Chapter 3

Topics in Linear Algebra

3.1 Introduction

Given an $n \times n$ matrix there are several decompositions, that one comes across in the theory of matrices, that are derived in an attempt to bring the matrix to a simpler level to make the analysis easier. We shall be dealing with only real or complex matrices. It is often convenient to treat a real matrix as a complex matrix. Some of the commonly used decompositions are

1. LU decomposition,
2. Triangular Decomposition (such as Schur decomposition, Jordan form),
3. diagonalization,
4. Tridiagonalization,
5. reduction to Hessenberg form, etc.

Each one has its advantages as well as disadvantages. There are some like diagonalization which are not possible for all $n \times n$ matrices. All these are applicable only for square matrices. However, one decomposition, which is a very comprehensive decomposition applicable for all matrices - whether square or rectangular - is widely used in several modern applications. This is known as the "**Singular Value Decomposition**" (written in short as SVD).

We shall, in this chapter discuss only the diagonalization and triangularization of an $n \times n$ complex matrix and SVD of a general $m \times n$ real and matrix.

3.2 Basic Notions in Linear Algebra

In this section we shall review some of the basic notions in Linear Algebra.

1. **Vector Space**
   Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. While analysing the linear system $Ax = b$ we find that we have to deal with two vectors, namely the known vector b given in $\mathbb{R}^m$ and the unknown vector x to be found in $\mathbb{R}^n$. So, in general, we shall consider a positive integer k and look at the structure of $\mathbb{R}^k$, the collection of all $k \times 1$ column vectors.

$$\mathbb{R}^k = \left\{ x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} : x_j \in \mathbb{R},\ 1 \leq j \leq k \right\} \tag{3.2.1}$$

   We first consider the simple operation of "Addition" on $\mathbb{R}^k$ defined as follows:
   For $x, y \in \mathbb{R}^k$ we define

$$x + y = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_k + y_k \end{pmatrix} \tag{3.2.2}$$

   We easily see that addition has the following properties:

   (a) $x, y \in \mathbb{R}^k \Longrightarrow x + y \in \mathbb{R}^k$

   (b) $x + y = y + x$ for all $x, y \in \mathbb{R}^k$

   (c) $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbb{R}^k$

(d) The vector

$$\theta_k = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{k\times 1}$$

is in $\mathbb{R}^k$ and is such that

$$x + \theta_k = x = \theta_k + x \text{ for all } x \in \mathbb{R}^k$$

(e) For every $x \in \mathbb{R}^k$ the vector

$$(-x) = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_k \end{pmatrix} \text{ is in } \mathbb{R}^k \text{ and is such that}$$

$$x + (-x) = \theta_k = (-x) + x$$

We next consider the operation of multiplying a vector in $\mathbb{R}^k$ by a scalar $\alpha \in \mathbb{R}$, defined as follows:

$$\alpha \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_k \end{pmatrix} \tag{3.2.3}$$

It is easy to see that this operation, called scalar multiplication, has the following properties:

(a) $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^k \Longrightarrow \alpha \cdot x \in \mathbb{R}^k$

(b) $\alpha \in \mathbb{R}$ and $x, y \in \mathbb{R}^k \Longrightarrow \alpha \cdot (x + y) = (\alpha \cdot x) + (\alpha \cdot y)$

(c) $\alpha,\ \beta \in \mathbb{R}$ and $x \in \mathbb{R}^k \Longrightarrow (\alpha + \beta) \cdot x = (\alpha \cdot x) + (\beta \cdot x)$

(d) $\alpha,\ \beta \in \mathbb{R}$ and $x \in \mathbb{R}^k \Longrightarrow (\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$

(e) $1 \cdot x = x$ for all $x \in \mathbb{R}^k$

The important ingredients we discussed above in $\mathbb{R}^k$ are the two basic operations of addition and scalar multiplication and their properties. Any system which has such a structure is called a "Vector Space". More precisely we have the following definition of a vector space over $\mathbb{R}$.

Definition 3.2.1 Let $\mathbb{C}^n$ be any nonempty set and let $+$ be a binary opertaion on V and $\cdot$ be a rule of combining an element of $\mathbb{R}$ and an element of $\mathbb{C}^n$, (called "scalar multiplication") such that the following properties are satisfied:

(Axioms for the operation $+$ on $\mathbb{C}^n$)

(a) $x, y \in \mathbb{C}^n \Longrightarrow x + y \in \mathbb{C}^n$

(b) $x + y = y + x$ for all $x, y \in \mathbb{C}^n$

(c) $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbb{C}^n$

(d) There exists a vector $\theta_{\mathbb{C}^n} \in \mathbb{C}^n$ such that

$$x + \theta_{\mathbb{C}^n} = x = \theta_{\mathbb{C}^n} + x \text{ for all } x \in \mathbb{C}^n$$

(e) For every $x \in \mathbb{C}^n$ there exists a vector in $\mathbb{C}^n$, which we denote by $(-x)$, such that

$$x + (-x) = \theta_{\mathbb{C}^n} = (-x) + x$$

(Axioms for scalar multiplication)

(f) $x \in \mathbb{C}^n$ and $\alpha \in \mathbb{R} \Longrightarrow \alpha \cdot x \in \mathbb{C}^n$

(g) $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$ for all $\alpha \in \mathbb{R}$ and for all $x, y \in \mathbb{C}^n$

(h) $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$ for all $\alpha, \beta \in \mathbb{R}$ and for all $x \in \mathbb{C}^n$

(i) $(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$ for all $\alpha, \beta \in \mathbb{R}$ and for all $x \in \mathbb{C}^n$

(j) $1x = x$ for all $x \in \mathbb{C}^n$

Then we say that $\mathbb{C}^n$ is **vector space over** $\mathbb{R}$ (with these operations of addition and scalar multiplication)

The important ingredients we discussed above in $\mathbb{R}^k$ are the two basic operations of addition and scalar multiplication and their properties. Any system which has such a structure is called a "Vector Space". More precisely we have the following definition of a vector space over $\mathbb{R}$.

Definition 3.2.2 Let $\mathbb{C}^n$ be any nonempty set and let $+$ be a binary opertaion on V and $\cdot$ be a rule of combining an element of $\mathbb{R}$ and an element of $\mathbb{C}^n$, (called "scalar multiplication") such that the following properties are satisfied:

(Axioms for the operation + on $\mathbb{C}^n$)

(a) $x, y \in \mathbb{C}^n \Longrightarrow x + y \in \mathbb{C}^n$

(b) $x + y = y + x$ for all $x, y \in \mathbb{C}^n$

(c) $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbb{C}^n$

(d) There exists a vector $\theta_{\mathbb{C}^n} \in \mathbb{C}^n$ such that

$$x + \theta_{\mathbb{C}^n} = x = \theta_{\mathbb{C}^n} + x \text{ for all } x \in \mathbb{C}^n$$

(e) For every $x \in \mathbb{C}^n$ there exists a vector in $\mathbb{C}^n$, which we denote by $(-x)$, such that

$$x + (-x) = \theta_{\mathbb{C}^n} = (-x) + x$$

(Axioms for scalar multiplication)

(f) $x \in \mathbb{C}^n$ and $\alpha \in \mathbb{R} \Longrightarrow \alpha \cdot x \in \mathbb{C}^n$

(g) $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$ for all $\alpha \in \mathbb{R}$ and for all $x, y \in \mathbb{C}^n$

(h) $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$ for all $\alpha, \beta \in \mathbb{R}$ and for all $x \in \mathbb{C}^n$

(i) $(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$ for all $\alpha, \beta \in \mathbb{R}$ and for all $x \in \mathbb{C}^n$

(j) $1x = x$ for all $x \in \mathbb{C}^n$

Then we say that $\mathbb{C}^n$ is **vector space over** $\mathbb{R}$ (with these operations of addition and scalar multiplication)

Remark 3.2.1 In the above definition, if we replace $\mathbb{R}$ by any field $\mathbb{F}$, then we get the notion of a vector space over $\mathbb{F}$. We shall be basically concerned about vector spaces over $\mathbb{R}$ and vector spaces over $\mathbb{C}$. In particular, we shall be dealing with the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$.

Remark 3.2.2 From the above axioms the following properties in a vector space $\mathbb{C}^n$ follow:

(a) For any $x \in \mathbb{C}^n$ we have $0x = \theta_n$.
This can be seen using the axioms of $+$ on $\mathbb{C}^n$ and those of scalar multiplication, as follows:

$$\begin{array}{rcl} 0x & = & (0+0)x \\ & = & 0x + 0x \\ \Longrightarrow & & \\ (-0x) + 0x & = & (-0x) + (0x + 0x) \\ \Longrightarrow & & \\ \theta_n & = & (-0x + 0x) + (0x) \\ \Longrightarrow & & \\ \theta_n & = & \theta_n + 0x \\ \Longrightarrow & & \\ \theta_n & = & 0x \end{array}$$

(b) Similarly we can show that

$$(-1)x = (-x) \text{ for all } x \in \mathbb{C}^n$$

Remark 3.2.3 We shall call elements of a vector space as vectors.

2. **Subspaces**
The basic idea in analysing a vector space is to break it up into smaller parts each of which is self contained with respect to the two basic operations of addition and scalar multiplication of the underlying vector space. This leads us to the notion of subspaces.

Definition 3.2.3 A nonempty subset $\mathcal{W}$ of $\mathbb{C}^n$ is said to be a **subspace** of $\mathbb{C}^n$ if

$$x, y \in \mathcal{W} \implies x + y \in \mathcal{W} \tag{3.2.4}$$

$$x \in \mathcal{W} \text{ and } \alpha \in \mathbb{C} \implies \alpha x \in \mathcal{W} \tag{3.2.5}$$

Similarly,
A nonempty subset $\mathcal{W}$ of $\mathbb{R}^n$ is said to be a **subspace** of $\mathbb{R}^n$ if

$$x, y \in \mathcal{W} \implies x + y \in \mathcal{W} \tag{3.2.6}$$

$$x \in \mathcal{W} \text{ and } \alpha \in \mathbb{R} \implies \alpha x \in \mathcal{W} \tag{3.2.7}$$

We now make some simple observations:

Remark 3.2.4 What the above definition says is that when we perform the two basic operations of the vector space with the $\mathcal{W}$ vectors then the resultant vectors are also $\mathcal{W}$ vectors. The main idea in analysing these vector spaces is to break the vector space into smaller subspaces, in a suitable manner, and analyse the problem in each subspace and then put all these together to get to the final analysis on the whole space.

