# Chapter 6

# General Matrices and Jordan Canonical Form

## 6.1 Primary Decomposition Theorem

We have seen in Chapter 5 the various decompositions that we get in the case of a diagonalizable matrix. The analogous decomposition for a general matrix is given by the Primary Decomposition Theorem. Let $A \in \mathcal{F}^{n\times n}$ be such that its characteristic polynomial and minimal polynomial are given with our usual notation as

$$c_{_A}(\lambda) = (\lambda-\lambda_1)^{a_1}(\lambda-\lambda_2)^{a_2}\cdots(\lambda-\lambda_k)^{a_k} \tag{6.1.1}$$
$$m_{_A}(\lambda) = (\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k} \tag{6.1.2}$$

In the diagonalizable case all the $r_j$ are 1. However, in the general case, one or more of the $r_j$ may be greater than 1. Analogous to the eigenspace corresponding to an eigenvalue $\lambda_j$ we define the following sequence of subspaces:

$$\begin{aligned}
V_1^{(j)} &= \mathit{Null\ Space}\ (A-\lambda_j I_{n\times n}) && (6.1.3)\\
&\quad \text{(which is the eigenspace corresponding to } \lambda_j \\
V_2^{(j)} &= \mathit{Null\ Space}\ (A-\lambda_j I_{n\times n})^2 && (6.1.4)\\
\cdots\ &\cdots\ \cdots\cdots \\
V_r^{(j)} &= \mathit{Null\ Space}\ (A-\lambda_j I_{n\times n})^r && (6.1.5)\\
\cdots\ &\cdots\ \cdots\cdots \\
V_{r_j}^{(j)} &= \mathit{Null\ Space}\ (A-\lambda_j I_{n\times n})^{r_j} && (6.1.6)
\end{aligned}$$

It can be shown that for each $j = 1, 2, \cdots, k$, we have $dim.(V_{r_j}^{(j)}) = a_j$, and:

$$\begin{array}{c} V_1^{(j)} \underset{\neq}{\subset} V_2^{(j)} \underset{\neq}{\subset} \cdots \underset{\neq}{\subset} V_{(r-1)}^{(j)} \underset{\neq}{\subset} V_r^{(j)} \underset{\neq}{\subset} V_{(r+1)}^{(j)} \underset{\neq}{\subset} \cdots \\ \underset{\neq}{\subset} V_{(r_j-1)}^{(j)} \underset{\neq}{\subset} V_{r_j}^{(j)} = V_{(r_j+1)}^{(j)} = V_{(r_j+2)}^{(j)} = \cdots \end{array} \tag{6.1.7}$$

Hence we have,

$$\begin{aligned} g_j &= dim.V_1^{(j)} < dim.V_2^{(j)} < \cdots < dim.V_r^{(j)} < dim.V_{r+1}^{(j)} < \cdots \\ &\quad < dim.V_{(r_j-1)}^{(j)} < dim.V_{r_j}^{(j)} = a_j = dim.V_{(r_j+1)}^{(j)} = dim.V_{(r_j+2)}^{(j)} = \cdots \end{aligned} \tag{6.1.8}$$

We shall denote by $V_j$ the subspace $V_{r_j}^{(j)}$, that is,

$$V_j \overset{def}{=} Null\ Space\ (A - \lambda_j I_{n\times n})^{r_j} \tag{6.1.9}$$

We observe that, in the case of a diagonalizable matrix $A$, for each $j$ we have $V_j = W_j$, the eigenspace corresponding to the eigenvalue $\lambda_j$. Just as the eigenspaces $W_j$ give a decomposition of $\mathcal{F}^n$ in the case of diagonalizable matrices, the subspaces $V_j$ give rise to a decomposition of $\mathcal{F}^n$ in the case of a general matrix. This is the Primary Decomposition Theorem. We have

