Chapter 5

Diagonalization, Eigenvalues and Eigenvectors

5.1 Introduction

We now study the structure of a diagonalizable matrix $A \in \mathcal{F}^{n \times n}$. Recall that we define the diagonalizability of a matrix as follows:

Definition 5.1.1 A matrix $\boldsymbol{A} \in \mathcal{F}^{n \times n}$ is said to be diagonalizable over $\mathcal{F}$ if there exists an ivertible matrix $\boldsymbol{P} \in \mathcal{F}^{n \times n}$ such that $\boldsymbol{P^{-1}AP}$ is a diagonal matrix $\boldsymbol{D} \in \mathcal{F}^{n \times n}$

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

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

We can write this as

$$AP = PD$$

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

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

From this we get

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

Comparing the jth columns on both sides we get,

$$AP_j \ = \ \lambda_j P_j \text{ for } 1 \leq j \leq n \tag{5.1.1}$$

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

Conclusion :

$A \in \mathcal{F}^{n\times n}$ is diagonalizable over $\mathcal{F}$

$\Longrightarrow$

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

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

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

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

then we can define $P \in \mathcal{F}^{n\times n}$ 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 5.1.1 **$A \in \mathcal{F}^{n\times n}$ is diagonalizable over $\mathcal{F}$**

$\Longleftrightarrow$

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

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

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

Remark 5.1.1 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.

5.2 Eigenvalues and Eigenvectors

We begin with the definition of eigenvalues and eigenvectors.

Definition 5.2.1 A scalar $\boldsymbol{\lambda} \in \mathcal{F}$ is said to be an eigenvalue of a matrix $\boldsymbol{A} \in \mathcal{F}^{n\times n}$ if there exists a nonzero vector $\boldsymbol{\varphi} \in \mathcal{F}^n$ such that

$$\boldsymbol{A\varphi} \;=\; \boldsymbol{\lambda\varphi} \tag{5.2.1}$$

If $\boldsymbol{\lambda} \in \mathcal{F}$ is an eigenvalue of $\boldsymbol{A}$ then any nonzero vector $\boldsymbol{\varphi} \in \mathcal{F}^n$ such that $\boldsymbol{A\varphi} = \boldsymbol{\lambda\varphi}$ is called an eigenvector corresponding to the eigenvalue $\boldsymbol{\lambda}$. We shall call an eigenvalue-eigenvector pair $(\boldsymbol{\lambda}, \boldsymbol{\varphi})$, as an eigenpair

Given any $A \in \mathcal{F}^{n\times n}$ can we find n such eigenpairs in which the vectors in the n pairs are all linearly independent? We shall first look at some examples

Example 5.2.1 Consider the matrix $A \in \mathbb{R}^{2\times 2}$ defined as follows:

$$A \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \tag{5.2.2}$$

Consider $\lambda_1 = 1 \in \mathbb{R}$ and the nonzero vector $\varphi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{R}^2$. Then we have

$$\begin{aligned} A\varphi_1 &= \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= 1\varphi_1 \\ &= \lambda_1\varphi_1 \end{aligned}$$

Hence $(1, \varphi_1)$ is one eigenpair for this matrix. Next, consider $\lambda_2 = 2$ and $\varphi_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then we have

$$A\varphi_1 \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\
&= 2\varphi_2 \\
&= \lambda_2 \varphi_2
\end{aligned}
$$

Thus we see that $(2, \varphi_2)$ is another eigenpair. Moreover φ_1, φ_2 are linearly independent in $\mathbb{R}^2$. Hence we have two eigenpairs as required. Hence the matrix A is diagonalizable over $\mathbb{R}$. If we define

$$
P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}
$$

then it is easy to check that $P \in \mathbb{R}^{2\times 2}$ is invertible and

$$
\begin{aligned}
P^{-1}AP &= \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \text{ a diagonal matrix in } \mathbb{R}^{2\times 2}
\end{aligned}
$$

Example 5.2.2 Consider the matrix $A \in \mathbb{R}^{2\times 2}$ defined as follows:

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

We shall see that this matrix is not diagonalizable over $\mathbb{R}$. Suppose A is diagonalizable over $\mathbb{R}$. Then there must exist an invertible $P \in \mathbb{R}^{2\times 2}$ such that $P^{-1}AP$ is a diagonal matrix $D \in \mathbb{R}^{2\times 2}$. Let

$$
\begin{aligned}
P &= \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in \mathbb{R}^{2\times 2} \\
D &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \in \mathbb{R}^{2\times 2}
\end{aligned}
$$

Then we have

$$
\begin{aligned}
P^{-1}AP &= D \\
&\Longrightarrow \\
AP &= PA
\end{aligned}
$$

$$
\begin{aligned}
& \Longrightarrow \\
\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} p & q \\ r & s \end{pmatrix} &= \begin{pmatrix} p & q \\ r & s \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\
& \Longrightarrow \\
\begin{pmatrix} -r & -s \\ p & q \end{pmatrix} &= \begin{pmatrix} \lambda_1 p & \lambda_2 q \\ \lambda_1 r & \lambda_2 s \end{pmatrix}
\end{aligned}
$$

Comparing the first column on both sides we get

$$
\begin{aligned}
-r &= \lambda_1 p \\
p &= \lambda_1 r \\
\Longrightarrow & \\
-r &= \lambda_1^2 r \\
\Longrightarrow & \\
(1+\lambda_1^2) r &= 0 \\
\Longrightarrow & \\
r &= 0 \text{ (since } \lambda_1 \text{ is real)}
\end{aligned}
$$

Similarly, comparison of the second columns gives

$$
s = 0
$$

Hence the second row of P is zero row and hence P is not invertible - a contradiction. Thus this matrix is not diagonalizable over $\mathbb{R}$

Example 5.2.3 Consider the same matrix A of the above example but now we treat A as a matrix in $\mathbb{C}^{2\times 2}$. Consider $\lambda_1 = i$ and $\varphi_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix}$. Then we have

$$
\begin{aligned}
A\varphi_1 &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ -i \end{pmatrix} \\
&= \begin{pmatrix} i \\ 1 \end{pmatrix} \\
&= i\begin{pmatrix} 1 \\ -i \end{pmatrix} \\
&= i\varphi_1
\end{aligned}
$$

Thus we see that (i, φ_1) is an eigenpair. Similarly, we can verify that $(-i, \varphi_2)$ is an eigenpair, where, $\varphi_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}$. We further note that φ_1, φ_2 are linearly independent in $\mathbb{C}^2$. Thus if we define

$$P = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$$

it is easy to verify that

$$P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \text{ a diagonal matrix over } \mathbb{C}^{2\times 2}$$

Thus A is diagonalizable over $\mathbb{C}$

From the above examples, it follows that we have the following possibilities:

1. There are some matrices for which we can get these requisite number of eigenpairs and hence these matrices are diagonalizable

2. There are some matrices for which we cannot get these requisite number of eigenpairs and hence we cannot diagonalize these

3. There are some real matrices which are diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$

The main question that remains still is that of finding these eigenpairs. If we know an eigenvalue λ then we can find the possible eigenvectors corresponding to this eigenvalue by solving the homogeneous system

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

Hence we search for the eigenvalues. How do we locate these eigenvalues? We shall discuss this next.

5.3 Characteristic Polynomial and Algebraic Multiplicity

Where should we look for the eigenvalues of a matrix? We have the following:

λ is an eigenvalue of $A \in \mathcal{F}^{n\times n} \iff \exists \varphi \in \mathcal{F}^n \ni A\varphi = \lambda\varphi$

$$\begin{aligned}
&\iff A_\lambda\varphi = \theta_n \\
&\iff \text{The homogeneous system } A_\lambda x = \theta_n \text{ has} \\
&\qquad \text{nontrivial solution } \varphi \\
&\iff |A_\lambda| = 0 \text{ where } A_\lambda = \lambda I - A
\end{aligned}$$

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

$$c_{_A}(\lambda) \;=\; |\lambda I - A| \tag{5.3.1}$$

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,

$$c_{_A}(\lambda) \;=\; \lambda^n - (Trace(A))\lambda^{(n-1)} + \cdots + (-1)^n|A| \tag{5.3.2}$$

where

$$Trace(A) \;=\; \sum_{i=1}^{n} a_{ii} \text{ (sum of the diagonal entries of } A) \tag{5.3.3}$$

$c_{_A}(\lambda)$ is a MONIC polynomial of degree n over $\mathcal{F}$, that is, $c_{_A}(\lambda) \in \mathcal{F}[\lambda]$. 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 $\mathcal{F}$. We shall now look at some examples.

Example 5.3.1 Consider the matrix $A \in \mathbb{R}^{2\times 2}$ defined below:

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

We have

$$
\begin{aligned}
c_A(\lambda) &= |\lambda I - A| \\
&= \begin{vmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 2 \end{vmatrix} \\
&= (\lambda - 1)(\lambda - 2)
\end{aligned}
$$

Hence we have

$$
c_A(\lambda) = (\lambda - 1)(\lambda - 2)
$$

The roots are $\lambda_1 = 1$ and $\lambda_2 = 2$ and both are real. Thus we are able to get two eigenvalues in $\mathbb{R}$ for this $A \in \mathbb{R}^{2\times 2}$.

Example 5.3.2 Consider the matrix $A \in \mathbb{R}^{2\times 2}$ defined below:

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

We have

$$
\begin{aligned}
c_A(\lambda) &= |\lambda I - A| \\
&= \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} \\
&= \lambda^2 + 1
\end{aligned}
$$

Thus the characteristic polynomial is given by

$$
c_A(\lambda) = \lambda^2 + 1
$$

The polynomial has no real roots and hence there are no eigenvalues in $\mathbb{R}$ for this matrix.
However, if we consider A as matrix in $\mathbb{C}^{2\times 2}$ then we have the two roots for the characteristic polynomial as $\lambda_1 = i$ and $\lambda_2 = -i$. Thus this matrix has eigenvalues over the field $\mathbb{C}$ but not over the field $\mathbb{R}$.

Example 5.3.3 Consider the matrix $A \in \mathbb{R}^{2\times 2}$ defined below:

$$
A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}
$$

We have

$$\begin{aligned} c_{_A}(\lambda) &= |\lambda I - A| \\ &= \begin{vmatrix} \lambda & -1 \\ 0 & \lambda \end{vmatrix} \\ &= \lambda^2 \end{aligned}$$

Thus the characteristic polynomial is given by

$$c_{_A}(\lambda) \;=\; \lambda^2$$

Hence the characteristic equation is

$$\lambda^2 \;=\; 0$$

It has repeated roots $\lambda_1 = \lambda_2 = 0$

From the above examples, it follows that, in general , the characteristic polynomial of $A \in \mathcal{F}^{n\times n}$ need not have n roots in $\mathcal{F}$, and that when there are roots these may be repeated. However, consider a diagonalisable matrix $A \in \mathcal{F}^{n\times n}$. Then we have an invertible matrix $P \in \mathcal{F}^{n\times n}$ such that $P^{-}AP = D$ where $D \in \mathcal{F}^{n\times n}$ is a diagonal matrix, whose diagonal entries, $d_1, d_2, \cdots, d_n$ are in $\mathcal{F}$ and may not be distinct. So we have,

$$\begin{aligned} A &= PDP^{-1} \\ &= P \begin{pmatrix} d_1 & & & & \\ & d_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & d_n \end{pmatrix} P^{-1} \\ \Longrightarrow & \\ c_{_A}(\lambda) &= \left|\lambda I - PDP^{-1}\right| \\ &= \left|\lambda PIP^{-1} - PDP^{-1}\right| \\ &= \left|P(\lambda I - D)P^{-1}\right| \\ &= |P|\,|\lambda I - D|\left|P^{-1}\right| \\ &= c_{_A}(D) \text{ (since } |P|\,|P^{-1}| = 1) \\ &= (\lambda - d_1)(\lambda - d_2)\cdots(\lambda - d_n) \end{aligned}$$

Hence we have that for a diagonalizable matrix $A \in \mathcal{F}^{n\times n}$, the characteristic polynomial, $c_{_A}(\lambda)$, can be completely factorized into linear factors, (some of which may be repeated). Hence for a diagonalisable matrix the characteristic polynomial can be written as

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

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct roots in $\mathcal{F}$ and $a_1, a_2, \cdots, a_k$ are their multiplicities. We shall, therefore, first look at matrices for which the characteristic polynomial can be factorized as above.
Thus we consider the following type of matrices for the rest of this chapter:
$A \in \mathcal{F}^{n\times n}$ such that the characteristic polynomial of A is of the form,

$$c_{_A}(\lambda) \;=\; (\lambda - \lambda_1)^{a_1}(\lambda - \lambda_2)^{a_2} \cdots (\lambda - \lambda_k)^{a_k} \tag{5.3.4}$$

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

$$a_1 + a_2 + \cdots + a_k \;=\; n \tag{5.3.5}$$

(Clearly, this is possible for all matrices in $\mathbb{C}^{n\times n}$).
For such matrices $\lambda_1, \lambda_2, \cdots, \lambda_k$ 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.