Remark 3.2.5 We have

$$\begin{aligned} \mathcal{W} \text{ is a subspace of } \mathbb{C}^n &\implies \mathcal{W} \text{ is nonempty} \\ &\implies \exists w \in \mathcal{W} \\ &\implies 0w \in \mathcal{W} \\ &\implies \theta_n \in \mathcal{W} \end{aligned}$$

Thus we see that the zero vector belongs to every subspace of $\mathbb{C}^n$. Similarly the zero vector belongs to every subspace of $\mathbb{R}^n$. Thus a nonempty subset will fail to be a subspace if it does not contain the zero vector. However, if a nonempty subset contains the zero vector it does not automatically assure that it is a subspace.

Remark 3.2.6 $\mathbb{C}^n$ is itself a subspace of $\mathbb{C}^n$ and $\mathbb{R}^n$ is itself a subspace of $\mathbb{R}^n$

Remark 3.2.7 $\mathbb{C}^n$ is the "largest" subspace of $\mathbb{C}^n$ and $\mathcal{W} = \{\theta_n\}$ is the smallest subspace of $\mathbb{C}^n$ and is also the smallest subspace of $\mathbb{R}^n$.

Example 3.2.1 Consider $\mathbb{C}^n = \mathbb{R}^3$.

(a) Let $\mathcal{W}$ be the subset of $\mathbb{R}^3$ defined as

$$\mathcal{W} = \left\{ x = \begin{pmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

Geometrically speaking, this is the plane $z = x + y$ in $\mathbb{R}^3$. It is easy to verify that this is a subspace of $\mathbb{R}^3$. We verify this fact as follows:

i. Since

$$\theta_3 = \begin{pmatrix} 0 \\ 0 \\ 0 + 0 \end{pmatrix}$$

it follows that $\theta_3 \in \mathcal{W}$ and hence $\mathcal{W}$ is nonempty

ii. We have

$$\begin{aligned} x, y \in \mathcal{W} \implies & x = \begin{pmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{pmatrix}, \; y = \begin{pmatrix} y_1 \\ y_2 \\ y_1 + y_2 \end{pmatrix}, \; x_j, y_j \in \mathbb{R}, \; 1 \leq j \leq 2 \\ \implies & x + y = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ (x_1 + y_1) + (x_2 + y_2) \end{pmatrix} \\ = & \begin{pmatrix} \alpha \\ \beta \\ \alpha + \beta \end{pmatrix} \text{ where } \alpha = x_1 + x_2, \; \beta = y_1 + y_2 \in \mathbb{R} \\ \implies & x + y \in \mathcal{W} \end{aligned}$$

iii. Further,

$$x \in \mathcal{W}, \alpha \in \mathbb{R} \implies x = \begin{pmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{pmatrix}, \alpha \in \mathbb{R}$$

$$\begin{aligned}
\implies \quad \alpha x &= \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha(x_1 + x_2) \end{pmatrix} \\
&= \begin{pmatrix} a \\ b \\ a+b \end{pmatrix} \text{ where } a = \alpha x_1, \ b = \alpha x_2 \in \mathbb{R} \\
\implies \quad \alpha x &\in \mathcal{W}
\end{aligned}$$

Thus we see that $\mathcal{W}$ is nonempty and is closed with respect to addition and scalar multiplication and hence $\mathcal{W}$ is a subspace of $\mathbb{R}^3$

(b) Let $\mathcal{W}$ be defined as

$$\mathcal{W} = \left\{ x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : 2x_1 - 3x_2 + x_3 = 0 \text{ and } 3x_1 - 4x_2 - x_3 = 0; x_1, x_2, x_3 \in \mathbb{R} \right\}$$

Once again it is easy to verify that this is a subspace of $\mathbb{R}^3$. Geometrically speaking, this subspace is the line of intersection of the two planes $2x - 3y + z = 0$ and $3x - 4y - z = 0$

Remark 3.2.8 In general, a subspace in $\mathbb{R}^3$ will be either $\mathbb{R}^3$ or $\{\theta_3\}$, or a plane through the origin or a line through the origin.

Example 3.2.2 Consider the vector space $\mathbb{R}^k$ (where we assume $k \geq 2$). Then we can easily verify that the following subsets are subspaces of $\mathbb{R}^n$:

(a) $\mathcal{W} = \left\{ x \in \mathbb{R}^k : x_1 = 0 \right\}$

(b) $\mathcal{W} = \left\{ x \in \mathbb{R}^k : x_k = 3x_1 \right\}$

Example 3.2.3 The following subset $\mathcal{W}$ of $\mathbb{R}^3$ is NOT a subspace of $\mathbb{R}^3$ (Why?):

$$\mathcal{W} \;=\; \left\{ x = \begin{pmatrix} a \\ b \\ 4 \end{pmatrix} : a, b \in \mathbb{R} \right\}$$

3. **Linear Independence**

We shall next introduce the notion of a linearly independent set. Consider a finite set of vectors

$$u_1, u_2, \cdots, u_r$$

in a vector space $\mathbb{C}^n$ (or $\mathbb{R}^n$). Any vector of the form

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r$$

where $\alpha_j \in \mathbb{C}$ (or $\mathbb{R}$), for $1 \leq j \leq r$ is called a linear combination of these given vectors. In particular,

$$0u_1 + 0u_2 + \cdots + 0u_r$$

is a linear combination of these vectors and is equal to θ_n. This linear combination is called the trivial linear combination of these vectors. Thus we find that given any finite set of vectors, we can obtain the zero vector θ_n, as a linear combination of these vectors.

Example 3.2.4 Consider the vector space $\mathbb{C}^n = \mathbb{R}^3$ and the set of vectors,

$$S = u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \; u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Then clearly we can write the zero vector θ_n as the trivial linear combination of these vectors as

$$\theta_n = 0u_1 + 0u_2$$

Further this is the only way we can express θ_n as a linear combination of u_1, u_2. For, if a linear combination gives θ_n, then we must have,

$$\begin{aligned} \alpha_1 u_1 + \alpha_2 u_2 &= \theta_n \\ &\Longrightarrow \\ \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \alpha_1 + \beta_1 \end{pmatrix} &= 0 \\ &\Longrightarrow \\ \alpha_1, \text{ and } \alpha_2 &= 0 \end{aligned}$$

On the other hand consider the set of vectors,

$$S = u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \ u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \ u_3 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

Then we have the trivial linear combination

$$\theta_n \quad = \quad 0u_1 + 0u_2 + 0u_3$$

We also have

$$1u_1 + 1u_2 + (-1)u_3 \quad = \quad \theta_n$$

In fact, for any $\alpha \in \mathbb{R}$ we have

$$\alpha u_1 + \alpha u_2 + (-\alpha)u_3 \quad = \quad \theta_n$$

Thus nontrivial linear combinations of u_1, u_2, u_3 also give rise to the zero vector.

From the above example it follows that given any finite subset S of a vector space $\mathbb{C}^n$, the following two possibilities arise:

(a) EITHER θ_n can be expressed ONLY as the trivial linear combination of the vectors in S,

(b) OR θ_n can also be expressed as a nontrivial linear combination of the vectors in S

We distinguish these two possibilities with the following definition:

Definition 3.2.4 A nonempty finite subset

$$\boldsymbol{S = u_1, u_2, \cdots, u_r}$$

of $\mathbb{C}^n$ is said to be linearly independent if

$$\boldsymbol{\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r = \theta_n \implies \alpha_j = 0, \ 1 \leq \jmath \leq r}$$

(that is, the only way to express the zero vector as a linear combination of the vectors in S is to express it as the trivial linear combination).
If S is not linearly independent it is said to be linearly dependent.

Remark 3.2.9 The set

$$S = u_1, u_2, \cdots, u_r$$

in $\mathbb{C}^n$ is linearly dependent means that there exist $\alpha_1, \alpha_2, \cdots, \alpha_r \in \mathbb{C}$, at least one of which is not zero, such that

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r = \theta_n$$

Remark 3.2.10 Any set that contains the zero vector is linearly dependent

4. <u>**Basis**</u>
We next introduce the notion of a basis.

Definition 3.2.5 A set of vectors $\boldsymbol{\mathcal{B}} = \boldsymbol{u_1, u_2, \cdots, u_k}$ is said to be a basis for a subspace $\boldsymbol{\mathcal{W}}$ of $\mathbb{C}^n$ if

(a) The vectors $\boldsymbol{u_1, u_2, \cdots, u_k}$ are all in $\boldsymbol{\mathcal{W}}$,

(b) The set of vectors $\boldsymbol{u_1, u_2, \cdots, u_k}$ is linearly independent, and

(c) Every vector in $\boldsymbol{\mathcal{W}}$ can be expressed as a linear combination of these vectors $\boldsymbol{u_1, u_2, \cdots, u_k}$

Remark 3.2.11 Not only every vector in $\mathcal{W}$ can be expressed as a linear combination of the basis vectors, but this representation is unique.

Remark 3.2.12 One can choose many bases for a subspace. The vectors in two different bases will be different. But the **number of vectors** in each basis will be the same. This number is called the "**dimension** of the subspace.

Example 3.2.5 Consider the subspace $\mathcal{W}$ of $\mathbb{R}^3$ defined as

$$\mathcal{W} \quad = \quad \left\{ x = \begin{pmatrix} \alpha \\ \beta \\ \alpha + 2\beta \end{pmatrix} : \alpha, \beta \in \mathbb{R} \right\}$$

Let $\mathcal{B}$ be the set of vectors

$$\mathcal{B} = \left\{ u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},\ u_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\}$$

We shall now see that this is a basis for $\mathcal{W}$.