**Theorem 6.1.1** **Primary Decomposition Theorem**
**Let $\boldsymbol{A} \in \boldsymbol{\mathcal{F}^{n\times n}}$ have characteristic polynomial and minimal polynomial as in (6.1.1) and (6.1.2). Let $\boldsymbol{V_j}$ be as in (6.1.9). Then every $\boldsymbol{x} \in \boldsymbol{\mathcal{F}^n}$ can be written uniquely as a sum**

$$\boldsymbol{x} = \boldsymbol{X_1} + \boldsymbol{X_2} + \cdots + \boldsymbol{X_j} + \cdots + \boldsymbol{X_k} \textbf{ where } \boldsymbol{X_j} \in \boldsymbol{V_j} \tag{6.1.10}$$

We write this symbolically as

$$\mathcal{F}^n = V_1 \oplus V_2 \oplus \cdots \oplus V_j \oplus \cdots \oplus V_k \tag{6.1.11}$$

(a "Direct Sum Decomposition" of $\mathcal{F}^n$).
Analogous to the Lagrange polynomials we constructed in the case of diagonalizable matrices, we shall now look at some polynomials in the case of

a general matrix. If $A \in \mathcal{F}^{n\times n}$ has characteristic polynomial and minimal polynomial as in (6.1.1) and (6.1.2), we define,

$$f_j(\lambda) \;=\; \prod_{\substack{i=1\\ i\neq j}}^{k}(\lambda-\lambda_i)^{r_i} \tag{6.1.12}$$

The polynomials $f_1(\lambda), f_2(\lambda), \cdots, f_k(\lambda)$ have gcd 1 and hence by (5.8.8) we have polynomials $g_1(\lambda), g_2(\lambda), \cdots, g_k(\lambda)$ such that

$$f_1(\lambda)g_1(\lambda)+f_2(\lambda)g_2(\lambda)+\cdots+f_k(\lambda)g_k(\lambda) \;=\; 1 \tag{6.1.13}$$

Let

$$L_j(\lambda) \;=\; f_j(\lambda)g_j(\lambda) \tag{6.1.14}$$

We then have

$$L_1(\lambda)+L_2(\lambda)+\cdots+L_k(\lambda) \;=\; 1 \tag{6.1.15}$$

We further observe that if $i \neq j$ the $L_i(\lambda)L_j(\lambda)$ has $m_{_A}(\lambda)$ as a factor, that is,

$$L_i(\lambda)L_j(\lambda) \;=\; q_{ij}(\lambda)m_{_A}(\lambda) \text{ if } i\neq j \text{ ( where } q_{ij}(\lambda)\in\mathcal{F}[\lambda]) \tag{6.1.16}$$

We now define the matrices $A_1, A_2, \cdots, A_j, \cdots, A_k$ as follows:

$$A_j \;=\; L_j(A) \tag{6.1.17}$$

From (6.1.15) we get,

$$A_1+A_2+\cdots+A_k \;=\; I_{n\times n} \tag{6.1.18}$$

Further from (6.1.16) we get

$$A_iA_j \;=\; q_{ij}(A)m_{_A}(A) \text{ if } i\neq j$$

and hence

$$A_iA_j \;=\; 0_{n\times n} \text{ if } i\neq j \tag{6.1.19}$$

From the above we see that

$$\begin{aligned} A_j^2 &= A_j \text{ for all } j && (6.1.20) \\ (A - \lambda_j I_{n\times n})^{r_j} A_j &= \left\{ \begin{array}{l} q_j(\lambda) m_{_A}(\lambda) \text{ where } q_j(\lambda) \in \mathcal{F}[\lambda] = \\ 0_{n\times n} \end{array} \right\} && (6.1.21) \end{aligned}$$

In the case of a diagonalizable matrix, the polynomials $L_j(\lambda)$ are the Lagrange interpolation polynomials $\ell_j(\lambda)$ and the matrices $A_j$ are then precisely the decomposition matrices we got for diagonalizable matrices. In that case we also get the decomposition

$$\lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k = A$$

However, in the general case, we may not get $\lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k$ to be equal to $A$. So we let,

$$D = \lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k \qquad (6.1.22)$$

Then we see that $D \in \mathcal{F}^{n\times n}$ is such that there exist $k$ distinct matrices $A_1, A_2, \cdots, A_k$ and $k$ distinct scalars $, \lambda_2, \cdots, \lambda_k$ such that

$$\begin{aligned} A_1 + A_2 + \cdots + A_k &= I_{n\times n}, \\ A_i A_j &= 0_{n\times n} \text{ if } i \neq j \text{ and} \\ \lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k &= D \end{aligned}$$