5.4 Eigenspaces and Geometric Multiplicity

Let $A \in \mathcal{F}^{n\times n}$ and let its characteristic polynomial be as in (5.3.4). Hence the distinct eigenvalues of A are $\lambda_1, \lambda_2, \cdots, \lambda_k$ with algebraic multiplicities, $a_1, a_2, \cdots, a_k$ respectively. Having found the eigenvalues we next look at the eigenvectors. We define

$$\mathcal{W}_j \;=\; Null\ Space\ of\ A - \lambda_j I \tag{5.4.1}$$

$$\;=\; \{x \in \mathcal{F}^n : Ax = \lambda_j x\} \tag{5.4.2}$$

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 $\mathcal{F}^n$. 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

$$\text{geometric multiplicity, } g_j \quad = \quad dimension\ of\ \mathcal{W}_j \tag{5.4.3}$$

We now look at some simple examples.

Example 5.4.1 For the matrix A in the Example 5.3.1 we have

$$\lambda_1 = 1,\ \lambda_2 = 2,\ a_1 = 1,\ a_2 = 1$$

For the eigenspaces we have,

$$\begin{aligned}
\mathcal{W}_1 &= Null\ Space\ of\ A_I \\
&= Null\ Space\ of \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \\
&= \left\{ x \in \mathbb{R}^2 : x = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \alpha \in \mathbb{R} \right\} \\
\mathcal{W}_2 &= Null\ Space\ of\ A - 2I \\
&= Null\ Space\ of \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \\
&= \left\{ x \in \mathbb{R}^2 : x = \begin{pmatrix} \beta \\ \beta \end{pmatrix}, \beta \in \mathbb{R} \right\}
\end{aligned}$$

We see that

$$\begin{aligned}
g_1 &= dimension\ of\ \mathcal{W}_1 = 1 = a_1 \\
g_2 &= dimension\ of\ \mathcal{W}_2 = 1 = a_2
\end{aligned}$$

Thus, in this case, for each eigenvalue, the algebraic multiplicity is the same as the geometric multiplicity.

Example 5.4.2 For the matrix A of Example 5.3.3 we have $\lambda_1 = 0$ is the only eigenvalue and its algebraic multiplicity is $a_1 = 2$. For the eigenspace we have

$$\begin{aligned}
\mathcal{W}_1 &= Null\ Space\ of\ A - 0I \\
&= Null\ Space\ of\ A
\end{aligned}$$

$$
\begin{aligned}
&= \quad Null\ Space\ of \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\
&= \quad \left\{ x \in \mathbb{R}^2 : x = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \alpha \in \mathbb{R} \right\}
\end{aligned}
$$

Thus the dimension of $\mathcal{W}_j$ is 1 and hence we heve $g_1 = 1$. Thus, in this case we have an eigenvalue whose geometric multiplicity is smaller than its algebraic multiplicity.

In the above two examples we found that in one example we have the geometric multiplicity of every eigenvalue same as the algebraic multiplicity, and in another example there is an eigenvalue with its geometric multiplicity smaller than its algebraic multiplicity. We shall now see that these are the only two possibilities and that the geometric multiplicity of no eigenvalue can exceed its algebraic multiplicity.

Since the dimension of $\mathcal{W}_j$ is g_j, any basis for $\mathcal{W}_j$ consists of g_j vectors. Let

$$\mathcal{B}_j \quad : \quad \varphi_1, \varphi_2, \cdots, \varphi_{g_j}$$

be a basis for $\mathcal{W}_j$. We can extend this to a basis for $\mathcal{F}^n$ by appending $n - g_j$ suitable linearly independent vectors. Let

$$\mathcal{B} \quad : \quad \varphi_1, \varphi_2, \cdots, \varphi_{g_j}, v_1, v_2, \cdots, v_{(n-g_j)}$$

be a basis for $\mathcal{F}^n$ which is an extension of $\mathcal{B}_j$. Let P be the matrix whose columns are these n basis vectors. We have

$$P \quad = \quad \left[\varphi_1\ \varphi_2 \cdots\ \varphi_{g_j}\ v_1\ v_2 \cdots\ v_{(n-g_j)}\right]$$

Clearly P is invertible since the columns are all linearly independent. Further, we have,

$$
\begin{aligned}
AP &= \left[A\varphi_1\ A\varphi_2 \cdots\ A\varphi_{g_j}\ Av_1\ Av_2 \cdots\ Av_{(n-g_j)}\right] \\
&= \left[\lambda_j\varphi_1\ \lambda_j\varphi_2 \cdots\ \lambda_j\varphi_{g_j}\ Av_1\ Av_2 \cdots\ Av_{(n-g_j)}\right] \\
&= \left[\varphi_1\ \varphi_2 \cdots\ \varphi_{g_j}\ v_1\ v_2 \cdots\ v_{(n-g_j)}\right] \left(\begin{array}{c|c} \lambda_j I_{(g_j \times g_j)} & \mathrm{K}_{g_j \times (n-g_j)} \\ \hline 0_{(n-g_j)\times g_j} & \mathrm{L}_{(n-g_j)\times(n-g_j)} \end{array}\right)
\end{aligned}
$$

$\Longrightarrow$

$$A = P \left(\begin{array}{c|c} \lambda_j I_{(g_j \times g_j)} & \mathrm{K}_{g_j \times (n-g_j)} \\ \hline 0_{(n-g_j)\times g_j} & \mathrm{L}_{(n-g_j)\times(n-g_j)} \end{array} \right) P^{-1}$$

Hence we get

$$c_{_A}(\lambda) = (\lambda - \lambda_j)^{g_j} c_{_L}(\lambda)$$

Thus we see that λ_j must be a root of $c_{_A}(\lambda)$ of multiplicity at least g_j. Thus we have $g_j \leq a_j$. Further, since the subspace $\mathcal{W}_j$ has at least one nonzero vector, (because λ_j is an eigenvalue), it follows that $1 \leq g_j$. Thus we have

$$1 \leq g_j \leq a_j \text{ for } 1 \leq j \leq k \tag{5.4.4}$$

Thus ,
The geometric multiplicity of any eigenvalue is less than or equal to its algebraic multiplicity, and the geometric multiplicity is a t least one.

5.5 Diagonalizability

If A is a diagonalizable matrix, then there exists an invertible P such that

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

Hence we get

$$AP = Pdiag(\underbrace{\lambda_1, \lambda_1 \cdots, \lambda_1}_{a_1\ times}, \underbrace{\lambda_2, \lambda_2, \cdots, \lambda_2}_{a_2\ times}, \cdots, \underbrace{\lambda_k, \lambda_k, \cdots, \lambda_k}_{a_k\ times})$$

From this we get that the,
first a_1 columns of P are linearly independent eigenvectors of A corresponding to eigenvalue λ_1,
next a_2 columns of P are linearly independent eigenvectors of A corresponding to eigenvalue λ_2,
$\cdots \cdots, \cdots \cdots, \cdots \cdots$

last a_k columns of P are linearly independent eigenvectors of A corresponding to eigenvalue λ_k.
Thus it follows that the dimension of $\mathcal{W}_j$ is at least a_j and hence $a_j \leq g_j$. But by (5.4.4) we have $g_j \leq a_j$. From these two we get

$$\boldsymbol{A \in \mathcal{F}^{n\times n}} \textbf{ is a diagonalizable matrix} \quad \Longrightarrow \quad \begin{array}{l}\boldsymbol{a_j = g_j} \textbf{ for every eigenvalue,} \\ \textbf{(algebraic multiplicity=} \\ \textbf{geometric multiplicity} \\ \textbf{for every eigenvalue)}\end{array} \tag{5.5.1}$$

We shall now see that the converse is also true.
Suppose $A \in \mathcal{F}^{n\times n}$ and its characteristic polynomial is as in (5.3.4). Let $\mathcal{W}_j$ be the eigenspace corresponding to the eigenvalue λ_j. Suppose the matrix is such that for every eigenvalue, the algebraic multiplicity is equal to its geometric multiplicity. Hence the dimension of $\mathcal{W}_j$ is a_j and we can find a basis

$$\mathcal{B}_j \quad : \quad v_1^{(j)}, v_2^{(j)}, \cdots, v_{a_j}^{(j)}$$

for $\mathcal{W}_j$. We now make a claim which we shall prove in the next section.
<u>**CLAIM 1**</u>:
If $\boldsymbol{x_1 + x_2 + \cdots + x_k = \theta_n}$, where $\boldsymbol{x_j \in \mathcal{W}_j}$ for each $\boldsymbol{j}$, then $\boldsymbol{x_j = \theta_n}$ for each $\boldsymbol{j}$
Using the above Claim we shall prove the following:
<u>**CLAIM 2**</u>:
$\boldsymbol{\mathcal{B} = \mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_k}$ is a basis for $\boldsymbol{\mathcal{F}^n}$
Suppose we have the above two claims. Then we construct the matrix P, whose,
first a_1 columns are $v_1^{(1)}, v_2^{(1)}, \cdots, v_{a_1}^{(1)}$,
next a_2 columns are $v_1^{(2)}, v_2^{(2)}, \cdots, v_{a_2}^{(1)}$, etc. and,
the last a_k columns are $v_1^{(k)}, v_2^{(k)}, \cdots, v_{a_1}^{(k)}$.
Since the columns of P are from the basis $\mathcal{B}$, it follows they are linearly independent. Hence P is invertible. We easily see that

$$\begin{aligned} AP &= P\, diagonal(\underbrace{\lambda_1, \lambda_1, \cdots, \lambda_1}_{a_1\ times}\ \underbrace{\lambda_2, \lambda_2, \cdots, \lambda_2}_{a_2\ times}\ \cdots\ \underbrace{\lambda_k, \lambda_k, \cdots, \lambda_k}_{a_k\ times}) \\ \Longrightarrow & \end{aligned}$$

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

Hence A is a diagonalisable matrix. Thus we have,
Let $A \in \mathcal{F}^{n\times n}$ and $c_A(\lambda)$ as in (5.3.4). Then

$$\left\{\begin{array}{l}\textbf{algebraic multiplicity} = \\ \textbf{geometric multiplicity} \\ \textbf{for every eigenvalue}\end{array}\right\} \Longrightarrow \boldsymbol{A} \textbf{ is diagonalizable over } \boldsymbol{\mathcal{F}} \tag{5.5.2}$$