(a) $u_1, u_2 \in \mathcal{W}$

(b) u_1, u_2 are linearly independent since

$$\begin{aligned} \alpha_1 u_1 + \alpha_2 u_2 = \theta_n &\implies \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_1 + 2\alpha_2 \end{pmatrix} = \theta_n = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ &\implies \alpha_1 = \alpha_2 = 0 \\ &\implies u_1, u_2 \text{ linearly independent} \end{aligned}$$

(c) Finally

$$\begin{aligned} x \in \mathcal{W} &\implies \exists\, \alpha, \beta \in \mathbb{R} \ni x = \begin{pmatrix} \alpha \\ \beta \\ \alpha + 2\beta \end{pmatrix} \\ &\implies x = \alpha u_1 + \beta u_2 \\ &\implies x \text{ is a linear combination of } u_1, u_2 \end{aligned}$$

Hence $\mathcal{B}$ is a basis for $\mathcal{W}$. Since there are two vectors in the basis the dimension of $\mathcal{W}$ is 2.
Note that the set of vectors

$$\mathcal{B}_1 = \left\{ v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},\ v_2 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \right\}$$

is also a basis for $\mathcal{W}$, But the set of vectors

$$\mathcal{B}_3 = \left\{ u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},\ u_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, u_3 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \right\}$$

is not a basis for $\mathcal{W}$ even though these vectors are in $\mathcal{W}$ and every evctor in $\mathcal{W}$ is a linear combination of these vectors. This is because this set is not linearly independent.

5. **Some Special Subspaces**
   We shall now look at some special types of subspaces.

   (a) Let $\mathcal{S} = \{u_1, u_2, \cdots, u_k\}$ be a set of vectors in $\mathbb{C}^n$. Then the set of all linear combinations of these vectors is a subspace of $\mathbb{C}^n$ and this is called the subspace spanned by $\mathcal{S}$ and is denoted by $\mathcal{L}[\mathcal{S}]$ or $\mathcal{L}[u_1, u_2, \cdots, u_k]$.

   (b) In particular, let $A \in \mathbb{C}^{m\times n}$ be a complex $n \times n$ matrix. Let $A = (a_{ij})_{m\times n}$. The jth column vector is

$$C_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \end{pmatrix}$$

   This is a vector in $\mathbb{C}^m$. Now consider the set of all column vectors

$$C_1, C_2, \cdots, C_n$$

   The subspace spanned by these vectors, namely $\mathcal{L}[C_1, C_2, \cdots, C_n]$ is called the "**Column Space**" of the matrix A. This subspace is the set of all linear combinations of $C_1, C_2, \cdots, C_n$. This is a subspace of $\mathbb{C}^m$. This subspace is denoted by $Col(A)$,

   (c) Similarly the ith row vector (written in the column vector format), is

$$R_j = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{in} \end{pmatrix}$$

   Now consider the set of all row vectors

$$R_1, R_2, \cdots, R_n$$

The subspace spanned by these vectors, namely $\mathcal{L}[R_1, R_2, \cdots, R_n]$ is called the "**Row Space**" of the matrix A. This subspace is the set of all linear combinations of $R_1, R_2, \cdots, R_n$. This is a subspace of $\mathbb{C}^n$. This subspace is denoted by $Row(A)$.

(d) Analogously we can define $Col(A^\star)$ and $Row(A^\star)$. In the case of real matrices we look at $Col(A^T)$ and $Row(A^T)$. We observe that

$$\begin{aligned} Col(A) &= Row(A^T) \\ Row(A) &= Col(A^T) \end{aligned}$$

(e) If $A \in \mathbb{R}^{m\times n}$ then we define

$$\begin{aligned} \mathcal{N}_A &= \{x \in \mathbb{R}^n : Ax = \theta_m\} \text{ (Null Space of } A) \\ \mathcal{R}_A &= \{b \in \mathbb{R}^m : \exists x \in \mathbb{R}^n \ni Ax = b\} \text{ (Range of } A) \\ &= \text{The set of all vectors of the form } Ax \text{ where } x \in \mathbb{R}^n \\ \mathcal{N}_{A^T} &= \left\{y \in \mathbb{R}^m : A^T y = \theta_n\right\} \text{ (Null Space of } A^T) \\ \mathcal{R}_A &= \left\{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^m \ni A^T y = x\right\} \text{ (Range of } A^T) \\ &= \text{The set of all vectors of the form } A^T y \text{ where } y \in \mathbb{R}^m \end{aligned}$$

Among these, $\mathcal{N}_A$ and $\mathcal{R}_{A^T}$ are subspaces of $\mathbb{R}^n$ and $\mathcal{N}_{A^T}$ and $\mathcal{R}_A$ are subspaces of $\mathbb{R}^m$.

The dimension of the range of a matrix is called the "**Rank**" of the matrix and the dimension of the null space of a matrix is called the "**nullity**" of the matrix. We have the following connections between the ranks and nullities:

i. For any matrix $A \in \mathbb{R}^{m\times n}$, we have

$$Rank + Nullity = Number\ of\ Columns$$

(**Rank-Nullity Theorem**)

ii. Rank of A=Rank of A^T

iii. $Col(A) = \mathcal{R}_A = Row(A^T)$

iv. $Col(A^T) = \mathcal{R}_{A^T} = Row(A)$

v. dimensions of $Col(A)$ and $Row(A^T)$, $Col(A^T)$ and $Row(A)$ are same as Rank of A

6. **Inner Product and Length**

For any two vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

in $\mathbb{R}^n$ we define their "**dot product**" - also referred to as "**inner product**" - as

$$(x, y) \;\; = \;\; y^T x = \sum_{j=1}^{n} x_j y_j$$

For any two vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

in $\mathbb{C}^n$ we define their "**inner product**" as

$$(x, y) \;\; = \;\; y^\star x = \sum_{j=1}^{n} x_j \overline{y_j}$$

where

$$y^\star \;\; = \;\; \overline{y^T}$$

Without loss of generality we will write $y^\star x$ in both cases, since in the real case $y^\star = y^T$. It is easy to see that the inner product satisfies the following properties:

(a) We have, for any $x \in \mathbb{C}^n$

$$\begin{aligned} (x, x) &= x^\star x \\ &= \sum_{j=1}^{n} |x_j|^2 \\ &\geq 0 \text{ and } = 0 \text{ if and only if } x = \theta_n \end{aligned}$$

Thus we have
$\boldsymbol{(x, x) \geq 0}$ **for all** $\boldsymbol{x \in \mathbb{C}^n}$ **and** $\boldsymbol{= 0}$ **if and only if** $\boldsymbol{x = \theta_n}$

(b) Next we observe that

$$\begin{aligned}(x, y) &= \sum_{j=1}^{n} x_j \overline{y_j} \\ &= \overline{\sum_{j=1}^{n} \overline{x_j} y_j} \\ &= \overline{(y, x)}\end{aligned}$$

Thus we have
$\boldsymbol{(x, y) = \overline{(y, x)}}$ **for any two vectors** $\boldsymbol{x, y \in \mathbb{C}^n}$

(c) It is easy to verify that
$\boldsymbol{(x, y + z) = (x, y) + (x, z)}$ **for any** $\boldsymbol{x, y, z \in \mathbb{C}^n}$ **and**
$\boldsymbol{(\alpha x, y) = \alpha(x, y)}$ **for any** $\boldsymbol{x, y \in \mathbb{C}^n}$ **and any** $\boldsymbol{\alpha \in \mathbb{C}}$
As a consequence of the above properties we also get
$\boldsymbol{(x + y, z) = (x, y) + (y, z)}$ **for any** $\boldsymbol{x, y, z \in \mathbb{C}^n}$ **and**
$\boldsymbol{(x, \alpha y) = \overline{\alpha}(x, y)}$ **for any** $\boldsymbol{x, y \in \mathbb{C}^n}$ **and any** $\boldsymbol{\alpha \in \mathbb{C}}$

(d) The "**length** of any vector is defined as

$$\|\boldsymbol{x}\| = \sqrt{(\boldsymbol{x}, \boldsymbol{x})}$$

(e) In $\mathbb{R}^2$ we have the formula

$$(x, y) = \|x\|\|y\|cos(\varphi)$$

where φ is the angle between the two vectors x and y. From this using the fact that $|cos(\varphi)| \leq 1$ we get

$$|(x, y)| \leq \|x\|\|y\|$$

and equality holds if and only if $|cos(\varphi)| = 1$, that is if and only if $\varphi = 0$ or π, that is if and only if x and y are along the same line - meaning they are linearly dependent. This is in general true, that is,
For any two vectors $\boldsymbol{x, y \in \mathbb{C}^n}$

$$|(\boldsymbol{x}, \boldsymbol{y})| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

where equality holds if and only if $\boldsymbol{x, y}$ **are linearly dependent** This is known as "**Cauchy-Schwarz inequality**".

(f) The length as defined in (d) above satisfies the following properties:

i. $\|\boldsymbol{x}\| \geq \mathbf{0}$ **for all** $\boldsymbol{x} \in \mathbb{C}^{\boldsymbol{n}}$ **and** $= \mathbf{0}$ **if and only if** $\boldsymbol{x} = \boldsymbol{\theta_n}$

ii. $\|\boldsymbol{\alpha x}\| = |\boldsymbol{\alpha}|\ \|\boldsymbol{x}\|$ **for all** $\boldsymbol{x} \in \mathbb{C}^{\boldsymbol{n}}$ **and for all** $\boldsymbol{\alpha} \in \mathbb{C}$

iii. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ **- (known as "Triangle Inequality") - where equality holds if and only if** $\boldsymbol{x}, \boldsymbol{y}$ **are linearly dependent.**

7. **Orthogonality**

Two vectors x and y in $\mathbb{C}^n$ are said to be orthogonal if $(x, y) = 0$, that is if

$$(x, y) \;=\; y^{\star} x = \sum_{j=1}^{n} x_j \overline{y_j} = 0$$

Note that the zero vector θ_n is orthogonal to all the vectors and that this is the only vector orthogonal to all the vectors.

8. **Orhonormal Sets and Orthonormal Basis**

As mentioned in (4) above, if $\mathcal{B}_{\mathcal{W}} = \{u_1, u_2, \cdots, u_d\}$ is a basis for a subspace $\mathcal{W}$ of $\mathbb{C}^N$ then every vector $w \in \mathcal{W}$ has a unique linear combination representation

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_d u_d$$

and we write

$$[w]_{\mathcal{B}_{\mathcal{W}}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{pmatrix}$$

α_j is called the j-th component of the vector w with respect to the basis $\mathcal{B}_{\mathcal{W}}$.