Hence by Theorem 5.7.1, we get that $D$ is a diagonalizable matrix. As observed above, in case $A$ is a diagonalizable matrix, $D$ will precisely be $A$. In the case of a general matrix, $A$ will not be equal to $D$. We shall now analyse by how much $A$ will differ from the diagonalizable matrix $D$. Let

$$N = A - D \qquad (6.1.23)$$

We have, from (6.1.18),

$$\begin{aligned} A &= A I_{n\times n} \\ &= AA_1 + AA_2 + \cdots + AA_k \end{aligned}$$

Combining this with (6.1.22) we get

$$\begin{aligned} A - D &= (AA_1 + AA_2 + \cdots + AA_k) - (\lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k) \\ &= (A - \lambda_1 I_{n\times n}) A_1 + (A - \lambda_2 I_{n\times n}) A_2 + \cdots + (A - \lambda_k I_{n\times n}) A_k \end{aligned}$$

Using (6.1.19), and (6.1.20), we get

$$(A - D)^r \;=\; (A - \lambda_1 I_{n\times n})^r A_1 + (A - \lambda_2 I_{n\times n})^r A_2 + \cdots + (A - \lambda_k I_{n\times n})^r A_k \tag{6.1.24}$$

If $r \geq\ Max.\{r_1, r_2, \cdots, r_j\}$ then $(A - \lambda_j I_{n\times n})^r A_j$ has a factor $m_{_A}(A)$, (by (6.1.21)), and hence

$$(A - \lambda_j I_{n\times n})^r A_j \;=\; 0_{n\times n} \text{ for all } j \text{ if } r \geq Max.\{r_1, r_2, \cdots, r_j\} \tag{6.1.25}$$

Hence we get from (6.1.23) and (6.1.24) ,

$$N^r = (A - D)^r \;=\; 0_{n\times n} \text{ for } r \geq Max.\{r_1, r_2, \cdots, r_j\} \tag{6.1.26}$$

Thus the difference between $A$ and the diagonalizable matrix $D$ is such that a power of it vanishes. This leads us to the notion of a nilpotent matrix. We have

**Definition 6.1.1 A matrix $N \in \mathbb{C}^{n\times n}$ is said to be a "NILPOTENT" matrix if there exists a positive integer $r$ such that $N^r = 0_{n\times n}$. For a nilpotent matrix the smallest power $r$ for which $N^r = 0_{n\times n}$ is called the "order of nilpotency" of $N$**

Thus we have that the matrix $N$ defined in (6.1.23) is a nilpotent matrix the order of nil potency being equal to $Max.\{r_1, r_2, \cdots, r_k\}$. We can summarise our analysis above as follows:

**Theorem 6.1.2 Every matrix $A \in \mathbb{C}^{n\times n}$ can be expressed as the sum,**

$$\boldsymbol{A} \;=\; \boldsymbol{D} + \boldsymbol{N}, \tag{6.1.27}$$

**of a diagonalizable matrix $D$ and a nilpotent matrix $N$**

We have already analysed diagonalizable matrices in the last chapter. We shall now analyse a nilpotent matrix.