Combining (5.5.1) and (5.5.2) we get,

Theorem 5.5.1 Let $A \in \mathcal{F}^{n\times n}$ and the characteristic polynomial of A is as in (5.3.4). Then

$$\begin{array}{lcl}\boldsymbol{A} \textbf{ is a diagonalizable matrix} & \Longleftrightarrow & \boldsymbol{a_j = g_j} \textbf{ for every eigenvalue,} \\ & & \textbf{(algebraic multiplicity=} \\ & & \textbf{geometric multiplicity} \\ & & \textbf{for every eigenvalue)}\end{array} \tag{5.5.3}$$

We shall now prove Claim 2 using Claim 1.
Proof of CLAIM 2:
In order to prove that $\mathcal{B}$ is a basis for $\mathcal{F}^n$ it is enough to show that $\mathcal{B}$ is linearly independent, (since any linearly independent set in $\mathcal{F}^n$ having n vectors will form a basis for $\mathcal{F}^n$). We have,

$$\begin{array}{rcl}\sum_{j=1}^{k}\sum_{r=1}^{a_j}\alpha_r^{(j)}v_r^{(j)} & = & \theta_n \\ \Longrightarrow & & \\ \sum_{j=1}^{k} x_j & = & \theta_n \text{ where } x_j = \sum_{r=1}^{a_j}\alpha_r^{(j)}v_r^{(j)} \in \mathcal{W}_j \\ \Longrightarrow & & \text{(by Claim 1)} \\ x_j & = & \theta_n \text{ for } 1 \leq j \leq k \\ \Longrightarrow & & \end{array}$$

$$\begin{aligned}
\sum_{r=1}^{a_j} \alpha_r^{(j)} v_r^{(j)} &= \theta_n \text{ for each } j,\ 1 \le j \le k \\
\Longrightarrow & \quad \text{(since } \mathcal{B}_j \text{ is a linearly independent set)} \\
\alpha_r^{(j)} &= 0 \text{ for } 1 \le r \le a_j, \text{ for each } j,\ 1 \le j \le k
\end{aligned}$$

Hence we get $\mathcal{B}$ is linearly independent, thus proving Claim 2.

5.6 Lagrange Interpolation Polynomials

Consider k distinct elements, $\lambda_1,\ \lambda_2, \cdots, \lambda_k$, in a field $\mathcal{F}$. (For any nonzero element $a \in \mathcal{F}$ we shall write its multiplicative inverse a^{-1} also as $\frac{1}{a}$). Any polynomial of degree $(k-1)$ is completely determined by its values at the k points $\lambda_1, \lambda_2, \cdots, \lambda_k$. Let $\ell_j(\lambda)$ be the polynomial of degree $k-1$ such that

$$\ell_j(\lambda_i) = \begin{cases} 0 & \text{if} \quad i \ne j \\ 1 & \text{if} \quad i = j \end{cases}$$

From the definition of $\ell_j(\lambda)$ above it follows that all the λ_i except λ_j are roots of $\ell_j(\lambda)$. Hence $\ell_j(\lambda)$ must have the factor

$$(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_{(j-1)})(\lambda - \lambda_{(j+1)}) \cdots (\lambda - \lambda_k)$$

Since $\ell_j(\lambda)$ is of degree $(k-1)$ it will be of the form

$$\alpha(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_{(j-1)})(\lambda - \lambda_{(j+1)}) \cdots (\lambda - \lambda_k)$$

where $\alpha \in \mathcal{F}$. The requirement that $\ell_j(\lambda_j) = 1$ now gives

$$\alpha(\lambda_j - \lambda_1)(\lambda_j - \lambda_2) \cdots (\lambda_j - \lambda_{(j-1)})(\lambda_j - \lambda_{(j+1)}) \cdots (\lambda_j - \lambda_k) \quad = \quad 1$$

Hence we get

$$\alpha \quad = \quad \frac{1}{(\lambda_j - \lambda_1)(\lambda_j - \lambda_2) \cdots (\lambda_j - \lambda_{(j-1)})(\lambda_j - \lambda_{(j+1)}) \cdots (\lambda_j - \lambda_k)}$$

Hence we get

$$\ell_j(\lambda) \quad = \quad \frac{(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_{(j-1)})(\lambda - \lambda_{(j+1)}) \cdots (\lambda - \lambda_k)}{(\lambda_j - \lambda_1)(\lambda_j - \lambda_2) \cdots (\lambda_j - \lambda_{(j-1)})(\lambda_j - \lambda_{(j+1)}) \cdots (\lambda_j - \lambda_k)}$$

We write this in short form as

$$\ell_j(\lambda) \;=\; \prod_{\substack{i=1\\ i\neq j}}^{k} \frac{(\lambda-\lambda_i)}{(\lambda_j-\lambda_i)}, \text{ for } 1 \leq j \leq k \tag{5.6.1}$$

Thus we have k polynomials of degree $(k-1)$, as defined , such that,

$$\ell_j(\lambda_i) \;=\; \begin{cases} 0 & \text{if} \quad i \neq j \\ 1 & \text{if} \quad i = j \end{cases} \tag{5.6.2}$$

These are called the **Lagrange Interpolation Polynomials** with respect to the distinct points $\lambda_1, \lambda_2, \cdots, \lambda_k$.
We shall now use these polynomials to prove Claim 1 of the previous section. First we observe the following property of the vectors in the eigenspace $\mathcal{W}_j$ corresponding to the eigenvalue λ_j. We have,

$$\begin{aligned} x_j \in \mathcal{W}_j \;&\Longrightarrow\; Ax_j = \lambda_j x_j \\ &\Longrightarrow\; A^2 x_j = \lambda_j^2 x_j \end{aligned}$$

In general, we get, for any nonnegative integer, m,

$$x_j \in \mathcal{W}_j \;\Longrightarrow\; Ax_j = \lambda_j^m x_j \tag{5.6.3}$$

Hence for any polynomial $p(\lambda) \in \mathcal{F}[\lambda]$ we get

$$x_j \in \mathcal{W}_j \;\Longrightarrow\; p(A)x_j = p(\lambda_j)x_j \tag{5.6.4}$$

In particular, we get, by (5.6.2),

$$\begin{aligned} \ell_i(A)x_j \;&=\; \ell_i(\lambda_j)x_j \\ &=\; \left\{ \begin{array}{lll} \theta_n & \text{if} & i \neq j \\ x_j & \text{if} & i = j \end{array} \right\} \end{aligned}$$

Thus we have, for $1 \leq i \leq k$,

$$\ell_i(A)(x_j) \;=\; \left\{ \begin{array}{lll} \theta_n & \text{if} & i \neq j \\ x_j & \text{if} & i = j \end{array} \right\} \text{ for every } x_j \in \mathcal{W}_j \tag{5.6.5}$$

We shall now use the above property to prove Claim 1 of the previous section.

$$\begin{aligned} x_1 + x_2 + \cdots + x_k &= \theta_n \\ &\Longrightarrow \\ \ell_j(A)(x_1 + x_2 + \cdots + x_k) &= \theta_n \\ &\Longrightarrow \\ x_j &= \theta_n \text{ by } (5.6.5) \end{aligned}$$

Thus we have

$$x_1 + x_2 + \cdots + x_k = \theta_n \Longrightarrow x_j = \theta_n, \ 1 \leq j \leq k$$

thus proving Claim 1

5.7 Decompositions for a Diagonalizable Matrix

We shall now discuss some decompositions that arise in the case of a diagonalizable matrix. As before let $A \in \mathcal{F}^{n\times n}$ be such that its characteristic polynomial is as in (5.3.4) and let A be diagonalisable. We have seen that (Claim 2 Section 5.5) if $\mathcal{B}_j = v_1^{(j)}, v_2^{(j)}, \cdots, v_{a_j}^{(j)}$ is a basis for the eigenspace $\mathcal{W}_j$ corresponding to the eigenvalue λ_j, then $\mathcal{B} = \mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_k$ is a basis for $\mathcal{F}^n$. Thus we have the first decomposition as follows:

Decomposition of Basis

Basis for $\mathcal{F}^n$ can be decomposed as k smaller bases one each for $\mathcal{W}_j$

We next observe that any $x \in \mathcal{F}^n$ can be expressed as a linear combination of these basis vectors. Thus,

$$\begin{aligned} x \in \mathcal{F}^n &\Longrightarrow \\ x &= \sum_{j=1}^{k}\sum_{r=1}^{a_j} \alpha_r^{(j)} v_r^{(j)} \\ &= \sum_{j=1}^{k} x_j \text{ where } x_j = \sum_{r=1}^{a_j} \alpha_r^{(j)} v_r^{(j)} \in \mathcal{W}_j \end{aligned}$$

Thus we see that every $x \in \mathcal{F}^n$ can be decomposed as the sum of k vectors, one each from $\mathcal{W}_j$. Further, using Claim 1 of Section 5. we can see that this decomposition is unique. Thus we have the next decomposition as follows:

Decomposition of a vector:

Every $\boldsymbol{x} \in \boldsymbol{\mathcal{F}^n}$ can be decomposed uniquely as the sum of a vector each from $\boldsymbol{\mathcal{W}_j}$, that is,
$\boldsymbol{x \in \mathcal{F}^n \Longrightarrow \exists}$ unique vectors $\boldsymbol{x_1, x_2, \cdots, x_k}$ such that $\boldsymbol{x_j \in \mathcal{W}_j}$ for $\boldsymbol{1 \leq j \leq k}$, and $\boldsymbol{x = x_1 + x_2 + \cdots = x_k}$

We shall now look at a consequence of this. Consider the polynomial $m(\lambda)$ defined as,

$$m(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_k) \tag{5.7.1}$$

We can write, for each j, $1 \leq j \leq l$,

$$m(\lambda) = q_j(\lambda)(\lambda - \lambda_j) \tag{5.7.2}$$

and hence

$$m(A) = q_j(A)(A - \lambda_j I) \tag{5.7.3}$$

Hence we have,

$$\begin{aligned} x \in \mathcal{F}^n &\Longrightarrow x = x_1 + x_2 + \cdots + x_j + \cdots + x_k \text{ a unique decomposition where} x_j \in \mathcal{W}_j \\ &\Longrightarrow \\ m(A)x &= \sum_{j=1}^{k} m(A)x_j \\ &= \sum_{j=1}^{k} q_j(A)(A - \lambda_j I)x_j \text{ (by (5.7.3))} \\ &= \theta_n \text{ since } x_j \in \mathcal{W}_j \end{aligned} \tag{5.7.5}$$

Thus we have

$$m(A)x = \theta_n \text{ for all } x \in \mathcal{F}^n \tag{5.7.6}$$

We next look at he decomposition of the matrix. As obtained in the previous section we have the Lagrange Polynomials, $\{\ell_j(\lambda)\}_{j=1}^{k}$ defined as

$$\ell_j(\lambda) = \prod_{\substack{i=1 \\ i\neq j}}^{k} \frac{(\lambda-\lambda_i)}{(\lambda_j-\lambda_i)}$$

These polynomials have the following properties:

$$\ell_1(\lambda)+\ell_2(\lambda)+\cdots+\ell_k(\lambda) = 1 \tag{5.7.7}$$
$$\lambda_1\ell_1(\lambda)+\lambda_2\ell_2(\lambda)+\cdots+\lambda_k\ell_k(\lambda) = \lambda \tag{5.7.8}$$

We now define k matrices $A_j \in \mathcal{F}^{n\times n}$ as follows:

$$A_j = \ell_j(A) \tag{5.7.9}$$

Since the matrices A_j are all polynomials in A they commute with each other, that is,

$$A_iA_j = A_jA_i \text{ for all } i \text{ and } j \tag{5.7.10}$$

From (5.7.7) and (5.7.8) we get the following:

$$A_1+A_2+\cdots+A_k = I_{n\times n} \tag{5.7.11}$$
$$\lambda_1A_1+\lambda_2A_2+\cdots+\lambda_kA_k = A \tag{5.7.12}$$