However,in general, it is not easy to find these components α_j since this may involve solving a large system of equations. In order to avoid such computations we introduce the notion of an orthonormal basis.

Definition 3.2.6 A set of vectors $\boldsymbol{\mathcal{S}} = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k}\}$ is said to be an orthonormal set if

$$(\boldsymbol{u_i}, \boldsymbol{u_j}) = \begin{cases} \mathbf{0} & \textbf{if} \quad \boldsymbol{i \neq j} \\ \mathbf{1} & \textbf{if} \quad \boldsymbol{i = j} \end{cases}$$

This means that the vectors in $\mathcal{S}$ are pairwise orthogonal and each vector in $\mathcal{S}$ has length one

Example 3.2.6 The set of vectors

$$\mathcal{S} = \left\{ u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},\ u_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, u_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

is s set of vectors which are pairwise orthogonal but they are not normalized to norm one and hence this is not an orthonormal set in $\mathbb{R}^4$. However the set

$$\mathcal{S}_1 = \left\{ u_1 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},\ u_2 = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, u_3 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

is an orthonormal set in $\mathbb{C}^4$

It is easy to verify that
Every Orthonormal set is linearly independent
We next introduce the notion of an orthonormal basis

Definition 3.2.7 An orthonormal set $\boldsymbol{\mathcal{S}}$ in a subspace $\boldsymbol{\mathcal{W}}$ which is a basis for $\boldsymbol{\mathcal{W}}$ is said to be an orthonormal basis for $\boldsymbol{\mathcal{W}}$

Thus for a set $\mathcal{S} = \{u_1, u_2, \cdots, u_d\}$ to be an orthonormal basis for $\mathcal{W}$ it must satisfy the following conditions:

(a) $\mathcal{S}$ must be an orthonormal set

(b) $\mathcal{S} \subset \mathcal{W}$, that is, $u_j \in \mathcal{W}$ for $1 \leq j \leq d$, and

(c) every vector in $\mathcal{W}$ is a linear combination of the vectors in $\mathcal{S}$

Example 3.2.7 Consider the subspace $\mathcal{W}$ of $\mathbb{C}^4$ defined as follows:

$$\mathcal{W} = \left\{ x = \begin{pmatrix} \alpha+\beta+\gamma \\ \alpha-\beta+\gamma \\ \alpha+\beta-\gamma \\ \alpha-\beta-\gamma \end{pmatrix} \right\}$$

The set

$$\mathcal{S}_1 = \left\{ u_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},\ u_2 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, u_3 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

is an orthonormal basis for $\mathcal{W}$.

If $\boldsymbol{\mathcal{S}} = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_d}\}$ is an orthonormal basis for a subspace $\boldsymbol{\mathcal{W}}$ then every vector $\boldsymbol{w}$ in $\boldsymbol{\mathcal{W}}$ can be expanded in terms of this orthonormal basis as

$$\begin{aligned} \boldsymbol{w} &= \sum_{j=1}^{d} (\boldsymbol{w}, \boldsymbol{u_j})\boldsymbol{u_j} \\ &= \sum_{j=1}^{d} (\boldsymbol{u_j^\star w})\boldsymbol{u_j} \end{aligned}$$

and this expansion in terms of this orthonormal basis is unique

The component $(\boldsymbol{u_j^\star w})$ is called the orthogonal projection of $\boldsymbol{w}$ onto $\boldsymbol{u_j}$

For any $\boldsymbol{x} \in \mathbb{C}^n$ the vector $\boldsymbol{x_{\mathcal{W}}} = \sum_{j=1}^{d} (\boldsymbol{u_j^\star x})\boldsymbol{u_j}$ is called the orthogonal projection of $\boldsymbol{x}$ onto the subspace $\boldsymbol{\mathcal{W}}$

9. <u>Gram Schmidt Orthonormalization</u>
We shall describe the process in $\mathbb{R}^n$. The extension to $\mathbb{C}^n$ is similar to this - wherever the transpose occurs in the process in $\mathbb{R}^n$, it has to be

replaced by $\star$ for $\mathbb{C}^n$.
Let

$$\{u_1, u_2, \cdots, u_r\}$$

be a linearly independent set in $\mathbb{R}^k$. Gram Schmidt orthonormalization provides an orthonormal basis

$$\{\boldsymbol{q_1}, \boldsymbol{q_2}, \cdots, \boldsymbol{q_r}\}$$

such that

$$\mathcal{L}[u_1, u_2, \cdots, u_j] \quad = \quad \mathcal{L}[\boldsymbol{q_1}, \boldsymbol{q_2}, \cdots, \boldsymbol{q_j}] \text{ for } 1 \leq j \leq r \qquad (3.2.8)$$

where $\mathcal{L}[u_j]$ denotes the subspace for which $u_1, u_2, \cdots, u_j$ is a basis. These are obtained as follows:
First we look at the vector u_1 and normalize it to length one, that is, we define $v_1 = u_1$ and $\boldsymbol{q_1} = \dfrac{v_1}{\|v_1\|}$.
We next look at the vector u_2 and remove that part of u_2 which is along $\boldsymbol{q_1}$, that is we remove the orthogonal projection of u_2 onto $\boldsymbol{q_1}$, and call the remaining as v_2. Thus we have

$$\begin{aligned} v_2 &= u_2 - (u_2, \boldsymbol{q_1})\boldsymbol{q_1} \\ &= u_2 - \frac{(u_2, v_1)}{\|v_1\|^2} v_1 \end{aligned}$$

Clearly the vector v_2 is orthogonal to $\boldsymbol{q_1}$ and we normalize it to norm one by defining

$$\boldsymbol{q_2} \quad = \quad \frac{v_2}{\|v_2\|}$$

The vectors $\boldsymbol{q_1}$,$\boldsymbol{q_2}$ are orthonormal and span the same subspace as u_1, u_2.
We next consider u_3 and remove the parts that are along $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ and then normalize the remaining to norm one. Thus we define

$$\begin{aligned} v_3 &= u_3 - (u_3, \boldsymbol{q_1})\boldsymbol{q_1} - (u_3, \boldsymbol{q_2})\boldsymbol{q_2} \\ &= u_2 - \frac{(u_3, v_1)}{\|v_1\|^2} v_1 - \frac{(u_3, v_2)}{\|v21\|^2} v_2 \\ \boldsymbol{q_3} &= \frac{v_3}{\|v_3\|} \end{aligned}$$

The vectors $\boldsymbol{q_1}$,$\boldsymbol{q_2}$ $\boldsymbol{q_2}$, are orthonormal and span the same subspace as u_1, u_2, u_3.
We continue this process and get the required orthonormal vectors recursively defined as follows:

$$\begin{aligned} \boldsymbol{v_1} &= \boldsymbol{u_1} && (3.2.9) \\ \boldsymbol{q_1} &= \frac{\boldsymbol{v_1}}{\|\boldsymbol{v_1}\|} && (3.2.10) \end{aligned}$$

For $\boldsymbol{j \geq 2}$ we define recursively,

$$\begin{aligned} \boldsymbol{v_j} &= \boldsymbol{u_j} - \sum_{k=1}^{j-1} \frac{(\boldsymbol{u_j}, \boldsymbol{v_k})}{\|\boldsymbol{v_k}\|^2} \boldsymbol{v_k} = \boldsymbol{u_j} - \sum_{k=1}^{j-1} (\boldsymbol{u_j}, \boldsymbol{q_k}) \boldsymbol{q_k} && (3.2.11) \\ \boldsymbol{q_j} &= \frac{\boldsymbol{v_j}}{\|\boldsymbol{v_j}\|} && (3.2.12) \end{aligned}$$

The process of defining the v_j is the orthogonalization part of the process and the process of defining the $\boldsymbol{q_j}$ from the v_j is the normalization part of the process.

Example 3.2.8 Consider the vectors

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \tag{3.2.13}$$

$$u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} \tag{3.2.14}$$

in $\mathbb{R}^4$. These are clearly independent. Note that these two vectors are already orthogonal to each other. Hence the the v_j are automatically the u_j themselves. So we have

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \tag{3.2.15}$$

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} \tag{3.2.16}$$

Consequently we have

$$\|v_1\|^2 = 3 \tag{3.2.17}$$
$$\|v_2\|^2 = 3 \tag{3.2.18}$$

Hence

$$\boldsymbol{q_1} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \tag{3.2.19}$$

$$\boldsymbol{q_2} = \frac{1}{\sqrt{3}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} \tag{3.2.20}$$

We then have

$$\begin{aligned} \mathcal{L}[\boldsymbol{q_1}] &= \mathcal{L}[u_1] \\ \mathcal{L}[\boldsymbol{q_1}, \boldsymbol{q_2}] &= \mathcal{L}[u_1, u_2] \end{aligned}$$

Example 3.2.9 Consider the following three vectors in $\mathbb{R}^3$:

$$u_1 = \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix},\ u_2 = \begin{pmatrix} 1 \\ 1 \\ -7 \end{pmatrix},\ u_3 = \begin{pmatrix} 3 \\ -1 \\ 8 \end{pmatrix}$$

It is easy to verify that these three vectors are linearly independent. (Verify this)
Let us apply Gram-Schmidt orthonormalization to this set. We have

$$v_1 = u_1 = \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix}$$

$$
\begin{aligned}
\|v_1\|^2 &= (9+9+36) = 54 \\
\|v_1\| &= 3\sqrt{6} \\
\boldsymbol{q_1} &= \frac{v_1}{\|v_1\|} \\
&= \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \\ \frac{1}{\sqrt{6}} \\ \\ \frac{2}{\sqrt{6}} \end{pmatrix}
\end{aligned}
$$