## 6.2 Nilpotent Matrices

Let $N \in \mathcal{F}^{n\times n}$ be a nilpotent matrix with order of nilpotency $\gamma_{_N}$. This means that

$$\begin{aligned} N^{\gamma_N} &= 0_{n\times n} \text{ and} && (6.2.1)\\ N^r &\neq 0_{n\times n} \text{ for any } r < \gamma_{_N} && (6.2.2) \end{aligned}$$

From (6.2.1) we also get,

$$N^r = 0_{n\times n} \text{ for all } r \geq \gamma_{_N} \qquad (6.2.3)$$

From this it follows that the polynomial $p(\lambda) = \lambda^{\gamma_N}$ is an annihilating polynomial for $N$. Hence the minimal polynomial must divide $\lambda^{\gamma_N}$. Thus the minimal polynomial must be of the form $m_{_N}(\lambda) = \lambda^r$ where $r \leq \gamma_{_N}$. However, from (6.2.2) we have that $N^r \neq 0_{nn}$ for any $r < \gamma_{_N}$. Hence we get that the minimal polynomial of $N$ is given by

$$m_{_N}(\lambda) = \lambda^{\gamma_N}$$

Thus we see that **$\boldsymbol{\lambda_1 = 0}$ is the onoy eigenvalue of any nilpotent matrix $\boldsymbol{N}$**. Since the characteristic polynomial of $N$ is a monic polynomial of degree $n$ and its roots are the eigenvalues we get that the characteristic polynomial of $N$ is given by

$$c_{_N}(\lambda) = \lambda^n$$

Hence we have,

**Theorem 6.2.1 If $\boldsymbol{N \in \mathcal{F}^{n\times n}}$ is a nilpotent matrix with order of nilpotency as $\boldsymbol{\gamma_{_N}}$ then its characteristic and minimal polynomials are given by**

$$\begin{aligned} \boldsymbol{c_{_N}(\lambda)} &= \boldsymbol{\lambda^n}, \textbf{ and} && (6.2.4)\\ \boldsymbol{m_{_N}(\lambda)} &= \boldsymbol{\lambda^{\gamma_N}} && (6.2.5) \end{aligned}$$

From the above simple properties we now obtain a useful result. Suppose $A \in \mathcal{F}^{n\times n}$ is both nilpotent and diagonalizable. Since it is nilpotent matrix its minimal polynomial is of tha form $m_{_A} = \lambda^{\gamma_A}$, where $\gamma_{_A}$ is its order of nilpotency. Combining this with the fact that $A$ is diagonalizable we get that its minimal polynomial must be $m_{_A}(\lambda) = \lambda$. Hence $m_{_A}(A) = 0_{n\times n}$ gives us that $A = 0_{n\times n}$. Thus it follows that

**Theorem 6.2.2 The only matrix in $\mathcal{F}^{n\times n}$ which is both diagonalizable and nilpotent is the zero matrix**

We shall now see an example to show that the sum and product of two nilpotent matrices need not be nilpotent.

**Example 6.2.1** Consider the matrices $N_1, N_2 \in \mathbb{C}^{n\times n}$ given below:

$$N_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{6.2.6}$$

$$N_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{6.2.7}$$

It is easy to see that

$$\begin{aligned} N_1^2 &= 0_{2\times 2} \\ N_2^2 &= 0_{2\times 2} \end{aligned}$$

Thus both are nilpotent matrices both with order of nilpotency as 2. However, we have,

$$N_1 + N_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

This matrix has characteristic polynomial given by

$$\lambda^2 - 1 = 0$$

Since there are two eigenvalues $\lambda_1 = 1,\ \lambda_2 = -1$, both eigenvalues having algebraic and geometric multiplicity both as 1, the matrix is diagonalizable. Since this is not the zero matrix, by the Theorem 6.2.2, this matrix cannot be nilpotent. Thus we see that the sum $N_1 + N_2$ of the two nilpotent matrices $N_1$ and $N_2$ is not nilpotent. Similarly we have,

$$N_1 N_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and this is a diagonal matrix. Again by Theorem 6.2.2 it follows that $N_1 N_2$ is not diagonalizable. Thus the product $N_1 N_2$ of the two nilpotent matrices $N_1$ and $N_2$ is not nilpotent.