Further, we have, for $i\neq j$, $\ell_i(\lambda)$ has factor $(\lambda-\lambda_j)$ and hence $\ell_i(A)$ has factor $(A-\lambda_jI)$. Thus we have,

$$\ell_i(\lambda) = q_{ij}(\lambda)(\lambda-\lambda_j)\text{for all } i\neq j, \text{ and} \tag{5.7.13}$$
$$\ell_i(A) = q_{ij}(A)(A-\lambda_jI) \text{ for all } i\neq j \tag{5.7.14}$$

Hence we get

$$x\in\mathcal{W}_j \implies A_i(x)=\theta_n \text{ for all } i\neq j \tag{5.7.15}$$

From this we get,

$$\begin{aligned} x\in\mathcal{W}_j &\implies x = Ix \\ &\implies \\ x &= (A_1+A_2+\cdots+A_j+\cdots+A_k)x \text{ (by (5.7.11))} \\ &\implies \\ x &= A_jx \text{ (by (5.7.15)} \end{aligned}$$

Thus

$$x \in \mathcal{W}_j \implies A_j x = x \tag{5.7.16}$$

From the above, we have,

$$x \in \mathcal{W}_j \implies x \in \mathcal{R}_{A_j} \tag{5.7.17}$$

We also observe that

$$\begin{array}{rcl} x \in \mathcal{R}_{A_j} & \implies & x = A_j y \text{ for some } y \in \mathcal{F}^n \\ & \implies & \\ (A-\lambda_j I)x & = & (A-\lambda_j I)A_j x \\ & = & p(A)m(A)x \text{ since } (\lambda-\lambda_j)\ell_j(\lambda) \text{ has a factor } m(\lambda) \\ & = & \theta_n \text{ by (5.7.6)} \\ & \implies & x \in \mathcal{W}_j \end{array}$$

Thus we have

$$x \in \mathcal{R}_{A_j} \implies x \in \mathcal{W}_j \tag{5.7.18}$$

Combining (5.7.17) and (5.7.18) we get

$$\mathcal{R}_{A_j} = \mathcal{W}_j \tag{5.7.19}$$

For $i \neq j$, $\ell_i(\lambda)\ell_j(\lambda)$ has a factor $m(\lambda)$, and hence we can write

$$\ell_i(\lambda)\ell_j(\lambda) = p_{ij} m(\lambda) \tag{5.7.20}$$

Hence we get

$$\begin{array}{rcl} \ell_i(A)\ell_j(A) & = & p_{ij}(A)m(A) \\ & \implies & \\ A_i A_j & = & p_{ij}(A)m(A) \implies \end{array}$$

$$x \in \mathcal{F}^n \implies A_i A_j x = p_{ij}(A)m(A)x = \theta_n \text{ by (5.7.6)}$$

Thus we have

$$A_i A_j = 0_{n\times n} \text{ for } i \neq j \tag{5.7.21}$$

Thus we have the following decomposition of the matrices $I_{n\times n}$ and A:
Decomposition of Matrix:
There exist k distinct matrices $A_1, A_2, \cdots, A_k$ and k scalars, (namely the distinct eigenvalues of A, $\lambda_1, \lambda_2, \cdots, \lambda_k$ such that

$$\boldsymbol{A_1 + A_2 + \cdots + A_k = I_{n\times n}} \tag{5.7.22}$$
$$\boldsymbol{\lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k = A} \tag{5.7.23}$$
$$\boldsymbol{A_i A_j = 0_{n\times n} \text{ for } i \neq j} \tag{5.7.24}$$

From the above, we can also easily see that

$$\boldsymbol{A_j^2 = A \text{ for all } j} \tag{5.7.25}$$

Further we have from (5.7.19),

$$\boldsymbol{\mathcal{R}_{A_j} = \mathcal{W}_j \text{ the eigenspace corresponding to } \lambda_j} \tag{5.7.26}$$

Conversely, If there exist k distinct matrices $A_1, A_2, \cdots, A_k \in \mathcal{F}^{n\times n}$ and k scalars, $\lambda_1, \lambda_2, \cdots, \lambda_k \in \mathcal{F}$ such that (5.7.22),(5.7.23) and (5.7.24) we define,

$$\mathcal{W}_j = \mathcal{R}_{A_j} \tag{5.7.27}$$

Then we can show the following: (We shall not give a proof of this, but we shall use this in the next section):

$$A_j^2 = I \text{ for each } j = 1, 2, \cdots, k \tag{5.7.28}$$
$$x \in \mathcal{F}^n \text{ unique } x_j \in \mathcal{W}_j \ni x = x_1 + x_2 + \cdots + x_k \tag{5.7.29}$$
$$\lambda_1, \lambda_2, \cdots, \lambda_k \text{ are precisely the distinct eigenvalues of } A \tag{5.7.30}$$
$$\mathcal{W}_j \text{ is the eigenspaces corresponding to } \lambda_j, j = 1, 2, \cdots, k \tag{5.7.31}$$
$$A \text{ is diagonalizable over } \mathcal{F} \tag{5.7.32}$$

Summarizing, we have the following theorem

Theorem 5.7.1 Let $A \in \mathcal{F}^{n\times n}$. Then

$$\boldsymbol{A \text{ is diagonalizable } \Longleftrightarrow \left\{\begin{array}{l} \text{there exist } k \text{ distinct matrices} \\ A_1, A_2, \cdots, A_k \in \mathcal{F}^{n\times n} \text{ and} \\ k \text{ scalars } \lambda_1, \lambda_2, \cdots, \lambda_k \in \mathcal{F} \\ \text{such that (5.7.22),(5.7.23) and (5.7.24) hold} \end{array}\right\}} \tag{5.7.33}$$

Further, then,(5.7.28),(5.7.29),(5.7.30),(5.7.31) also hold.

5.8 Minimal Polynomial

We shall now look at diagonalizability from the point of view of certain polynomials associated with a matrix. Given an $n \times n$ matrix consider the $n^2 + 1$ matrices,

$$I, A, A^2, \cdots, A^{n^2}.$$

Since these are $n^2 + 1$ elements in the n^2 dimensional vector space $\mathcal{F}^{n\times n}$, (of all $n \times n$ matrices), these must be linearly independent. Thus we must have

$$a_0 I + a_1 A + a_2 A^2 + \cdots + a_{n^2} A^{n^2} = 0$$

for suitable scalars $a_0, a_1, \cdots, a_{n^2}$ not all of which are zero. This means that there is a nontrivial polynomial of degree n^2, namely,

$$p(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{n^2}\lambda^{n^2}$$

such that it annihilates the matrix A, i.e.,

$$p(A) = 0_{n\times n}, \text{ (the } n \times n \text{ zero matrix)}$$

Thus given any $n \times n$ matrix there is at least one polynomial of degree n^2 that annhilates A. Clearly any polynomial which is a (nonzero) polynomial multiple of this $p(\lambda)$ will also annihilate A (and such a polynomial will have degree greater than or equal to n^2). The question is whether there are nontrivial polynomials of lower degree that annihilate A. To this end we have the well known

Theorem 5.8.1 CAYLEY-HAMILTON THEOREM
Every matrix satisfies its characteristic equation, that is, for any matrix $A \in \mathcal{F}^{n\times n}$,

$$c_A(A) = 0 \tag{5.8.1}$$

We know that the characteristic polynomial is of degree n and hence what the above theorem gives us is that there is certainly a nontrivial polynomial of degree n that annihilates A. Thus we have come down from degreee n^2 to degree n polynomial that annihilates A. The natural question is what is the lowest degree polynomial we can get that annihilates A. Before we try to find the answer to this question, we shall digress a little to recollect some properties of polynomials.
Let $\mathcal{F}[\lambda]$ be the collection of all polynomials in λ with coefficients from the field $\mathcal{F}$. We have the following properties:

1. $\mathcal{F}[\lambda]$ is a vector space over $\mathcal{F}$.

2. $p(\lambda), q(\lambda) \in \mathcal{F}[\lambda] \Rightarrow p(\lambda)q(\lambda) \in \mathcal{F}[\lambda]$

3. A polynomial is said to be a monic polynomial if the leading coefficient (the coefficient of the highest degree term) is 1.

4. (Division Algorithm): If $p_1(\lambda), p_2(\lambda) \in \mathcal{F}[\lambda]$ and degree of $p_2(\lambda)$ is less than or equal to that of $p_1(\lambda)$ we have polynomials $q(\lambda)$ and $r(\lambda)$ in $\mathcal{F}[\lambda]$ such that

$$p_1(\lambda) = q(\lambda)p_2(\lambda) + r(\lambda) \tag{5.8.2}$$

where

EITHER degree $r(\lambda) <$ degree $p_2(\lambda)$ OR $r(\lambda) = 0$ (the zero polynomial)

If $r(\lambda) = 0$ we say that $p_2(\lambda)$ is a divisor of $p_1(\lambda)$ or $p_1(\lambda)$ is divisible by $p_2(\lambda)$.

5. A polynomial of degree $k-1$ is uniquely determined by its values at k distinct points. Hence given k distinct points there exists a unque polynomial of degree $k-1$ that takes prescribed values at these points. The construction of such a polynomial is achieved through Lagrange Interpolation. Suppose $\lambda_1, \lambda_2, \cdots, \lambda_k$ are k distinct points. Then the polynomial $p_j(\lambda)$ that takes the value 1 at λ_j and 0 at all other λ_r (where $1 \leq r \leq k$ and $r \neq j$), is given by

$$p_j(\lambda) = \frac{(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_{j-1})(\lambda-\lambda_{j+1})\cdots(\lambda-\lambda_k)}{(\lambda_j-\lambda_1)(\lambda_j-\lambda_2)\cdots(\lambda_j-\lambda_{j-1})(\lambda_j-\lambda_{j+1})\cdots(\lambda_j-\lambda_k)} \tag{5.8.3}$$

We can easily see that

$$p_1(\lambda) + p_2(\lambda) + \cdots + p_k(\lambda) = 1 \tag{5.8.4}$$

$$\lambda_1 p_1(\lambda) + \lambda_2 p_2(\lambda) + \cdots + \lambda_k p_k(\lambda) = \lambda \tag{5.8.5}$$

and for any polynomial $p(\lambda)$ of degree $\leq (k-1)$.

$$p(\lambda_1)p_1(\lambda) + p(\lambda_2)p_2(\lambda) + \cdots + p(\lambda_k)p_k(\lambda) = p(\lambda) \tag{5.8.6}$$

6. If $f_1(\lambda), f_2(\lambda), \cdots, f_k(\lambda) \in \mathcal{F}[\lambda]$ then a polynomial $h(\lambda) \in \mathcal{F}[\lambda]$ is said to be a gcd for these polynomials if it is a common divisor of all these polynomials and the degree of any other common divisor cannot be greater than that of $h(\lambda)$. We call that gcd which is monic as <u>the</u> gcd.