We next get $\boldsymbol{q_2}$

$$
\begin{aligned}
v_2 &= u_2 - \frac{(u_2, v_1)}{\|v_1\|^2} v_1 \\
&= \begin{pmatrix} 1 \\ 1 \\ -7 \end{pmatrix} - \frac{-36}{54} \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix} \\
&= \begin{pmatrix} 3 \\ 3 \\ -3 \end{pmatrix} \\
\|v_2\|^2 &= 27 \\
\|v_2\| &= 3\sqrt{3} \\
\boldsymbol{q_2} &= \frac{v_2}{\|v_2\|} \\
&= \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \\ \frac{1}{\sqrt{3}} \\ \\ -\frac{1}{\sqrt{3}} \end{pmatrix}
\end{aligned}
$$

Finally we get $\boldsymbol{q_3}$

$$
\begin{aligned}
v_3 &= u_3 - \frac{(u_3, v_1)}{\|v_1\|^2} v_1 - \frac{(u_3, v_2)}{\|v_2\|^2} v_2 \\
&= \begin{pmatrix} 3 \\ -1 \\ 8 \end{pmatrix} - \frac{54}{54} \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix} - \frac{-18}{27} \begin{pmatrix} 3 \\ 3 \\ -3 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix} \\
\|v_3\|^2 &= 8 \\
\|v_3\| &= 2\sqrt{2} \\
\boldsymbol{q_3} &= \frac{v_3}{\|v_3\|} \\
&= \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}
\end{aligned}
$$

Thus, starting from the ordered basis u_1, u_2, u_3 above, the Gram-Schmidt orthonormalization gives rise to the ordered orthonormal basis $\boldsymbol{q_1, q_2, q_3}$ for $\mathbb{R}^3$, where

$$
\boldsymbol{q_1} = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}, \ \boldsymbol{q_2} = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}, \ \boldsymbol{q_3} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}
$$

3.3 Diagonalization

We now study the structure of a diagonalizable matrix $A \in \mathbb{C}^{n\times n}$. We define the diagonalizability of a matrix as follows:

Definition 3.3.1 A matrix $\boldsymbol{A} \in \mathbb{C}^{\boldsymbol{n\times n}}$ is said to be diagonalizable over $\mathbb{C}$ if there exists an ivertible matrix $\boldsymbol{P} \in \mathbb{C}^{\boldsymbol{n\times n}}$ such that $\boldsymbol{P^{-1}AP}$ is a diagonal matrix $\boldsymbol{D} \in \mathbb{C}^{\boldsymbol{n\times n}}$

Let us consider a diagonalizable matrix $A \in \mathbb{C}^{n\times n}$ and analyse what are the ingredients that make the matrix a diagonalizable matrix. Suppose A is diagonalizble. Then we must have an invertible $P \in \mathbb{C}^{n\times n}$ such that

$$P^{-1}AP = D = \begin{pmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \lambda_n \end{pmatrix} \in \mathbb{C}^{n\times n}$$

We can write this as

$$AP = PD$$

If we now denote the jth column of P as P_j then we get

$$A[P_1\ P_2\ \cdots\ \cdots\ P_j \cdots\ P_n] \ = \ [P_1\ P_2\ \cdots\ \cdots\ P_j \cdots\ P_n]D$$

From this we get

$$[AP_1\ AP_2\ \cdots\ \cdots\ AP_j \cdots\ AP_n] \ = \ [\lambda_1 P_1\ \ \lambda_2 P_2\ \cdots\ \cdots\ \lambda_j P_j \cdots\ \lambda_n P_n]$$

Comparing the jth columns on both sides we get,

$$AP_j \ = \ \lambda_j P_j \text{ for } 1 \leq j \leq n \tag{3.3.1}$$

We note that the column vectors $P_1, P_2, \cdots P_n$ are linearly independent vectors in $\mathbb{C}^n$, (since P is invertible). Thus we have,

Conclusion :

$\boldsymbol{A \in \mathbb{C}^{n\times n}}$ **is diagonalizable over** $\mathbb{C}$

$\Longrightarrow$

There exist $\boldsymbol{n}$ scalars $\boldsymbol{\lambda_1, \lambda_2, \cdots, \lambda_n}$ in $\mathbb{C}$ and $\boldsymbol{n}$ linearly independent vectors $\boldsymbol{P_1, P_2, \cdots, P_n}$ in $\mathbb{C}^n$ such that

$$\boldsymbol{AP_j = \lambda_j P_j \textbf{ for } 1 \leq j \leq n}$$

Conversely, it is easy to see that if there exists n scalars $\lambda_1, \lambda_2, \cdots, \lambda_n$ in $\mathbb{C}$ and n linearly independent vectors $P_1, P_2, \cdots, P_n$ in $\mathbb{C}^n$ such that

$$AP_j = \lambda_j P_j \text{ for } 1 \leq j \leq n$$

then we can define $P \in \mathbb{C}^{n\times n}$ as the matrix whose jth column is P_j, and then we get $P^{-1}AP$ as the diagonal matrix whose n diagonal entries are respectively $\lambda_1, \lambda_2, \cdots \lambda_n$. Combining this with the above Conclusion we get

Theorem 3.3.1 $\boldsymbol{A \in \mathbb{C}^{n\times n}}$ **is diagonalizable over** $\mathbb{C}$
$\Longleftrightarrow$
There exist $\boldsymbol{n}$ **scalars** $\boldsymbol{\lambda_1, \lambda_2, \cdots, \lambda_n}$ **in** $\mathbb{C}$ **and** $\boldsymbol{n}$ **linearly independent vectors** $\boldsymbol{P_1, P_2, \cdots, P_n}$ **in** $\mathbb{C}^n$ **such that**

$$\boldsymbol{AP_j = \lambda_j P_j} \textbf{ for } \boldsymbol{1 \le j \le n}$$

Thus if we have to diagonalize a matrix A we need n pairs (λ_j, P_j). where $\lambda_j \in \mathbb{C}$ and $P_j \in \mathbb{C}^n$ such that $AP_j = \lambda_j P_j$. This leads us to the notion of eigenvalues and eigenvectors.

Remark 3.3.1 While seeking these n pairs (λ_j, P_j), it is not necessary that the scalars λ_j be distinct. Some of them may even be repeated. However, the vectors P_j that we are seeking must be linearly independent and hence they must all be nonzero vectors.

Thus we must search for these n scalars and n linearly independent vectors as above. This leads us to the notion of eigenvalues and eigenvectors. We begin with the definition of eigenvalues and eigenvectors.

Definition 3.3.2 A scalar $\boldsymbol{\lambda \in \mathbb{C}}$ **is said to be an eigenvalue of a matrix** $\boldsymbol{A \in \mathbb{C}^{n\times n}}$ **if there exists a nonzero vector** $\boldsymbol{\varphi \in \mathbb{C}^n}$ **such that**

$$\boldsymbol{A\varphi} \;=\; \boldsymbol{\lambda\varphi} \tag{3.3.2}$$

If $\boldsymbol{\lambda \in \mathbb{C}}$ **is an eigenvalue of** $\boldsymbol{A}$ **then any nonzero vector** $\boldsymbol{\varphi \in \mathbb{C}^n}$ **such that** $\boldsymbol{A\varphi = \lambda\varphi}$ **is called an eigenvector corresponding to the eigenvalue** $\boldsymbol{\lambda}$**. We shall call an eigenvalue-eigenvector pair** $\boldsymbol{(\lambda, \varphi)}$**, as an eigenpair**
Given any $A \in \mathbb{C}^{n\times n}$ can we find n such eigenpairs in which the vectors in the n pairs are all linearly independent? The answer is, in general, this may not br possible.
Where should we look for the eigenvalues of a matrix? We have the following:

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } A \in \mathbb{C}^{n\times n} &\Longleftrightarrow \exists \varphi \in \mathbb{C}^n \ni A\varphi = \lambda\varphi \\
&\Longleftrightarrow A_\lambda \varphi = \theta_n \\
&\Longleftrightarrow \text{The homogeneous system } A_\lambda x = \theta_n \text{ has} \\
&\qquad \text{nontrivial solution } \varphi \\
&\Longleftrightarrow |A_\lambda| = 0 \text{ where } A_\lambda = \lambda I - A
\end{aligned}$$

Thus the eigenvalues are the roots of the function $c_{_A}(\lambda)$ where

$$c_{_A}(\lambda) = |\lambda I - A| \tag{3.3.3}$$

We have

$$c_{_A}(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & \cdots & -a_{1j} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & \cdots & -a_{2j} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{j1} & -a_{j2} & \cdots & \cdots & \lambda - a_{jj} & \cdots & -a_{jn} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \cdots & -a_{nj} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

When we expand this determinant we get a polynomial,

$$c_{_A}(\lambda) = \lambda^n - (Trace(A))\lambda^{(n-1)} + \cdots + (-1)^n |A| \tag{3.3.4}$$

where

$$Trace(A) = \sum_{i=1}^{n} a_{ii} \text{ (sum of the diagonal entries of } A) \tag{3.3.5}$$

$c_{_A}(\lambda)$ is a MONIC polynomial of degree n whose coefficients are in $\mathbb{C}$, that is, $c_{_A}(\lambda) \in \mathbb{C}[\lambda]$. This polynomial is called the **Characteristic Polynomial** of the matrix A. The eigenvalues that we are looking for are precisely the roots of this polynomial in $\mathbb{C}$. By the fundamental theorem of algebra, we can factorize this polynomial into products of linear factors as,

$$c_{_A}(\lambda) = (\lambda - \lambda_1)^{a_1}(\lambda - \lambda_2)^{a_2} \cdots (\lambda - \lambda_k)^{a_k} \tag{3.3.6}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are distinct elements of $\mathbb{C}$ and $a_1, a_2, \cdots, a_k$ are positive integers such that

$$a_1 + a_2 + \cdots + a_k = n \tag{3.3.7}$$

(Clearly, this is possible for all matrices in $\mathbb{C}^{n\times n}$).
$\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct eigenvalues and $a_1, a_2, \cdots, a_k$ are their multiplicities as roots of the characteristic polynomial. The multiplicity a_j is called the **Algebraic Multiplicity** of the eigenvalue λ_j.