However, we shall now see that when two nilpotent matrices commute with each other, then their sum and product will also be nilpotent. Suppose $N_1, N_2 \in \mathcal{F}^{n\times n}$ are nilpotent matrices with orders of nilpotency as $r_1$ and $r_2$. Then we have

$$\begin{aligned} N_1^r &= 0 \text{ for all } r \geq r_1 && (6.2.8)\\ N_2^r &= 0 \text{ for all } r \geq r_2 && (6.2.9)\\ & && (6.2.10) \end{aligned}$$

We then have for any positive integer $m$, using commutativity,

$$(N_1 + N_2)^m = \sum_{r=0}^{m} N_1^{(m-r)} N_2^r$$

If $m \geq r_1 + r_2$, then in the terms in the sum with $r = 0, 1, 2, \cdots, r_2$, we have $m - r \geq (r_1 + r_2) - (r_2)$ and hence $(m - r) \geq r_1$. Consequently the first factor $N_1^{(m-r)}$ is zero in all these terms. For the terms in the sum with $r \geq r_2$ we have the second factor is $0_{n\times n}$. Thus, if $m \geq r_1 + r_2$, then each term is zero and hence $(N_1 + N_2)^m = 0_{n\times n}$ and hence $N_1 + N_2$ is nilpotent.
Similarly we have for any positive integer $r$, using ommutativity,

$$(N_1 N_2)^r = N_1^r N_2^r$$

Hence if $m = Min.\{r_1, r_2\}$ we get $(N_1 N_2)^m = 0_{n\times n}$ and hence $N_1 N_2$ is nilpotent. Thus we have,

**The sum and product of any two commuting nilpotent matrices are also nilpotent**

We shall now use this result to investigate the uniqueness of the decomposition of a matrix as the sum of a diagonalizable matrix $D$ and a nilpotent matrix $N$ obtained in Theorem 6.1.2. In Theorem 6.1.2 the diagonalizable matrix $D$ we got was given by (6.1.22), and hence it is a polynomial in $A$. Further the nilpotent matrix $N$ we obtained was defined as $N = A - D$, and hence $N$ is also a polynomial in $A$. Thus $D$, $N$ and $A$ all commute with each other. Thus we have the decomposition

$$\begin{aligned} A &= D + N \text{ where} && (6.2.11)\\ DN &= ND && (6.2.12) \end{aligned}$$

Suppose now there exists another decomposition of $A$ as

$$A = D_1 + N_1 \qquad (6.2.13)$$

where $D_1$ is diagonalizable and $N_1$ is nilpotent and $D_1$ and $N_1$ commute with each other, that is,

$$D_1N_1 \;=\; N_1D_1 \tag{6.2.14}$$

Then we get from (6.2.13),

$$\begin{aligned} D_1A &= D_1(D_1+N_1) \\ &= D_1^2 + D_1N_1 \\ &= D_1^2 + N_1D_1 \text{ by (6.2.14)} \\ &= (D_1+N_1)D_1 \\ &= AD_1 \end{aligned}$$

Hence $D_1$ commutes with $A$ and hence with any polynomial in $A$. Thus $D_1$ commutes with $D$ and $N$. Similarly it can be shown that $N_1$ commutes with $D$ and $N$. Thus we have $N - N_1$ is nilpotent. We can also show that the sum of two commuting diagonalizable matrices is diagonalizable and hence $D - D_1$ is diagonalizable. Now we have from (6,2,11) and (6,2,13),

$$\begin{aligned} A - A &= (D-N) - (D_1 - N_1) \\ \Longrightarrow & \\ 0_{n\times n} &= (D-D_1) - (N-N_1) \\ \Longrightarrow & \\ D - D_1 &= N - N_1 \end{aligned}$$

The lhs above is a diagonalizable whereas the rhs is nilpotent, and since they are equal we must have both equal to $0_{n\times n}$, since we have seen that the only matrix which is both diagonalizable and nilpotent is the zero matrix. Thus we have,

$$D \;=\; D_1 \tag{6.2.15}$$

$$N \;=\; N_1 \tag{6.2.16}$$

Thus the decomposition of a matrix $A \in \mathcal{F}^{n\times n}$ as the sum of a diagonalizable and nilpotent matrix is unique if these two matrices have to commute. Thus we have