7. The gcd of $f_1(\lambda), f_2(\lambda), \cdots, f_k(\lambda) \in \mathcal{F}[\lambda]$ can be expressed as a polynomial linear combination of these polynomials. This means that if $g_0(\lambda)$ is the gcd of $f_1(\lambda), f_2(\lambda), \cdots, f_k(\lambda)$ then there exist polynomials $g_1(\lambda), g_2(\lambda), \cdots, g_k(\lambda)$ such that

$$g_0(\lambda) \;=\; g_1(\lambda)f_1(\lambda) + g_2(\lambda)f_2(\lambda) + \cdots + g_k(\lambda)f_k(\lambda) \quad (5.8.7)$$

8. In particular, if $f_1(\lambda), f_2(\lambda), \cdots, f_k(\lambda)$ are coprime, i.e, have gcd as 1 then there exist polynomials $g_1(\lambda), g_2(\lambda), \cdots, g_k(\lambda)$ such that

$$1 \;=\; g_1(\lambda)f_1(\lambda) + g_2(\lambda)f_2(\lambda) + \cdots + g_k(\lambda)f_k(\lambda) \quad (5.8.8)$$

   In particular, if

   $f_j(\lambda) = (\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_{j-1})(\lambda-\lambda_{j+1})\cdots(\lambda-\lambda_k)$ for $1 \leq j \leq k$,

   where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are distinct then these polynomials are coprime and the $g_j(\lambda)$ are given by the Lagrange interpolation formula, as,

$$g_j(\lambda) \;=\; f_j(\lambda_j) = \left\{ \prod_{\substack{i=1 \\ i \neq j}}^{k} (\lambda_j - \lambda_i) \right\}^{-1} \quad (5.8.9)$$

We shall now continue our search for the lowest degree polynomial that annihilates A. We already know by Cayley-Hamilton theorem that the nth degree polynomial, $c(\lambda)$, namely the characteristic polynomial, annihilates A. Let $\mathcal{A}$ denote the set of all polynomials that annihilate A. We have

$$\mathcal{A} \;=\; \{p(\lambda) \in \mathcal{F}[\lambda] : p(A) = 0\} \quad (5.8.10)$$

Let $\hat{\mathcal{A}} = \mathcal{A} - \{\text{the zero polynomial}\}$. Thus

$$\hat{\mathcal{A}} = \{p(\lambda) \in \mathcal{F}[\lambda] : p(A) = 0 \text{ and } p(\lambda) \text{ nontrivial}\} \tag{5.8.11}$$

Let $s(\lambda)$ be in $\hat{\mathcal{A}}$ be such that it has lowest degree, i.e.,

$$\text{degree } s(\lambda) \leq \text{ degree } p(\lambda) \forall\, p(\lambda) \in \hat{\mathcal{A}}$$

Now if $p(\lambda) \in \hat{\mathcal{A}}$ then by the Division Algorithm (see Property 4 above) there exist polynomials $q(\lambda)$ and $r(\lambda)$ in $\mathcal{F}[\lambda]$ such that

$$p(\lambda) = q(\lambda)s(\lambda) + r(\lambda) \tag{5.8.12}$$

where

$$\text{EITHER degree } r(\lambda) < \text{ degree } s(\lambda) \text{ OR } r(\lambda) = 0$$

We have

$$r(\lambda) = p(\lambda) - q(\lambda)s(\lambda)$$

and hence

$$\begin{aligned} r(A) &= p(A) - q(A)s(A) \\ &= 0 (\text{since } p(A) = s(A) = 0 \text{ as } p(\lambda), s(\lambda) \in \hat{\mathcal{A}}) \end{aligned}$$

Thus if $r(\lambda)$ were not the zero polynomial this would mean that $r(\lambda) \in \hat{\mathcal{A}}$ and has degree less than that of $s(\lambda)$ contradicting the choice of $s(\lambda)$. Thus we must have $r(\lambda) = 0$. Hence from (5.8.12) we get that for any $p(\lambda) \in \hat{\mathcal{A}}$ there exists a corresponding polynomial $q(\lambda)$ such that

$$p(\lambda) = q(\lambda)s(\lambda) \tag{5.8.13}$$

Thus every polynomial in $\hat{\mathcal{A}}$ is a polynomial multiple of $s(\lambda)$. Let

$$m(\lambda) = \frac{s(\lambda)}{\alpha}$$

where α is the coefficient of the highest degree term in $s(\lambda)$. Thus $m(\lambda)$ is monic and has the same degree as $s(\lambda)$. Since every polynomial in $\hat{\mathcal{A}}$ is a polynomial multiple of $s(\lambda)$ it follows that every polynomial in $\hat{\mathcal{A}}$ is also a polynomial multiple of $m(\lambda)$, (since $s(\lambda)$ is itself a constant multiple of $m(\lambda)$). Thus $m(\lambda)$ is a monic polynomial in $\hat{\mathcal{A}}$ and has lowest degree. Suppose $m_1(\lambda)$ is any other monic polynomial in $\hat{\mathcal{A}}$ having lowest degree, then as before we have every polynomial in $\hat{\mathcal{A}}$ is a polynomial multiple of $m_1(\lambda)$. Thus $m(\lambda)$ and $m_1(\lambda)$ are both monic, have same degree and each is a polynomial multiple of the other and hence we get $m(\lambda) = m_1(\lambda)$. Thus $m(\lambda)$ is the unique monic polynomial in $\hat{\mathcal{A}}$ which has the lowest degree and every polynomial in $\hat{\mathcal{A}}$ is a polynomial multiple of this $m(\lambda)$. This polynomial is called the **Minimal Polynomial** of the matrix A. From now on we shall denote the minimal polynomial of A as $m_A(\lambda)$. We have

Definition 5.8.1 The minimal polynomial of a matrix $A \in \mathcal{F}^{n\times n}$ is the lowest degree monic polynomial in $\mathcal{F}[\lambda]$ that annihilates A

Clearly the zero polynomial is also a polynomial multiple of $m_A(\lambda)$, (namely zero polynomial multiple), and hence every polynomial in $\mathcal{A}$ (i.e., every polynomial that annihilates A) is a polynomial multiple of $m_A(\lambda)$. Since the characteristic polynomial $c_{A}(\lambda)$ annihilates A we must have $c_{A}(\lambda)$ is a polynomial multiple of $m_A(\lambda)$. If $\mathcal{F} = \mathbb{C}$ then the characteristic polynomial $c_{A}(\lambda)$ has the form given in (5.3.4) and hence the minimal polynomial being a divisor of $c_{A}(\lambda)$ must have the form

$$m_A(\lambda) \quad = \quad (\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k} \tag{5.8.14}$$

where

$$0 \leq r_j \leq a_j \text{ for } 1 \leq j \leq k \tag{5.8.15}$$

We shall now show that each r_j must be atleast 1. For this, we must show that each eigenvalue value of A is a root of $m_A(\lambda)$. Consider any eigenvalue value λ_j of A. Then there exists a nonzero vector u such that

$$Au = \lambda_j u$$

From this we can easily conclude that

$$\begin{array}{rcl} A^2 u & = & \lambda_j^2 u \\ A^3 u & = & \lambda_j^3 u \\ \cdots & \cdots & \cdots \\ A^t u & = & \lambda_j^t u \text{ for any nonnegative integer } t, \end{array}$$

and hence, in general, for any polynomial $p(\lambda)$ we have

$$p(A)u \quad = \quad p(\lambda_j)u \tag{5.8.16}$$

Thus in particular, for the polynomial $m_A(\lambda)$ we get

$$m_A(A)u = m_{_A}(\lambda_j)u$$

But $m_A(A) = 0_{n\times n}$ since $m_A(\lambda) \in \hat{\mathcal{A}}$ and hence we get

$$m_{_A}(\lambda_j)u = \theta_n,$$

and this gives us that $m_{_A}(\lambda_j) = \theta_n$ since $u \neq \theta_n$. Thus every characteristic value of A is a root of $m_A(\lambda)$. Hence we can modify (5.8.15) as

$$1 \leq r_j \leq a_j \text{ for } 1 \leq j \leq k \tag{5.8.17}$$

We can summarise all these as follows:
The minimal polynomial $m_A(\lambda)$ has the following properties

1. $m_A(\lambda)$ is a monic polynomial.

2. $m_A(\lambda)$ annihilates A, i.e., $m(A) = 0$.

3. $m_A(\lambda)$ has the lowest degree among all nontrivial polynomials that annihilate A.

4. Every polynomial that annihilates A is a polynomial multiple of $m_A(\lambda)$.

5. Every characteristic value of A is a root of $m_A(\lambda)$.

6. $m_A(\lambda)$ has the form

   $$m_A(\lambda) \quad = \quad (\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k} \tag{5.8.18}$$

   where

   $$1 \leq r_j \leq a_j \text{ for } 1 \leq j \leq k \tag{5.8.19}$$

   It is the lowest degree polynomial of this form that annihilates A.

We shall now look at another criterion for diagonalizability of A in terms of the minimal polynomial. Suppose A is a diagonalizable matrix. Then we have that the algebraic multiplicity of each eigenvalue value is equal to its geometric multiplicity. Hence corresponding to the eigen value λ_j there exist a_j linearly independent characteristic vectors $u_1^{(j)}, u_2^{(j)}, \cdots, u_{a_j}^{(j)}$. By the decompositions that we obtained in Section 5.7, any vector x in $\mathbb{C}^n$ can be expanded as

$$x = x_1 + x_2 + \cdots + x_k$$

where

$$x_j \in W_j = \text{Null Space of } A - \lambda_j I$$

Hence we get that, if $p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_k)$ then

$$\begin{aligned} p(A)x &= p(A)x_1 + p(A)x_2 + \cdots + p(A)x_k \\ &= \theta_n \end{aligned}$$

since for each j, $p(A)$ has a factor $A - \lambda_j I$ and hence $p(A)x_j = \theta_n$, (as $x_j \in$ Null Space of $A - \lambda_j I$). Hence $p(A)$ annihilates A and it is the lowest degree polynomial satisfying (5.8.18) and (5.8.19) that annihilates A. Hence this is the minimal polynomial. Thus if A is diagonalizable then its minimal polynomial is a product of linear factors, one factor from each eigenvalue, i.e.

$$m_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_k)$$

It is interseting to mention here that the converse is also true (we shall not prove this), i.e., if the minimal polynomial of the matrix A is of the above form then it is diagonalizable. We sum up these observations as the following,

Theorem 5.8.2 Let $A \in \mathcal{F}^{n\times n}$ be such that its characteristic polynomial is as in (5.3.4). Then

$$\boldsymbol{A \text{ is diagonalizble} \iff m_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_k)} \tag{5.8.20}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct eigenvalues of A.

5.9 Functions of a Diagonalizable Matrix

We shall now look at the process of defining functions of matrices. Consider a diagonal matrix, $D \in \mathbb{C}^{n\times n}$

$$
\begin{aligned}
D &= \begin{pmatrix} d_1 & & & & \\ & d_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & d_n \end{pmatrix} \\
&\overset{notation}{=} diag.(d_1, d_2, \cdots, d_n)
\end{aligned}
$$

Then we have, for any nonnegative integer r,

$$
D^r = diag.(d_1^r, d_2^r, \cdots, d_n^r) \tag{5.9.1}
$$

Hence for any polynomial

$$
p(\lambda) = \sum_{r=0}^{m} a_r \lambda^r \text{ in } \mathbb{C}[\lambda]
$$

we have

$$
\begin{aligned}
p(D) &= \sum_{r=0}^{m} a_r D^r \\
&= \sum_{r=0}^{m} a_r diag.(d_1^r, d_2^r, \cdots, d_n^r) \text{ by } (5.8.1) \\
&= diag.\left(\sum_{r=0}^{m} a_r d_1^r, \sum_{r=0}^{m} a_r d_2^r, \cdots, \sum_{r=0}^{m} a_r d_n^r\right) \\
&= diag.\left(p(d_1), p(d_2), \cdots, p(d_n)\right)
\end{aligned}
$$

Thus for any diagonal matrix $\boldsymbol{D = diag.(d_1, d_2, \cdots, d_n)} \in \mathbb{C}^{\boldsymbol{n\times n}}$, we can define for any polynomial $\boldsymbol{p(\lambda)} \in \boldsymbol{\mathcal{F}[\lambda]}$, the matrix $\boldsymbol{p(D)}$ as

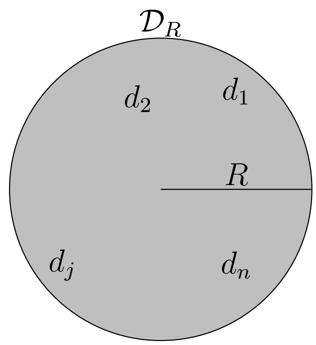
$$
\boldsymbol{p(D) = diag.\left(p(d_1), p(d_2), \cdots, p(d_n)\right)} \tag{5.9.2}
$$