Having found where to look the eigenvalues we next look at the eigenvectors. Let λ_j be an eigenvalue. There should be at least one nonzero vector $u \in \mathbb{C}^n$ such that $Au = \lambda_j u$. We define

$$\begin{aligned} \mathcal{W}_j &= Null\ Space\ of\ A - \lambda_j I && (3.3.8) \\ &= \{x \in \mathbb{C}^n : Ax = \lambda_j x\} && (3.3.9) \end{aligned}$$

Any nonzero vector in $\mathcal{W}_j$ is an eigenvector corresponding to the eigenvalue λ_j. Since $\mathcal{W}_j$ is the Null Space of the matrix $A - \lambda_j I$, it is a subspace of $\mathbb{C}^n$. This subspace is called the **Eigenspace** corresponding to the eigenvalue λ_j. The dimension of this subspace is called the **Geometric Multiplicity** of the eigenvalue λ_j and we denote this by g_j. Thus we have

$$\text{geometric multiplicity, } g_j = dimension\ of\ \mathcal{W}_j \qquad (3.3.10)$$

We shall state the following facts without proof:

1. **For every eigenvalue $\boldsymbol{\lambda_j}$ the geometric multiplicity $\boldsymbol{g_j}$ is at least one and at most the algebbraic multiplicity $\boldsymbol{a_j}$, that is,**

   $$\mathbf{1 \leq g_j \leq a_j \text{ for every } j,\ (1 \leq j \leq k)}$$

2. **A matrix $\boldsymbol{A} \in \mathbb{C}^{\boldsymbol{n\times n}}$ is diagonalizable**
   $\Longleftrightarrow$
   $\boldsymbol{g_j = a_j}$ for every $\boldsymbol{j}$, $(\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{k})$,
   that is, a matrix $\boldsymbol{A} \in \mathbb{C}^{\boldsymbol{n\times n}}$ is diagonalizable if and only if for every eigenvalue the algebraic multiplicity is the same as the geometric multiplicity

Suppose now that the algebraic multiplicity of each eigenvalue is equal to its geometric multiplicity. Then as observed above the matrix A is diagonalizable. This means that there exists an onvertible matrix P in $\mathbb{C}^{n\times n}$ such that $P^{-1}AP$ is a diagonal matrix $D \in \mathbb{C}^{n\times n}$. The matrix is constructed from the eigenvectors of A. Let $\lambda_1, \lambda_2, \cdots, \lambda_k$ be the distnct eigenvalues with $a_j = g_j$ for $1 \leq j \leq k$. Then there are a_j linearly independent eigenvectors $u_1^{(j)}, u_2^{(j)}, , \cdots, u_{a_j}^{(j)}$ corresponding to the eigenvalue λ_j. We now construct the matrix P whose first a_1 columns are the a_1 eigenvectors $\left\{u_r^{(1)}\right\}_{r=1}^{a_1}$ corresponding to the eigenvalue λ_1, next a_2 columns are the a_2 eigenvectors $\left\{u_r^{(2)}\right\}_{r=1}^{a_2}$

corresponding to the eigenvalue λ_2, and so on, and the last a_k columns are the a_k eigenvectors $\left\{u_r^{(k)}\right\}_{r=1}^{a_k}$ corresponding to the eigenvalue λ_k. We have

$$P = \left[\begin{array}{ccc|c|ccc|c|ccc} u_1^{(1)} & \cdots & u_{a_1}^{(1)} & \cdots & u_1^{(j)} & \cdots & u_{a_j}^{(j)} & \cdots & u_1^{(k)} & \cdots & u_{a_k}^{(k)} \end{array}\right]$$

Then

$$P^{-1}AP = diagonal(\underbrace{\lambda_1, \lambda_1, \cdots, \lambda_1}_{a_1\ times}\ \underbrace{\lambda_1, \lambda_2, \cdots, \lambda_2}_{a_2\ times} \cdots \underbrace{\lambda_k, \lambda_k, \cdots, \lambda_k}_{a_k\ times})$$

$$= \left(\begin{array}{cccc|cc|ccc} \lambda_1 & & & & & & & & \\ & \lambda_1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & \lambda_1 & & & & & \\ \hline & & & & \ddots & & & & \\ & & & & & \ddots & & & \\ \hline & & & & & & \lambda_k & & \\ & & & & & & & \ddots & \\ & & & & & & & & \lambda_k \end{array}\right)$$

$$= \begin{pmatrix} \lambda_1 I_{a_1\times a_1} & & & & & \\ & \lambda_2 I_{a_2\times a_2} & & & & \\ & & \ddots & & & \\ & & & \lambda_j I_{a_j\times a_j} & & \\ & & & & \ddots & \\ & & & & & \lambda_k I_{a_k\times a_k} \end{pmatrix}$$

We shall now look at some examples.

Example 3.3.1 Consider the matrix

$$A = \begin{pmatrix} 4 & 8 & -2 \\ -3 & -6 & 1 \\ 9 & 12 & -5 \end{pmatrix}$$

The characteristic polynomial of this matrix is given by

$$c(\lambda) = (\lambda+3)(\lambda+2)^2$$

Hence the distinct eigenvalues are

$$\lambda_1 = -3 \text{ and } \lambda_2 = -2$$

with respective algebraic multiplicities

$$a_1 = 1 \text{ and } a_2 = 2$$

We next find the eigenspaces and geometric multiplicities.
<u>Eigenspace corresponding to Eigenvalue $\lambda_1 = -3$:</u>
We have to solve the system $(A - \lambda_1 I)x = \theta_3$, that is $(A + 3I)x = \theta_3$. We have

$$A + 3I \;\;=\;\; \begin{pmatrix} 7 & 8 & -2 \\ -3 & -3 & 1 \\ 9 & 12 & -2 \end{pmatrix}$$

Hence we get the system $(A + 3I)x = \theta_3$ as

$$\begin{array}{rcrcrcl} 7x_1 & + & 8x_2 & - & 2x_3 & = & 0 \\ -3x_1 & - & 3x_2 & + & x_3 & = & 0 \\ 9x_1 & + & 12x_2 & - & 2x_3 & = & 0 \end{array}$$

Solving this system we get that the eigenspace corresponding to the eigenvalue $\lambda_1 = -3$ as

$$W_1 \;\;=\;\; \left\{ x = \alpha \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix} : \alpha \in \mathbb{C} \right\}$$

The vector

$$u_1 = \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_1 = -3$ and this forms a basis for W_1. Hence

$$g_1 = \mathit{dimension}\; W_1 = 1$$

<u>Eigenspace corresponding to Eigenvalue $\lambda_2 = -2$</u>
We have to solve the system $(A - \lambda_2 I)x = \theta_3$, that is $(A + 2I)x = \theta_3$. We have

$$A + 2I \;\;=\;\; \begin{pmatrix} 6 & 8 & -2 \\ -3 & -4 & 1 \\ 9 & 12 & -3 \end{pmatrix}$$

Hence we get the system $(A+2I)x = \theta_3$ as

$$\begin{array}{rcrcrcl} 6x_1 & + & 8x_2 & - & 2x_3 & = & 0 \\ -3x_1 & - & 4x_2 & + & x_3 & = & 0 \\ 9x_1 & + & 12x_2 & - & 3x_3 & = & 0 \end{array}$$

Solving this system we get that the eigenspace corresponding to the eigenvalue $\lambda_1 = -3$ as

$$W_1 = \left\{ x = \alpha \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix} : \alpha, \beta \in \mathbb{C} \right\}$$

The vectors

$$u_2 = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \; u_3 = \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix}$$

are linearly independent eigenvectors corresponding to $\lambda_2 = -2$ and these forms a basis for W_2. Hence

$$g_2 = \textit{dimension}\; W_2 = 2$$

Thus we have

$$\begin{array}{lll} \lambda_1 = -3 & a_1 = 1 & g_1 = 1 \\ \lambda_2 = -2 & a_2 = 2 & g_2 = 2 \end{array}$$

Since the algebraic and geometric multiplicities tally for both the eigenvalues the matrix is diagonalizable. The diagonalizing matrix P is obtained as follows:

$$\begin{aligned} P &= \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \\ &= \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ -3 & 3 & 4 \end{pmatrix} \end{aligned}$$

It can be easily verified that

$$P^{-1}AP = D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

We get the decomposition

$$A = PDP^{-1}$$

Example 3.3.2 Consider the matrix

$$A = \begin{pmatrix} 0 & -6 & -4 \\ 5 & -11 & -6 \\ -6 & 9 & 4 \end{pmatrix}$$

The characteristic polynomial of this matrix is given by

$$c(\lambda) = (\lambda+3)(\lambda+2)^2$$

The eigenvalues and their algebraic multiplicities are given by

$$\lambda_1 = -3, a_1 = 1 \text{ and } \lambda_2 = -2, a_2 = 2$$

We next find the eigenspaces and the geometric multiplicities.
<u>Eigenspace corresponding to Eigenvalue $\lambda_1 = -3$</u>
We have to solve the system $(A+3I)x = \theta_3$. We have

$$A + 3I = \begin{pmatrix} 3 & -6 & -4 \\ 5 & -8 & -6 \\ -6 & 9 & 7 \end{pmatrix}$$

Hence we get the system

$$\begin{array}{rcrcrcl} 3x_1 & - & 6x_2 & - & 4x_3 & = & 0 \\ 5x_1 & - & 8x_2 & - & 6x_3 & = & 0 \\ -6x_1 & + & 9x_2 & + & 7x_3 & = & 0 \end{array}$$

Solving this system we get the eigenspace W_1 corresponding to the eigenvalue $\lambda_1 = -3$ as

$$W_1 = \left\{ x = \alpha \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix} \right\}$$

The vector

$$u_1 = \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix}$$

is an eigenvector corresponding to the eigenvalue $\lambda_1 = -3$ and forms a basis for W_1. Hence

$$g_1 = dimension\ W_1 = 1$$

Eigenspace corresponding to Eigenvalue $\lambda_2 = -2$
We have to solve the system $(A + 2I)x = \theta_3$. We have

$$A + 2I = \begin{pmatrix} 2 & -6 & -4 \\ 5 & -9 & -6 \\ -6 & 9 & 6 \end{pmatrix}$$

Hence we get the system

$$\begin{array}{rcrcrcl} 2x_1 & - & 6x_2 & - & 4x_3 & = & 0 \\ 5x_1 & - & 9x_2 & - & 6x_3 & = & 0 \\ -6x_1 & + & 9x_2 & + & 6x_3 & = & 0 \end{array}$$

Solving this system we get the eigenspace W_2 corresponding to the eigenvalue $\lambda_1 = -2$ as

$$W_2 = \left\{ x = \alpha \begin{pmatrix} 0 \\ 2 \\ -3 \end{pmatrix} : \alpha \in \mathbb{C} \right\}$$

The vector

$$u_2 = \begin{pmatrix} 0 \\ 2 \\ -3 \end{pmatrix}$$

is an eigenvector corresponding to the eigenvalue $\lambda_2 = -2$ and forms a basis for W_2. Hence

$$g_2 = dimension\ W_2 = 1$$

Thus we have

$$\begin{array}{lll} \lambda_1 = -3 & a_1 = 1 & g_1 = 1 \\ \lambda_2 = -2 & a_2 = 2 & g_2 = 1 \end{array}$$

Since the algebraic and geometric multiplicities of $\lambda_2 = -2$ do not tally the matrix is not diagonalizable. We shall later see that this matrix can be triangularized.