**Theorem 6.2.3 Every matrix $\boldsymbol{A} \in \boldsymbol{\mathcal{F}^{n\times n}}$, can be decomposed as the sum,**

$$\boldsymbol{A} \;=\; \boldsymbol{D} + \boldsymbol{N} \tag{6.2.17}$$

**of a diagonalizable matrix $D$ and a nilpotent matrix $N$. Moreover such a decomposition is unique if**

$$DN = ND \tag{6.2.18}$$

**We the call $D$ as the "Diagonalizable Part" of $A$ and $N$ as the "Nilpotent Part" of $A$**

## 6.3 Structure of Nilpotent Matrcies

Suppose $N \in \mathcal{F}^{n\times n}$ is nilpotent. Then we know that its characteristic polynomial must be $c_{_A}(\lambda) = \lambda^n$. The order of nilpotency $\gamma_{_N}$ satisfies,

$$1 \leq \gamma_{_N} \leq n \tag{6.3.1}$$

When the order of nilpotency is 1 we have $N^1 = 0_{n\times n}$ and hence $N$ is the zero matrix. Thus the only nilpotent matrix with order of nilpotency 1 is the zero matrix.
We shall next look at nilpotent matrices whose order of nilpotency is $n$, that is $\gamma_{_N} = n$. For such matrices we have

$$\begin{aligned} N^n &= 0_{n\times n} \text{ and} && (6.3.2)\\ N^r &\neq 0_{n\times n} \text{ for } 1 \leq r \leq n-1 && (6.3.3)\end{aligned}$$

In particular we have $N^{(n-1)} \neq 0_{n\times n}$. Hence there exists a nonzero vector $x \in \mathcal{F}^n$ such that $N^{(n-1)}x \neq \theta_n$. Now we define $n$ nonzero vectors $\{v_j\}_{j=1}^n$ as follows:

$$\left.\begin{array}{ccl} v_1 & = & N^{(n-1)}x \\ v_2 & = & N^{(n-2)}x \\ \cdots & \cdots & \cdots \\ v_j & = & N^{((n-j))}x \\ \cdots & \cdots & \cdots \\ v_n & = & x \end{array}\right\} \tag{6.3.4}$$

We now claim the following:
<u>**CLAIM**</u>:
**$v_1, v_2, \cdots, v_n$ are linearly independent**

We shall prove this Claim later. But suppose this claim is true then we get that $v_1, v_2, \cdots, v_n$ form a basis for $\mathcal{F}^n$. We then have, from (6.3.4),

$$\left.\begin{array}{lcl} Nv_1 & = & \theta_n \\ Nv_2 & = & v_1 \\ \cdots & \cdots & \cdots \\ Nv_j & = & v_{(j-1)} \\ \cdots & \cdots & \cdots \\ Nv_n & = & v_{(n-1)} \end{array}\right\} \tag{6.3.5}$$

If we now define the matrix $P$ as the matrix whose $j$th column is $v_j$, for $j = 1, 2, \cdots, n$, that is,

$$P \quad = \quad [v_1 \; v_2 \; \cdots \; v_j \; \cdots \; v_n] \tag{6.3.6}$$

then $P$ is invertible, (since the columns are linearly independent), and,

$$\begin{array}{rcl} NP & = & [Nv_1 \; Nv_2 \; \cdots \; Nv_j \; \cdots \; Nv_n] \\ & = & [\theta_n \; v_1 \; v_2 \; \cdots \; v_{(j-1)} \; \cdots \; v_{(n-1)}] \\ & = & P \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix} \\ \Longrightarrow & & \\ P^{-1}NP & = & \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix} \end{array}$$

We now define the matrix $N_n$ as

$$N_n \quad \stackrel{def}{=} \quad \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix} \tag{6.3.7}$$

and call it the "**canonical** $n \times n$ **nilpotent matrix**". This matrix is an $n \times n$ matrix and has 0 all its entries, except the entries immediately to the right of the diagonal which are all 1. Thus we have,