Next, let $\mathcal{D}_R$ be the open disc in the complex plane defined as,

$$
\mathcal{D}_R = \{\lambda \in \mathbb{C} : |\lambda| < R\} \tag{5.9.3}
$$

Let R be large enough such that the disc encloses all the diagonal entries of the diagonal matrix $D = diag.(d_1, d_2, \cdots, d_n)$. This means we have

$$|d_j| < R \text{ for } 1 \leq j \leq n \tag{5.9.4}$$



Any function $f(\lambda)$ which is analytic in $\mathcal{D}_R$ can be expanded in a Taylor series as,

$$f(\lambda) = \sum_{r=0}^{\infty} \frac{f^{(r)}(0)}{r!}\lambda^r \tag{5.9.5}$$

which converges at all points in $\mathcal{D}_R$ and hence, in particular, at the points $d_1, d_2, \cdots, d_n$. Let us define the partial sums $p_N(\lambda)$ as

$$p_N(\lambda) = \sum_{r=0}^{N} \frac{f^{(r)}(0)}{r!}\lambda^r \tag{5.9.6}$$

Then we have

$$\lim_{N\to\infty} p_N(\lambda) = f(\lambda) \text{ for every } \lambda \in \mathcal{D}_R \tag{5.9.7}$$

In particular, we have,

$$\lim_{N\to\infty} p_N(d_j) = f(d_j) \tag{5.9.8}$$

These partial sums $p_N(\lambda)$ are all polynomials in λ, that is $p_N(\lambda) \in \mathbb{C}[\lambda]$. Hence by (5.9.2) we can define $p_N(D)$ as

$$p_N(D) = diag.\left(p_N(d_1), p_N(d_2) \cdots, p_N(d_n)\right) \tag{5.9.9}$$

But by (5.9.8), it follows that, (for every j, $1 \leq j \leq n$), the jth diagonal entry of $p_N(D)$ converges to $f(d_j)$ as $N \to \infty$, and all nondiagonal entries, being zero for all N, converge to zero as $N \to \infty$. Thus we have,

$$\lim_{N\to\infty} diag.\,(p_N(d_1), p_N(d_2)\cdots, p_N(d_n)) \;=\; diag.\,(f(d_1), f(d_2), \cdots, f(d_n))$$

Hence, by (5.8.7), it follows that it is natural to define $f(A)$ as

$$\begin{aligned} f(D) &= \lim_{N\to\infty} p_N(D) \\ &= diag.\,(f(d_1), f(d_2), \cdots, f(d_n)) \end{aligned}$$

Thus we have the following result:
For any diagonal matrix, $\boldsymbol{D = diag.(d_1, d_2, \cdots, d_N)}$ in $\mathbb{C}^{n\times n}$, we can define, $\boldsymbol{f(D)}$ as,

$$\boldsymbol{f(D) \;=\; diag.\,(f(d_1), f(d_2), \cdots, f(d_n))} \tag{5.9.10}$$

for any function $\boldsymbol{f(\lambda)}$ which is analytic in a disc containing all the diagonal entries of $\boldsymbol{D}$.
Now that we have defined the $f(D)$ for a diagonal matrix $D \in \mathbb{C}^{n\times n}$, we shall now define $f(A)$ for a diagonalizable matrix $A \in \mathbb{C}^{n\times n}$. Let the characteristic polynomial be as in (5.3.4). We have,

$$\begin{aligned} A \text{ is diagonalizable} &\Longrightarrow \exists \text{ invertible } P \in \mathbb{C}^{n\times n} \ni P^{-1}AP = D, \text{ a diagonal matrix} \\ &\Longrightarrow A = PDP^{-1} \\ &\Longrightarrow A^r = PD^rP^{-1} \text{ for all nonnegative integers } r \\ &\Longrightarrow p(A) = \sum_{r=0}^{m} a_r P^{-1} D^r P^{-1} \\ &\qquad \text{for any polynomial } p(\lambda) = \sum_{r=0}^{m} a_r \lambda^r \in \mathbb{C}[\lambda] \\ &\Longrightarrow A = \sum_{r=0}^{m} P^{-1}(a_r D^r) P^{-1} \\ &\Longrightarrow A = P\left(\sum_{r=0}^{m} a_r D^r\right) P^{-1} \\ &\Longrightarrow A = P\; p(D)\; P^{-1} \end{aligned}$$

In particular, we have invertible $P \in \mathbb{C}^{n\times n}$ such that,

$$P^{-1}AP \;=\; D = diag(\underbrace{\lambda_1, \lambda_1 \cdots, \lambda_1}_{a_1\ times}, \underbrace{\lambda_2, \lambda_2, \cdots, \lambda_2}_{a_2\ times}, \cdots, \underbrace{\lambda_k, \lambda_k, \cdots, \lambda_k}_{a_k\ times}) \tag{5.9.11}$$

Thus we have,

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

Hence we have,
For any diagonalizable matrix $\boldsymbol{A} \in \mathbb{C}^{\boldsymbol{n\times n}}$ there exists a $\boldsymbol{P} \in \mathbb{C}^{\boldsymbol{n\times n}}$ such that

$$\boldsymbol{P^{-1}AP = (\underbrace{\lambda_1, \cdots, \lambda_1}_{a_1\ times}, \underbrace{\lambda_2, \cdots, \lambda_2}_{a_2\ times}, \cdots, \underbrace{\lambda_k, \cdots, \lambda_k}_{a_k\ times})} \tag{5.9.12}$$

and for any polynomial $\boldsymbol{p(\lambda)} \in \mathbb{C}[\boldsymbol{\lambda}]$ we define $\boldsymbol{p(A)}$ as

$$\boldsymbol{p(A) = P\ diag(\underbrace{p(\lambda_1), \cdots, p(\lambda_1)}_{a_1\ times}, \underbrace{p(\lambda_2), \cdots, p(\lambda_2)}_{a_2\ times}, \cdots, \underbrace{p(\lambda_k), \cdots, p(\lambda_k)}_{a_k\ times})P^{-1}} \tag{5.9.13}$$

Next, as before, we consider a function $f(\lambda)$ which is analytic in a disc $\mathcal{D}_R$ in the complex plane, such that it encloses all the eigenvalues of A. In this disc we have the Taylor series expansion,

$$f(\lambda) \;=\; \sum_{r=0}^{\infty} \frac{f^{(r)}(0)}{r!}\lambda^r \tag{5.9.14}$$

For any positive integer N, we define the partial sums, $p_N(\lambda)$ as

$$p_N(\lambda) \;=\; \sum_{r=0}^{m} \frac{f^{(r)}(0)}{r!}\lambda^r \tag{5.9.15}$$

We have

$$\lim_{N\to\infty} p_N(\lambda) \;=\; f(\lambda) \text{ for all } \in \mathcal{D}_R \tag{5.9.16}$$

In particular we have

$$\lim_{N\to\infty} p_N(\lambda_j) \;=\; f(\lambda_j) \text{ for every eigenvalue } \lambda_k \text{ of } A \qquad (5.9.17)$$

Now we define

$$f(A) \;=\; \lim_{N\to\infty} p_N(A) \qquad (5.9.18)$$

Since we have already defined $P(A)$ for any polynomial $p(\lambda)$, we get

$$\begin{aligned} f(A) &= P \lim_{N\to} p_N(A) P^{-1} \\ &= P\, diag(\underbrace{f(\lambda_1),\cdots,f(\lambda_1)}_{a_1\ times},\underbrace{f(\lambda_2),\cdots,f(\lambda_2)}_{a_2\ times},\cdots,\underbrace{f(\lambda_k),\cdots,f(\lambda_k)}_{a_k\ times})P^{-1} \end{aligned}$$

(Since the disc $\mathcal{D}_R$ has to enclose all the eigenvalues of A we must have $R > Max.\{|\lambda_1|,|\lambda_2|,\cdots,|\lambda_k|\}$). Thus we have,
If $A \in \mathbb{C}^{n\times n}$ is a diagonalizable matrix with its characteristic polynomial as in (5.3.4), then, there exists a $P \in \mathbb{C}^{n\times n}$ such that

$$\boldsymbol{P^{-1}AP = (\underbrace{\lambda_1,\cdots,\lambda_1}_{a_1\ times},\underbrace{\lambda_2,\cdots,\lambda_2}_{a_2\ times},\cdots,\underbrace{\lambda_k,\cdots,\lambda_k}_{a_k\ times})} \qquad \mathbf{(5.9.19)}$$

and for any function $f(\lambda)$ analytic in a disc $\mathcal{D}_R$, (where the radius R of the disc is $> |\lambda_j|$ for all the eigenvalues λ_j of A), we define $f(A)$ as,

$$\boldsymbol{f(A) = P\ diag(\underbrace{f(\lambda_1),\cdots,f(\lambda_1)}_{a_1\ times},\underbrace{f(\lambda_2),\cdots,f(\lambda_2)}_{a_2\ times},\cdots,\underbrace{f(\lambda_k),\cdots,f(\lambda_k)}_{a_k\ times})P^{-1}} \qquad (5.9.20)$$

5.10 Examples

We shall now illustrate the various aspects of diagonalizability described in the previous sections by some examples.

Example 5.10.1 Consider the matrix $A \in \mathbb{C}^{2\times 2}$ defined as,

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \tag{5.10.1}$$

The characteristic polynomial is given by

$$c_{_A}(\lambda) = (\lambda - 1)(\lambda - 2) \tag{5.10.2}$$

Hence the distinct and eigenvalues and their algebraic multiplicities are given by

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

The eigenspaces are given by

$$W_1 = Null\ Space\ of\ A - I \text{ and } W_2 = Null\ Space\ of\ A - 2I$$

It is easy to see that,

$$W_1 = \left\{ x \in \mathbb{C}^2 : x = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \alpha \in \mathbb{C} \right\} \tag{5.10.3}$$

$$W_2 = \left\{ x \in \mathbb{C}^2 : x = \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \alpha \in \mathbb{C} \right\} \tag{5.10.4}$$

Hence we see that

$$dim.W_1 = 1,\ dim.W_2 = 1$$

Thus the geometric multiplicities of both the eigenvalues are 1, that is, $g_1 = 1 = g_2$. Since $a_1 = g_1$ and $a_2 = g_2$ we must have that A is diagonalizable over $\mathbb{R}$. We choose the basis

$$\varphi_1^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{5.10.5}$$

for W_1 and the basis,

$$\varphi_1^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{5.10.6}$$

for W_2. If we now set the matrix P as these two basis vectors as columns we get $P \in \mathbb{R}^{2\times 2}$ and,

$$P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{5.10.7}$$

and

$$P^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \tag{5.10.8}$$

It is easy to verify that,

$$P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \tag{5.10.9}$$

which is a diagonal matrix in $R^{2\times 2}$. Thus A is diagonalizable over $\mathbb{C}$. We shall next look at the Lagrange interpolation polynomials associated with these two eigenvalues. We have

$$\begin{aligned} \ell_1(\lambda) &= \frac{\lambda - 2}{1 - 2} = -(\lambda - 2) \\ \ell_2(\lambda) &= \frac{\lambda - 1}{2 - 1} = \lambda - 1 \end{aligned}$$

Hence we get

$$\begin{aligned} A_1 &= \ell_1(A) = 2I - A \\ &= \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} A_2 &= \ell_2(A) = A - I \\ &= \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

It is easy gerify that the following are true:

$$\begin{aligned} A_1 + A_2 &= I_{2\times 2} \\ 1A_1 + 2A_2 &= A \\ A_1A_2 = A_2A_1 &= 0_{2\times 2} \\ A_1^2 &= A_1 \\ A_2^2 &= A_2 \end{aligned}$$