3.4 Schur's Similarity Triangularization

In the last section we saw that not all matrices in $\mathbb{C}^{n \times n}$ can be diagonalized. We needed for diagonalizability the matrix has to satisfy the condition that

for every eigenvalue the algebraic multiplicity is the same as the geometric multiplicity. However, in this section, we shall see that all matrices can be triangularized. We shall see a method, due to Schur, to triangularize a matrix.
Consider a matrix $A \in \mathbb{C}^{n\times n}$. We shall show that this can be reduced to a triangular matrix by a similarity transformation, that is, we shall show that there exists an invertible matrix $P \in \mathbb{C}^{n\times n}$ such that $P^{-1}AP$ is an upper triangular matrix $U \in \mathbb{C}^{n\times n}$. We get the similarity transformation as follows:
Let λ_1 be an eigenvalue of A and let u_1 be a corresponding eigenvector. Then we have u_1 is a nonzero vector in $\mathbb{C}^n$ such that

$$Au_1 = \lambda_1 u_1 \tag{3.4.1}$$

We now choose any $(n-1)$ linearly independent vectors $u_2, u_3, \cdots, u_n$ such that

$$u_1, u_2, u_3, \cdots, u_n$$

is a basis for $\mathbb{C}^n$. Let $P^{(1)}$ be the matrix whose columns are these n vectors. Thus

$$P^{(1)} = \begin{bmatrix} u_1 & u_2 & \cdots & u_j & \cdots & u_n \end{bmatrix}$$

Since the columns are independent, the matrix P is invertible. We now observe that

$$\begin{aligned} AP^{(1)} &= A\begin{bmatrix} u_1 & u_2 & \cdots & u_j & \cdots & u_n \end{bmatrix} \\ &= \begin{bmatrix} Au_1 & Au_2 & \cdots & Au_j & \cdots & Au_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 u_1 & Au_2 & \cdots & Au_j & \cdots & Au_n \end{bmatrix} \end{aligned}$$

Thus we have

$$AP^{(1)} = \begin{bmatrix} \lambda_1 u_1 & Au_2 & \cdots & Au_j & \cdots & Au_n \end{bmatrix} \tag{3.4.2}$$

Consider the vector Au_2 which forms the second column of $AP^{(1)}$. This is a vector which is in $\mathbb{C}^n$ and hence can be expanded in terms of the basis $u_1, u_2, \cdots, u_n$ as

$$Au_2 = \alpha_{12}u_1 + \alpha_{22}u_2 + \cdots + \alpha_{j2}u_j + \alpha_{n2}u_n \tag{3.4.3}$$

Similarly,, for each j, $(2 \leq j \leq n)$, the vector Au_j which forms the j-th column of $AP^{(1)}$ can be expanded as

$$Au_j = \alpha_{1j}u_1 + \alpha_{2j}u_2 + \cdots + \alpha_{jj}u_j + \alpha_{nj}u_n \tag{3.4.4}$$

Using this in equation 3.4.2 we get

$$\begin{aligned} AP^{(1)} &= \left[\begin{array}{ccccccc} \lambda_1 u_1 & Au_2 & \cdots & Au_j & \cdots & Au_n \end{array}\right] \\ &= \left[\begin{array}{cccccc} u_1 & u_2 & \cdots & u_j & \cdots & u_n \end{array}\right] T_1 \end{aligned}$$

where T_1 is the matrix

$$T_1 = \left(\begin{array}{c|ccccc} \lambda_1 & \alpha_{12} & \cdots & \alpha_{1j} & \cdots & \alpha_{1n} \\ \hline 0 & \alpha_{22} & \cdots & \alpha_{2j} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \alpha_{j2} & \cdots & \alpha_{jj} & \cdots & \alpha_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \alpha_{n2} & \cdots & \alpha_{nj} & \cdots & \alpha_{nn} \end{array}\right) \tag{3.4.5}$$

We shall now write T_1 as

$$T_1 = \left(\begin{array}{c|c} \lambda_1 & K_{1\times(n-1)} \\ \hline 0_{(n-1)\times 1} & A_2 \end{array}\right) \tag{3.4.6}$$

where

$$K_{1\times(n-1)} = \left[\begin{array}{ccccc} \alpha_{12} & \cdots & \alpha_{1j} & \cdots & \alpha_{1n} \end{array}\right] \tag{3.4.7}$$

$$A_2 = \left(\begin{array}{ccccc} \alpha_{22} & \cdots & \alpha_{2j} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{j2} & \cdots & \alpha_{jj} & \cdots & \alpha_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{n2} & \cdots & \alpha_{nj} & \cdots & \alpha_{nn} \end{array}\right) \tag{3.4.8}$$

and

$$0_{(n-1)\times 1} = \left(\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \end{array}\right)_{(n-1)\times 1} \tag{3.4.9}$$

Thus we have

$$\begin{aligned} AP^{(1)} &= P^{(1)}T_1 \\ &\Longrightarrow \\ (P^{(1)})^{-1}AP^{(1)} &= T_1 \end{aligned}$$

Observe that the characteristic polynomial of A and T_1 are same and hence the characteristic polynomial of A is given by

$$\begin{aligned} c(\lambda &= |\lambda I - T_1| \\ &= \left| \left(\begin{array}{c|c} \lambda - \lambda_1 & -K_{1\times(n-1)} \\ \hline 0_{(n-1)\times 1} & \lambda I_{(n-1)\times(n-1)} - A_2 \end{array} \right) \right| \\ &= (\lambda - \lambda_1)c_2(\lambda) \end{aligned}$$

where $c_2(\lambda)$ is the characteristic polynomial of A_2. Thus λ_1 is the eigenvalue which we have already chosen and the other $(n-1)$ eigenvalues are obtained as the eigenvalues of the $(n-1)\times(n-1)$ matrix A_2. Now choose an eigenvalue λ_2 of A_2, (which as observed above will also be an eigenvalue of A), and an eigenvector (for A_2 corresponding to this eigenvalue) $u_2 \in \mathbb{C}^{n-1}$. Apply the above procedure to the matrix A_2 to get a matrix $P_2 \in \mathbb{C}^{(n-1)\times(n-1)}$ such that we get

$$\begin{aligned} P_2^{-1}A_2P_2 &= T_2 \\ &= \left(\begin{array}{c|c} \lambda_2 & (K_2)_{1\times(n-2)} \\ \hline 0_{(n-2)\times 1} & A_3 \end{array} \right) \end{aligned}$$

Let

$$P^{(2)} = \left(\begin{array}{c|c} 1 & 0_{1\times(n-1)} \\ \hline 0_{(n-1)\times 1} & P_2 \end{array} \right)$$

Then $P^{(2)} \in \mathbb{C}^{n\times n}$ and we can easily verify that

$$(P^{(2)})^{-1}(P^{(1)})^{-1}AP^{(1)}P^{(2)} = \left(\begin{array}{c|c|c} \lambda_1 & 0 & 0_{1\times(n-2)} \\ \hline 0 & \lambda_2 & K_{(1\times(n-2)} \\ \hline 0_{(n-2)\times 1} & 0_{(n-2)\times 1} & A_3 \end{array} \right)$$

where $A_3 \in \mathbb{C}^{(n-2)\times(n-2)}$. Continuing this process we get $P^{(3)}, \cdots, P^{(n-1)}$ such that

$$(P^{(n-1)})^{-1}(P^{(n-2)})^{-1} \cdots (P^{(2)})^{-1}(P^{(1)})^{-1}AP^{(1)}P^{(2)} \cdots P^{(n-2)}P^{(n-1)} = U$$

where U is an upper triangular matrix with the eigenvalues of A along the diagonal. We now set

$$Q = P^{(1)}P^{(2)} \cdots P^{(n-2)}P^{(n-1)}$$

We can also write this as

$$A = QUQ^{-1} \tag{3.4.10}$$

We shall now give an example of the above similarity triangularization of an $n \times n$ complex matrix.