**Theorem 6.3.1 If $N \in \mathcal{F}^{n\times n}$ is nilpotent with order of nilpotency as $n$, then it is similar to the canonical $n \times n$ nilpotent matrix $N_n$, that is, there exists an invertible matrix $P \in \mathcal{F}^{n\times n}$ such that**

$$P^{-1}NP = N_n \tag{6.3.8}$$

**We say that "canonical form of $N$ is $N_n$".**

We shall complete the proof of the above argument by proving Claim 2:
<u>**Proof of Claim**</u>:
We have

$$\begin{aligned}
\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n &= \theta_n \Longrightarrow \\
\alpha_1 N^{(n-1)}x + N^{(n-2)}x + \cdots + \alpha_{(n-1)}Nx + \alpha_n x &= \theta_n \Longrightarrow \\
N^{(n-1)}(\alpha_1 N^{(n-1)}x + N^{(n-2)}x + \cdots + \alpha_{(n-1)}Nx + \alpha_n x) &= \theta_n
\end{aligned}$$

(since all terms except the last have factor $N^n$ which is $0_{n\times n}$) $\Longrightarrow$

$$\begin{aligned}
\alpha_n N^{(n-1)}x &= \theta_n \Longrightarrow \\
(\text{since } N^{(n-1)}x \neq \theta_n)\ \alpha_n &= 0 \Longrightarrow \\
\alpha_1 N^{(n-1)}x + N^{(n-2)}x + \cdots + \alpha_{(n-1)}Nx &= \theta_n \Longrightarrow \\
N^{(n-2)}(\alpha_1 N^{(n-1)}x + N^{(n-2)}x + \cdots + \alpha_{(n-1)}Nx) &= \theta_n \Longrightarrow \\
(\text{as above})\ \alpha_{(n-1)} &= 0
\end{aligned}$$

Continuing this process step by step we get all $\alpha_j$ as 0 and hence $v_1, v_2, \cdots, v_n$ are linearly independent, thus proving the Claim.
We next look at nilpotent matrices for which the order of nilpotency is between 1 and $n$, that is,

$$1 < \gamma_{_N} < n \tag{6.3.9}$$

We then have,

$$\begin{aligned}
N^{\gamma_N} &= 0_{n\times n} \text{ and} && (6.3.10)\\
N^r &\neq 0_{n\times n} \text{ for } 1 \leq r \leq \gamma_{_N} - 1 && (6.3.11)
\end{aligned}$$

From this we get

$$\begin{aligned} Null\ Space\ of\ N^{\gamma_N} &= \mathcal{F}^n && (6.3.12)\\ Null\ Space\ of\ N^{r} &\neq \mathcal{F}^n \text{ for } 1 \leq r \leq \gamma_N - 1 && (6.3.13) \end{aligned}$$

We further have,

$$\left.\begin{array}{l} Null\ Space\ of\ N^{r} \text{ is a proper subset of } Null\ Space\ of\ N^{(r+1)} \\ \text{for } 1 \leq r \leq \gamma_N - 1 \end{array}\right\} \quad (6.3.14)$$

Thus if we define,

$$V_j = Null\ Space\ of\ N^{j} \text{ for } 1 \leq j \leq \gamma_N \qquad (6.3.15)$$

then we have

$$V_1 \overset{\subset}{\neq} V_2 \overset{\subset}{\neq} \cdots \overset{\subset}{\neq} V_j \overset{\subset}{\neq} V_{(j+1)} \overset{\subset}{\neq} \cdots \overset{\subset}{\neq} V_{\gamma_N - 1} \overset{\subset}{\neq} V_{\gamma_N} \qquad (6.3.16)$$