Further we see that the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is a basis for the column space of A and hence for the Range of A. However, this vector is also a basis for W_1. Hence we get

$$\textit{Range of } A_1 = W_1$$

Similarly the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a basis for the column space of A_2 and hence of Range of A_2, and it is also a basis for W_2. Hence we get,

$$Range\ of\ A_2 = W_2$$

We shall next see the decomposition of a vector $x \in \mathbb{C}^2$ as the sum of a vector in W_1 and a vector in W_2.
Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. We want to write x as

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = X_1 + X_2 \text{ where } X_1 \in W_1, X_2 \in W_2$$

Since $X_1 \in W_1$ and $X_2 \in W_2$ they must be of the form

$$\begin{aligned} X_1 &= \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \\ X_2 &= \begin{pmatrix} \beta \\ \beta \end{pmatrix} \end{aligned}$$

Thus, given $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, we want to find $\alpha, \beta \in \mathbb{C}$ such that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} \beta \\ \beta \end{pmatrix}$$

Hence we get

$$\begin{aligned} \alpha &= x_1 - x_2 \\ \beta &= x_2 \end{aligned}$$

Hence,

$$\begin{aligned} X_1 &= \begin{pmatrix} x_1 - x_2 \\ 0 \end{pmatrix} \in W_1 \\ X_2 &= \begin{pmatrix} x_2 \\ x_2 \end{pmatrix} \in W_2 \\ x &= X_1 + X_2 \end{aligned}$$

We shall now look at the minimal polynomial of A. Since the matrix is diagonalizable we must have the minimal polynomial as

$$m_{_A}(\lambda) \;=\; (\lambda-1)(\lambda-2)$$

Let us verify that this polynomial annihilates A. We have

$$\begin{aligned} m_{_A}(A) &= (A-I)(A-2I) \\ &= \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \\ &= 0_{2\times 2} \end{aligned}$$

We shall now look at some functions of A. We have, as obtained in (5.10.9),

$$A \;=\; P\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}P^{-1} \tag{5.10.10}$$

Hence for any function $f(\lambda)$ analytic in a disc $\mathcal{D}_R$ where $R > Max.\{1,2\}$, that is $R > 2$, we define

$$f(A) \;=\; P\begin{pmatrix} f(1) & 0 \\ 0 & f(2) \end{pmatrix}P^{-1}$$

For example, if $f(\lambda) = sin\left(\frac{\pi}{2}\lambda\right)$ then we have

$$\begin{aligned} sin\left(\frac{\pi}{2}A\right) &= P\begin{pmatrix} sin\left(\frac{\pi}{2}\right) & 0 \\ 0 & sin(\pi) \end{pmatrix}P^{-1} \\ &= P\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}P^{-1} \end{aligned}$$

Similarly, if $f(\lambda) = cos(\pi\lambda)$ then we get

$$\begin{aligned} cos(\pi A) &= P\begin{pmatrix} cos(\pi) & 0 \\ 0 & cos(2\pi) \end{pmatrix}P^{-1} \\ &= P\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}P^{-1} \end{aligned}$$

Example 5.10.2 Consider the matrix $A \in \mathbb{C}^{2\times 2}$ defined as,

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{5.10.11}$$

The characteristic polynomial is given by,

$$c_{_A}(\lambda) = \lambda^2 + 1 \tag{5.10.12}$$

The eigenvalues and their algebraic multiplicities are given by,

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

For the eigenspaces we have,

$$\begin{aligned}
W_1 &= Null\ Space\ of\ A_i I \\
&= Null\ Space\ of \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \\
&= \left\{ x \in \mathbb{C}^2 : x = \alpha \begin{pmatrix} 1 \\ -i \end{pmatrix}; \alpha \in \mathbb{C} \right\}, \text{and} \\
W_2 &= Null\ Space\ of\ A + iI \\
&= Null\ Space\ of \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \\
&= \left\{ x \in \mathbb{C}^2 : x = \beta \begin{pmatrix} 1 \\ i \end{pmatrix}; \in \mathbb{C} \right\}, \text{and}
\end{aligned}$$

We have the geometric multiplicities of the eigenvalues as

$$\begin{aligned}
g_1 &= dim.W_1 = 1 \\
g_2 &= dim.W_2 = 1
\end{aligned}$$

Since, for each eigenvalue, the algebraic and geometric multiplicities are equal, the matrix is diagonalizable over $\mathbb{C}$. We choose the basis for W_1 and W_2 as,

$$\begin{aligned}
\varphi_1^{(1)} &= \begin{pmatrix} 1 \\ -i \end{pmatrix} \text{ and} \\
\varphi_1^{(2)} &= \begin{pmatrix} 1 \\ i \end{pmatrix}
\end{aligned}$$

If we now set,

$$P = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \in \mathbb{C}^{2\times 2} \tag{5.10.13}$$

then,

$$P^{-1} = \frac{1}{2i}\begin{pmatrix} i & -1 \\ i & 1 \end{pmatrix} \in \mathbb{C}^{2\times 2} \tag{5.10.14}$$

It is easy to verify that

$$P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \text{ a diagonal matrix in } C^{2\times 2} \tag{5.10.15}$$

The Lagrange interpolation polynomials corresponding to these eigenvalues are given by,

$$\ell_1(\lambda) = \frac{\lambda+i}{i+i} = \frac{\lambda+i}{2i} \tag{5.10.16}$$

$$\ell_2(\lambda) = \frac{\lambda-i}{-i-i} = -\frac{\lambda-i}{2i} \tag{5.10.17}$$

We now define

$$\begin{aligned} A_1 &= \ell_1(A) \\ &= \frac{1}{2i}(A+iI) \\ &= \frac{1}{2i}\begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \\ A_2 &= \ell_2(A) \\ &= -\frac{1}{2i}(A-iI) \\ &= -\frac{1}{2i}\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \end{aligned}$$

It is now easy to verify that A_1 and A_2 satisfy the following:

$$\begin{aligned} A_1 + A_2 &= I_{2\times 2} \\ iA_1 + (-i)A_2 &= A \end{aligned}$$

$$\begin{aligned} A_1A_2 = A_2A_1 &= 0_{2\times 2} \\ A_1^2 &= A_1 \\ A_2^2 &= A_2 \\ W_1 &= Range\ of\ A_1 \\ W_2 &= Range\ of\ A_2 \end{aligned}$$

The decomposition of any vector $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{C}^2$ as the sum of a vector in W_1 and a vector in W_2 is given by,

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{x_1+ix_2}{2} \\ -i\frac{x_1+ix_2}{2} \end{pmatrix} + \begin{pmatrix} \frac{x_1-ix_2}{2} \\ i\frac{x_1-ix_2}{2} \end{pmatrix} \tag{5.10.18}$$

Since the matrix is diagonalizable the minimal polynomial is given by

$$m_{_A}(\lambda) = (\lambda - i)(\lambda + i) = \lambda^2 + 1 \tag{5.10.19}$$

It is easy to verify that

$$m_{_A}(A) = A^2 + I_{2\times 2} = 0_{2\times 2} \tag{5.10.20}$$

We shall now look at some functions of this matrix. Since $\lambda_1 = i$ and $\lambda_2 = -i$ are the eigenvalues of A, for any function $f(\lambda)$ which is analytic in a disc $\mathcal{D}_R$ of radius $R > Max.\{|\lambda_1|, |\lambda_2|\}$, (that is $R > 1$), $f(A)$ can be defined as

$$f(A) = P\begin{pmatrix} f(i) & 0 \\ 0 & f(-i) \end{pmatrix} P^{-1} \tag{5.10.21}$$

where P is as in (5.10.3). For example if $f(\lambda) = exp.(\pi\lambda)$ we have,

$$\begin{aligned} exp.(\pi A) &= P\begin{pmatrix} exp.(i\pi) & 0 \\ 0 & exp.(-i\pi) \end{pmatrix} P^{-1} \\ &= P\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} P^{-1} \end{aligned}$$

Similarly, if $f(\lambda) = exp.\left(\frac{\pi}{2}\lambda\right)$ then we have,

$$\begin{aligned} f(A) &= P\begin{pmatrix} exp.\left(\frac{\pi}{2}i\right) & 0 \\ 0 & exp.\left(-i\frac{\pi}{2}\right) \end{pmatrix} \\ &= P\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} P^{-1} \end{aligned}$$

Example 5.10.3 Consider the matrix $A \in \mathbb{C}^{2\times 2}$ defined as given below:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{5.10.22}$$

The characteristic polynomial is given by

$$c_{_A}(\lambda) = \lambda^2 \tag{5.10.23}$$

The only eigenvalue is $\lambda_1 = 0$ and its algebraic multiplicity is 2. For the corresponding eigenspace, we have,

$$\begin{aligned} W_1 &= Null\ Space\ of\ A - 0I \\ &= Null\ Space\ of\ A \\ &= \left\{ x \in \mathbb{C}^2 : x = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \end{aligned}$$

Hence we have,

$$g_1 = dim.W_1 = 1$$

Thus we have $g_1 < a_1$ and hence this matrix is not diagonalizable. For this matrix, since the minimal polynomial must divide the characteristic polynomial, the minimal polynomial has to be λ or λ^2. But since λ does not annihilate A we get that the minimal polynomial is given by

$$m_{_A}(\lambda) = \lambda^2$$

Example 5.10.4 Consider the matrix $A \in \mathbb{C}^{2\times 2}$ defined below:

$$A = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \tag{5.10.24}$$

Its characteristic polynomial is given by

$$c_{_A}(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 3) \tag{5.10.25}$$

The eigenvalues and their algebraic multiplicities are given by,

$$\lambda_1 = 1,\ a_1 = 1,\ \lambda_2 = 2,\ a_2 = 1,\ \lambda_3 = 3,\ a_3 = 1$$

The eigenspaces are given by,

$$
\begin{aligned}
W_1 &= Null\ Space\ of\ A - I \\
&= \left\{ x \in \mathbb{C}^2 : x = \alpha \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} ; \alpha \in \mathbb{C} \right\} \\
W_2 &= Null\ Space\ of\ A - 2I \\
&= \left\{ x \in \mathbb{C}^2 : x = \beta \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} ; \beta \in \mathbb{C} \right\} \\
W_3 &= Null\ Space\ of\ A - 3I \\
&= \left\{ x \in \mathbb{C}^2 : x = \gamma \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} ; \gamma \in \mathbb{C} \right\}
\end{aligned}
$$

Hence we get

$$g_1 = dim.W_1 = 1 \tag{5.10.26}$$

$$g_2 = dim.W_2 = 1 \tag{5.10.27}$$

$$g_3 = dim.W_3 = 1 \tag{5.10.28}$$

Since, for every eigenvalue, the algebraic multiplity and geometric multiplicity are equal, the matrix is diagonalizable. We chose the basis for the eigenspaces as,

$$\varphi_1^{(1)} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \text{ for } W_1, \tag{5.10.29}$$

$$\varphi_1^{(2)} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \text{ for } W_2 \text{ and} \tag{5.10.30}$$

$$\varphi_1^{(3)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \text{ for } W_3 \tag{5.10.31}$$

If we now define,

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \tag{5.10.32}$$

then P is invertible and

$$P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \text{ a diagonal matrix in } \mathbb{C}^{2\times 2} \qquad (5.10.33)$$

The Lagrange interpolation polynomials corresponding to these eigenvalues are given by,

$$\begin{aligned}
\ell_1(\lambda) &= \frac{(\lambda-2)(\lambda-3)}{(1-2)(1-3)} \\
&= \frac{1}{2}(\lambda-2)(\lambda-3) \\
\ell_2(\lambda) &= \frac{(\lambda-1)(\lambda-3)}{(2-1)(2-3)} \\
&= -(\lambda-1)(\lambda-3) \\
\ell_3(\lambda) &= \frac{(\lambda-1)(\lambda-2)}{(3-1)(3-2)} \\
&= \frac{1}{2}(\lambda-1)(\lambda-2)
\end{aligned}$$