Example 3.4.1 Consider the 3×3 matrix,

$$A = \begin{pmatrix} 0 & -6 & -4 \\ 5 & -11 & -6 \\ -6 & 9 & 4 \end{pmatrix}$$

In Example 3.3.2 we saw that this matrix is not diagonalizable. We shall now see that this can be triangularized by Schur's method described above. The characteristic polynomial of this matrix is given by

$$c(\lambda) = (\lambda + 3)(\lambda + 2)^2$$

We begin with one of the eigenvalues, say $\lambda_1 = -3$.
We next find its eigenvector. We have to solve the system

$$(A - (-3)I)x = \theta_3$$

This gives us

$$\begin{aligned} 3x_1 - 6x_2 - 4x_3 &= 0 \\ 5x_1 - 8x_2 - 6x_3 &= 0 \\ -6x_1 + 9x_2 + 7x_3 &= 0 \end{aligned}$$

The solutions of this system are of the form

$$x = \alpha \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix}$$

Hence an eigenvector corresponding to the eigenvalue λ_1 is given by

$$u_1 = \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix}$$

Next we find two linearly independent vectors u_2, u_3 such that u_1, u_2, u_3 form a basis for $\mathbb{C}^3$. We choose

$$u_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$u_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

We set

$$P^{(1)} = \left[\begin{array}{ccc} u_1 & u_2 & u_3 \end{array}\right]$$

We shall show that then we get

$$(P^1)^{-1}AP^{(1)} = T$$

where T is an upper triangular matrix. We shall construct the upper traingular matrix as follows:

$$T = \left(\begin{array}{r|cc} -3 & \alpha_1 & \beta_1 \\ \hline 0 & \alpha_2 & \beta_2 \\ 0 & \alpha_3 & \beta_3 \end{array}\right)$$

We find $\alpha_1, \alpha_2, \alpha_3$ and $\beta_1, \beta_2, \beta_3$ by expanding Au_2 and Au_3 in terms of the basis u_1, u_2, u_3 as follows:

$$\begin{aligned} Au_2 &= \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 \\ Au_3 &= \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 \end{aligned}$$

First we have

$$
\begin{aligned}
Au_2 &= \begin{pmatrix} 0 & -6 & -4 \\ 5 & -11 & -6 \\ -6 & 9 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 5 \\ -6 \end{pmatrix} \\
&= \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 \\
&= \alpha_1 \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
\Longrightarrow & \\
\begin{pmatrix} 0 \\ 5 \\ -6 \end{pmatrix} &= \begin{pmatrix} -2\alpha_1 + \alpha_2 \\ \alpha_1 + \alpha_3 \\ -3\alpha_1 \end{pmatrix} \\
\Longrightarrow & \\
\alpha_1 &= 2 \\
\alpha_2 &= 4 \\
\alpha_3 &= 3
\end{aligned}
$$

Similarly we have

$$
\begin{aligned}
Au_3 &= \begin{pmatrix} 0 & -6 & -4 \\ 5 & -11 & -6 \\ -6 & 9 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -6 \\ -11 \\ 9 \end{pmatrix} \\
&= \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 \\
&= \beta_1 \begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix} + \beta_2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \beta_3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
\Longrightarrow & \\
\begin{pmatrix} -6 \\ -11 \\ 9 \end{pmatrix} &= \begin{pmatrix} -2\beta_1 + \beta_2 \\ \beta_1 + \beta_3 \\ -3\beta_1 \end{pmatrix} \\
\Longrightarrow &
\end{aligned}
$$

$$\begin{aligned}
\beta_1 &= -3 \\
\beta_2 &= -12 \\
\beta_3 &= -8
\end{aligned}$$

Thus

$$T = \left(\begin{array}{r|rr} -3 & 2 & -3 \\ \hline 0 & 4 & -12 \\ 0 & 3 & -8 \end{array}\right)$$

We shall now verify that

$$AP^{(1)} = P^{(1)}T$$

We have

$$\begin{aligned}
AP^{(1)} &= \begin{pmatrix} 0 & -6 & -4 \\ 5 & -11 & -6 \\ -6 & 9 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ -3 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 6 & 0 & -6 \\ -3 & 5 & -11 \\ 9 & -6 & 9 \end{pmatrix}
\end{aligned}$$

Further

$$\begin{aligned}
P^{(1)}T &= \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ -3 & 0 & 0 \end{pmatrix} \left(\begin{array}{r|rr} -3 & 2 & -3 \\ \hline 0 & 4 & -12 \\ 0 & 3 & -8 \end{array}\right) \\
&= \begin{pmatrix} 6 & 0 & -6 \\ -3 & 5 & -11 \\ 9 & -6 & 9 \end{pmatrix}
\end{aligned}$$

Thus we have

$$\begin{aligned}
AP^{(1)} &= P^{(1)}T \\
&\Longrightarrow \\
(P^1)^{-1}AP^{(1)} &= T
\end{aligned}$$

Now we write T as

$$T = \left(\begin{array}{c|c} -3 & K_{1\times 2} \\ \hline 0_{1\times 2} & A^{(2)} \end{array}\right)$$

where

$$\begin{aligned} K_{1\times 2} &= \begin{pmatrix} 2 & -3 \end{pmatrix} \\ A^{(2)} &= \begin{pmatrix} 4 & -12 \\ 3 & -8 \end{pmatrix} \end{aligned}$$

We now repeat the process for $A^{(2)}$. The characteristic polynomial of $A^{(2)}$ is

$$c(\lambda) = (\lambda + 2)^2$$

We choose the eigenvalue $\lambda_2 = -2$. We then find an eigenvector. We solve the system

$$(A^{(2)} + 2I)x = \theta_2$$

We get

$$\begin{aligned} 6x_1 - 12x_2 &= 0 \\ 3x_1 - 6x_2 &= 0 \end{aligned}$$

Hence we get $x_1 = 2x_2$ and the solutions are all of the form

$$x = \alpha \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

We choose an eigenvector for $A^{(2)}$ corresponding to $\lambda_2 = -2$ as

$$v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

We now choose a vector v_2 so that v_1, v_2 is a basis for $\mathbb{C}^2$. We choose

$$v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

We set

$$P_2 = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$$

We then define

$$T^{(2)} = \begin{pmatrix} -2 & \alpha_1 \\ 0 & \alpha_2 \end{pmatrix}$$

where

$$
\begin{aligned}
A^{(2)}v_2 &= \alpha_1 v_1 + \alpha_2 v_2 \\
&\Longrightarrow \\
\begin{pmatrix} 4 & -12 \\ 3 & -8 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \alpha_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&\Longrightarrow \\
\begin{pmatrix} 4 \\ 3 \end{pmatrix} &= \begin{pmatrix} 2\alpha_1 + \alpha_2 \\ \alpha_1 \end{pmatrix} \\
&\Longrightarrow \\
\alpha_1 &= 3 \\
\alpha_2 &= -2
\end{aligned}
$$

Hence

$$
T^{(2)} = \begin{pmatrix} -2 & 3 \\ 0 & -2 \end{pmatrix}
$$

It is now easy to verify that

$$
A^{(2)}P_2 = P_2 T^{(2)} = \begin{pmatrix} -4 & 4 \\ -2 & 3 \end{pmatrix}
$$

Thus we have

$$
P_2^{-1}A^{(2)}P_2 = T^{(2)} = \begin{pmatrix} -2 & 3 \\ 0 & -2 \end{pmatrix}
$$

Now if we let

$$
P^{(2)} = \left(\begin{array}{c|c} 1 & 0_{1\times 2} \\ \hline 0_{2\times 1} & P^{(2)} \end{array}\right)
$$

then

$$
\begin{aligned}
(P^{(2)})^{-1}\left(\begin{array}{c|c} -3 & K_{2\times 1} \\ \hline 0_{2\times 1} & A^{(2)} \end{array}\right)P^{(2)} &= \left(\begin{array}{c|c} -3 & K_{2\times 1} \\ \hline 0_{2\times 1} & T^{(2)} \end{array}\right) \\
&= \left(\begin{array}{r|rr} -3 & 2 & -3 \\ \hline 0 & -2 & 3 \\ 0 & 0 & -2 \end{array}\right)
\end{aligned}
$$

Hence we get

$$(P^{(2)})^{-1}(P^{(1)})^{(-1)}AP^{(1)}P^{(2)} = \left(\begin{array}{r|r|r} -3 & 2 & -3 \\ \hline 0 & -2 & 3 \\ \hline 0 & 0 & -2 \end{array}\right)$$

and hence if we define $Q = P^{(1)}P^{(2)}$ we get

$$Q^{-1}AQ = \left(\begin{array}{r|r|r} -3 & 2 & -3 \\ \hline 0 & -2 & 3 \\ \hline 0 & 0 & -2 \end{array}\right)$$

thus showing that A is similar to a Triangular matrix.

Thus every $A \in \mathbb{C}^{n\times n}$ is similar to an upper triangular matrix. This is Schur decomposition of a matrix $A \in \mathbb{C}^{n\times n}$ by similarity transformation.
In the above method, in each step, we started with an eigenvalue and a corresponding eigenvector of a matrix A_k and appended enough linearly independent vectors to form a basis for $\mathbb{C}^k$. Instead if we start with a eigenvalue and a corresponding unit eigenvector of the matrix A_k and append enough orthonormal vectors to form an orthonormal basis for $\mathbb{C}^k$, the resulting matrices $P^{(k)}$ are all unitary matrices and if as before we define

$$Q = P^{(1)}P^{(2)}\cdots P^{(n-1)}$$

we get

$$Q^{\star}AQ = U$$

where U is an upper triangular matrix in $\mathbb{C}^{n\times n}$. This is Schur's unitary reduction to an upper triangular matrix.
We illustrate this by an example below:

Example 3.4.2 Consider the same matrix as in Example 3.4.1. We start with $\lambda_1 = -3$ and the unit eigenvector

$$u_1 = \frac{1}{\sqrt{14}}\begin{pmatrix} -2 \\ 1 \\ -3 \end{pmatrix}$$

We first two vectors

$$v_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$$

such that v_2 and v_3 are linearly independent and orthogonal to u_1. We then find the orthonormal vectors u_2 and u_3 by applying Gram-Schmidt procedure to these two vectors. We get

$$u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \quad u_3 = \frac{1}{\sqrt{70}} \begin{pmatrix} -6 \\ 3 \\ 5 \end{pmatrix}$$

Then u_1, u_2, u_3 form an orthonormal basis for $\mathbb{C}^3$. We define

$$P^{(1)} \quad = \quad \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$$

Then $P^{(1)}$ is such that $(P^{(1)})^\star = (P^{(1)})^T = (P^{(1)})^{-1}$. Hence $P^{(1)}$ is unitary and we get

$$(P^{(1)})^T A P^{(1)} \quad = \quad \left(\begin{array}{r|r} -3 & K_{1\times 2} \\ \hline 0_{2\times 1} & A_2 \end{array}\right)$$

We now repeat this process for A_2. Clearly both the eigenvalues of A_2 are (-2) and hence the triangular form we get at the end will be

$$\left(\begin{array}{r|r|r} -3 & \star & \star \\ \hline 0 & -2 & \star \\ \hline 0 & 0 & -2 \end{array}\right)$$