$ for $j = 1, \ldots, n$, then

$$A^* = \begin{bmatrix} A_1^* \\ A_2^* \\ \cdots \\ A_n^* \end{bmatrix}, \quad A^*A = \begin{bmatrix} A_1^*A_1 & A_1^*A_2 & \ldots & A_1^*A_n \\ A_2^*A_1 & A_2^*A_2 & \ldots & A_2^*A_n \\ \cdots & \cdots & \cdots & \cdots \\ A_n^*A_1 & A_n^*A_2 & \ldots & A_n^*A_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

so that $A_i^*A_i = 1$ for all i and $A_i^*A_j = 1$ for all $i \neq j$.

PROPOSITION 9.10. *If $S := \{u^{(1)}, u^{(2)}, \ldots, u^{(k)}\}$ is an orthogonal set which does not contain the zero vector, then S is a linearly independent set. In particular, every orthonormal set is linearly independent.*

THEOREM 9.11. *(i) Eigenvalues of a hermitian matrix are real numbers.*
(ii) Eigenvalues of a unitary matrix are with absolute value 1.

Proof. (i) Suppose A is a hermitian matrix. Let λ be such that $Au = \lambda u$ with $u \neq 0$. Then $u^*Au = \lambda u^*u$. Since both u^*Au and u^*u are reals, it follows that λ is also a real number.

(ii) Suppose A is a unitary matrix. Let λ be such that $Au = \lambda u$ with $u \neq 0$. Then, we have $u^*A^* = \bar{\lambda}u^*$ so that $u^*A^*Au = \bar{\lambda}u^*Au = \bar{\lambda}\lambda u^*u$. But, $A^*A = I$. Hence, we have $u^*u = \bar{\lambda}\lambda u^*u = |\lambda|^2 u^*u$ so that $|\lambda| = 1$. □

THEOREM 9.12. *Suppose A is hermitian matrix. If λ_1 and λ_2 are distinct eigenvalues of A with corresponding eigenvectors u and v respectively. Then u and v are orthogonal to each other.*

Proof. Suppose $Au = \lambda_1 u$ and $Av = \lambda_2 v$. Then we have $v^*A^* = \bar{\lambda}_2 v^*$. But, $A^* = A$ and $\bar{\lambda}_2 = \lambda_2$ (by (i)) so that we have $v^*A = \lambda_2 v^*$. Hence,

$$\lambda_1 v^*u = v^*Au = \lambda_2 v^*u.$$

Thus,

$$(\lambda_1 - \lambda_2)v^*u = 0.$$

From this, it follows that if $\lambda_1 \neq \lambda_2$, then u and v are orthogonal to each other. □

Remark 9.13. *If A is a real symmetric matrix, then in view of Proposition 9.9, the vectors u and v in Theorem 9.12 can be taken to be with real entries.*

COROLLARY 9.14. *Suppose A is a hermitian matrix. If A has n distinct eigenvalues, then there exists a unitary matrix U such that U^*AU is a diagonal matrix with eigenvalues of A as diagonal entries.*

COROLLARY 9.15. *Suppose A is a real symmetric matrix. If A has n distinct eigenvalues, then there exists an (real) orthogonal matrix Q such that Q^TAQ is a diagonal matrix with eigenvalues of A as diagonal entries.*

We state the following two theorems without proof.

THEOREM 9.16. *Every hermitian matrix of order n has n linearly independent eigenvectors.*

THEOREM 9.17. *Every real symmetric matrix of order n has n linearly independent eigenvectors with real entries.*

COROLLARY 9.18. *Suppose A is a hermitian matrix. Then there exists a unitary matrix U such that U^*AU a diagonal matrix with eigenvalues of A as diagonal entries.*

COROLLARY 9.19. *Suppose A is a real symmetric matrix. Then there exists an (real) orthogonal matrix Q such that Q^TAQ is a diagonal matrix with eigenvalues of A as diagonal entries.*

Exercise 9.20. *1. Give proof for Proposition 9.10, Corollaries 9.14, 9.15, 9.18, 9.19*

2. Justify the statements in the following, where A is a square matrix.

(a) A scalar λ is an eigenvalue of A iff λ is an eigenvalue of A^T.
If your justification involves determinants, give a determinant-free justification.

(b) A scalar λ is an eigenvalue of A iff $\bar{\lambda}$ is an eigenvalue of A^*.

(c) If A is a real matrix, then λ is an eigenvalue of A iff $\bar{\lambda}$ is an eigenvalue of A^*.

(d) If A is a real matrix of odd degree, then A has at least one real eigenvalue.