We then define,

$$\begin{aligned}
A_1 &= \ell_1(A) \\
&= \frac{1}{2}(A-2I)(A-3I) \\
&= \frac{1}{2}\begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ -1 & 0 & 1 \end{pmatrix} \\
A_2 &= \ell_1(A) \\
&= -(A-I)(A-3I) \\
&= \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix} \\
A_3 &= \ell_3(A) \\
&= \frac{1}{2}(A-I)(A-2I)
\end{aligned}$$

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

It is now easy to check that these matrices satisfy the following properties:

$$\begin{aligned}
A_1 + A_2 + A_3 &= I_{3\times 3} \\
1A_1 + 2A_2 + 3A_3 &= A \\
A_iA_j &= 0_{3\times 3} \text{ for } i \neq j,\ 1 \leq i, j \leq 3 \\
A_1^2 &= A_1 \\
A_2^2 &= A_2 \\
A_3^2 &= A_3 \\
W_1 &= Range\ A_1 \\
W_2 &= Range\ A_2 \\
W_3 &= Range\ A_3
\end{aligned}$$

The decomposition of any vector $x \in \mathbb{R}^3$ is obtained as follows:

$$\begin{aligned}
x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} \alpha \\ 2\alpha \\ -\alpha \end{pmatrix} + \begin{pmatrix} \beta \\ -\beta \\ \beta \end{pmatrix} + \begin{pmatrix} \gamma \\ 0 \\ \gamma \end{pmatrix} \\
&\Longrightarrow \\
\alpha + \beta + \gamma &= x_1 \\
2\alpha - \beta &= x_2 \\
-\alpha + \beta + \gamma &= x_3 \\
&\Longrightarrow \\
\alpha &= \frac{x_1 - x_3}{2} \\
\beta &= x_1 - x_2 - x_3 \\
\gamma &= -\frac{x_1}{2} + x_2 + 3\frac{x_3}{2}
\end{aligned}$$

Thus we have,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{x_1}{2} - \frac{x_3}{2} \\ x_1 - x_3 \\ -\frac{x_1}{2} + \frac{x_3}{2} \end{pmatrix}}_{\in W_1} + \underbrace{\begin{pmatrix} x_1 - x_2 - x_3 \\ -x_1 + x_2 + x_3 \\ x_1 - x_2 - x_3 \end{pmatrix}}_{\in W_2} + \underbrace{\begin{pmatrix} -\frac{x_1}{2} + x_2 + 3\frac{x_3}{2} \\ 0 \\ -\frac{x_1}{2} + x_2 + 3\frac{x_3}{2} \end{pmatrix}}_{\in W_3}$$

Since the matrix is diagonalizable its minimal polynomial is given by

$$m_{_A}(\lambda) \quad = \quad (\lambda-1)(\lambda-2)(\lambda-3) \tag{5.10.34}$$

It is easy to verify that

$$m_{_A}(A) = (A - I_{2\times 2})(A - 2I_{2\times 2})(A - 3I_{2\times 2}) = 0_{3\times 3}$$

For any function $f(\lambda)$ analytic in a disc $\mathcal{D}_R$, where $R > 3$ we can define $f(A)$ as

$$f(A) \quad = \quad P \begin{pmatrix} f(1) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{pmatrix} P^{-1} \tag{5.10.35}$$

where P is as in (5.10.32). For example we have

$$\begin{aligned}
sin(\pi A) &= P \begin{pmatrix} sin(\pi) & 0 & 0 \\ 0 & sin(2\pi) & 0 \\ 0 & 0 & sin(3\pi) \end{pmatrix} P^{-1} \\
&= P \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} P^{-1} \\
&= 0_{3\times 3} \\
sin\left(\frac{\pi}{2}A\right) &= P \begin{pmatrix} sin\left(\frac{\pi}{2}\right) & 0 & 0 \\ 0 & sin(\pi) & 0 \\ 0 & 0 & sin\left(3\frac{\pi}{2}\right) \end{pmatrix} P^{-1} \\
&= P \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} P^{-1} \\
exp.(At) &= P \begin{pmatrix} exp.(t) & 0 & 0 \\ 0 & exp.(2t) & 0 \\ 0 & 0 & exp.(3t) \end{pmatrix} P^{-1} \text{ for any } t \in \mathbb{C}
\end{aligned}$$

Example 5.10.5 Consider the matrix $A \in \mathbb{C}^{2\times 2}$ as defined below:

$$A \quad = \quad \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix} \tag{5.10.36}$$

The characteristic polynomial is given by

$$c_{_A}(\lambda) \;=\; (\lambda-1)(\lambda+3)^2 \tag{5.10.37}$$

The eigenvalues and their algebraic multplicities are given by

$$\lambda_1 = 1,\; a_1 = 1,\; \lambda_2 = -3,\; a_2 = 2$$

We have

$$\begin{aligned} A - I_{3\times 3} &= \begin{pmatrix} 0 & -4 & -4 \\ 8 & -12 & -8 \\ -8 & 8 & 4 \end{pmatrix} \\ A + 3I_{3\times 3} &= \begin{pmatrix} 4 & -4 & -4 \\ 8 & -8 & -8 \\ -8 & 8 & 8 \end{pmatrix} \end{aligned}$$

For the eigenspaces we have

$$\begin{aligned} W_1 &= \mathit{Null\ Space\ of}\ A - I_{3\times 3} \\ &= \left\{ x \in \mathbb{C}^3 : x = \alpha \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix},\ \alpha \in \mathbb{C} \right\} \\ W_2 &= \mathit{Null\ Space\ of}\ A + 3I_{3\times 3} \\ &= \left\{ x \in \mathbb{C}^3 : x = \beta \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix},\ \beta, \gamma \in \mathbb{C} \right\} \end{aligned}$$

We choose bases for W_1 and W_2 as follows:

$$\begin{aligned} \varphi_1^{(1)} &= \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix} \text{ for } W_1;\ \text{and} \\ \varphi_1^{(2)} &= \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{ and} \\ \varphi_2^{(2)} &= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \text{ for } W_2 \end{aligned}$$

We have for the geometric multiplicities,

$$\begin{aligned} g_1 = dim.W_1 &= 1 \\ g_2 = dim.W_2 &= 2 \end{aligned}$$

Since the algebraic multiplicity is equal to the geometric multiplicity for each eigenvalue, the matrix is diagonalizable. We now define,

$$P = \begin{pmatrix} -1 & 1 & 0 \\ -2 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix} \tag{5.10.38}$$

We can easily verify that

$$AP = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

and hence

$$P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \text{ a diagonal matrix in } \mathbb{C}^{3\times 3} \tag{5.10.39}$$

$$A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} P^{-1} \tag{5.10.40}$$

For the Lagrange interpolation polyomials corresponding to these eigenvalues we have

$$\begin{aligned} \ell_1(\lambda) &= \frac{(\lambda+3)}{(1+3)} \\ &= \frac{1}{4}(\lambda+3) \\ \ell_2(\lambda) &= \frac{(\lambda-1)}{(-3-1)} \\ &= -\frac{1}{4}(\lambda-1) \end{aligned}$$

We get the decomposition of the matrix as follows: We define,

$$A_1 = \ell_1(A)$$

$$
\begin{aligned}
&= \frac{1}{4}(A + 3I_{3\times 3}) \\
&= \begin{pmatrix} 1 & -1 & -1 \\ 2 & -2 & -2 \\ -2 & 2 & 2 \end{pmatrix} A_2 \qquad =
\end{aligned}
$$

$$
\begin{aligned}
\ell_2(A) & \\
&= -\frac{1}{4}(A - I_{3\times 3}) \\
&= \begin{pmatrix} 0 & 1 & 1 \\ -2 & 3 & 2 \\ 2 & -2 & -1 \end{pmatrix}
\end{aligned}
$$

It is now easy to verify the following:

$$
\begin{aligned}
A_1 + A_2 &= I_{3\times 3} \\
1A_1 + (-3)A_2 &= A \\
A_1A_2 = A_2A_1 &= 0_{3\times 3} \\
A_1^2 &= A_1 \\
A_2^2 &= A_2 \\
W_1 &= Range\ of\ A_1 \\
W_2 &= Range\ of\ A_2
\end{aligned}
$$

Since the matrix is diagonalizable the minimal polynomial must be

$$
m_{_A}(\lambda) = (\lambda - 1)(\lambda + 3) \tag{5.10.41}
$$

It is easy to verify that

$$
\begin{aligned}
m_{_A}(A) &= (A - I_{3\times 3})(A + 3I_{3\times 3}) \\
&= 0_{3\times 3}
\end{aligned}
$$

For any function $f(\lambda)$ analytic in a disc $\mathcal{D}_R$, where $R > Max.\{|1|, |-3|\}$, (that is $R > 3$), we can define $f(A)$ as,

$$
f(A) = P \begin{pmatrix} f(1) & 0 & 0 \\ 0 & f(3) & 0 \\ 0 & 0 & f(3) \end{pmatrix} P^{-1} \tag{5.10.42}
$$

where P is as defined in (5.10.38). For example we have

$$\begin{aligned}
cos(\pi A) &= P\begin{pmatrix} cos(\pi) & 0 & 0 \\ 0 & cos(3\pi) & 0 \\ 0 & 0 & cos(3\pi) \end{pmatrix} P^{-1} \\
&= P\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} P^{-1} \\
&= P(-I_{3\times 3})P^{-1} \\
&= -I_{3\times 3}
\end{aligned}$$

The decomposition of a vector $x \in \mathbb{C}^3$ as the sum of a vector in W_1 and a vector in W_2 is obtained as follows:
We have to find α, β and $\gamma \in \mathbb{C}$ such that

$$\begin{aligned}
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} -\alpha \\ -2\alpha \\ 2\alpha \end{pmatrix} + \begin{pmatrix} \beta \\ \beta+\gamma \\ \gamma \end{pmatrix} \\
&\Longrightarrow \\
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} -\alpha+\beta \\ -2\alpha+\beta+\gamma \\ 2\alpha+\gamma \end{pmatrix} \\
&\Longrightarrow \\
-\alpha+\beta &= x_1 \\
-2\alpha+\beta+\gamma &= x_2 \\
2\alpha+\gamma &+ x_3
\end{aligned}$$

Solving fr α, β, γ we get

$$\begin{aligned}
\alpha &= \frac{x_1}{3} - \frac{x_2}{3} + \frac{x_3}{3} \\
\beta &= \frac{4x_1}{3} - \frac{x_2}{3} + \frac{x_3}{3} \\
\gamma &= -\frac{2x_1}{3} + \frac{2x_2}{3} + \frac{x_3}{3}
\end{aligned}$$

Hence we get

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad = \quad \underbrace{\begin{pmatrix} -\frac{x_1}{3} + \frac{x_2}{3} - \frac{x_3}{3} \\ -\frac{2x_1}{3} + \frac{2x_2}{3} - \frac{2x_3}{3} \\ \frac{2x_1}{3} - \frac{2x_2}{3} + \frac{2x_3}{3} \end{pmatrix}}_{\in W_1} + \underbrace{\begin{pmatrix} \frac{4x_1}{3} - \frac{x_2}{3} + \frac{x_3}{3} \\ \frac{2x_1}{3} + \frac{x_2}{3} + \frac{2x_3}{3} \\ -\frac{2x_1}{3} + \frac{2x_2}{3} + \frac{x_3}{3} \end{pmatrix}}_{\in W_2} \tag{5.10.43}$$