Let us define

$$d_j = dim.(V_j),\ 1 \leq j \leq \gamma_N \qquad (6.3.17)$$

Then we have

$$d_1 < d_2 < \cdots < d_{\gamma_N - 1} < d_{\gamma_N} \qquad (6.3.18)$$

We then define

$$\alpha_j = d_j - d_{j-1} \text{ for } 2 \leq j \leq \gamma_N \qquad (6.3.19)$$

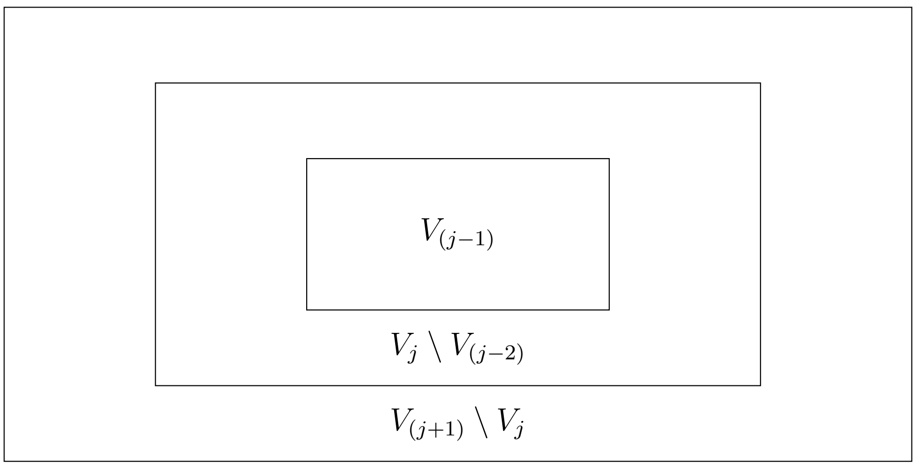
These $\alpha_j$ give us the extra dimension we get when we move up from $V_{(j-1)}$ to $V_j$. We shall define $\alpha_1 = g_1$ and $\alpha_{\gamma_N+1} = 0$. Then we can write

$$\left.\begin{array}{ll} d_1 = & \alpha_1 \text{ (the geometric multiplicity of the eiegnvalue 0)} \\ d_2 = & d_1 + \alpha_2 \\ d_3 = & d_1 + \alpha_2 + \alpha_3 \\ \cdots & \cdots\cdots \\ d_{(j-1)} = & d_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{(j-1)} \\ d_j = & d_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{(j-1)} + \alpha_j \\ d_{(j+1)} = & d_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{(j-1)} + \alpha_j + \alpha_{(j+1)} \\ \cdots & \cdots\cdots\cdots\cdots\cdots \\ d_{(\gamma_N-1)} = & d_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{(j-1)} + \alpha_j + \alpha_{(j+1)} + \cdots + \alpha_{\gamma_N-1} \\ d_{\gamma_N} = & d_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{(j-1)} + \alpha_j + \alpha_{(j+1)} + \cdots + \alpha_{\gamma_N-1} + \alpha_{\gamma_N} = n \\ d_{(\gamma_N+1)} = & d_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{(j-1)} + \alpha_j + \alpha_{(j+1)} + \cdots + \alpha_{\gamma_N-1} + \alpha_{\gamma_N} + \alpha_{(\gamma_N+1)} \end{array}\right\} \quad (6.3.20)$$

These $\alpha_j$ have a nice structure as follows:

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_j \geq \alpha_{(j+1)} \geq \cdots \geq \alpha_{\gamma_N - 1} \geq \alpha_{\gamma_N} = \alpha_{(\gamma_N + 1)} \tag{6.3.21}$$

This means that the extra dimension that we acquire from moving up from a $V_{(j-1)}$ to $V_j$ is at least as mush as the extra dimension that we acquire in the next stage of moving up from $V_j$ to $V_{(j+1)}$. We shall now prove this fact:




Let $\mathcal{B} = v_1, v_2, \cdots, v_{d_j}$ be a basis for $V_j$. We can now find suitable $\alpha_{(j+1)}$ linearly independent vectors in $V_{(j+1)} \setminus V_j$ and append it to $\mathcal{B}$ to get a basis for $V_{(j+1)}$ as follows:
We first pick a vector $u_1$ in $V_{(j+1)} \setminus V_j$. This is obviously nonzero. We then look at the subspace $K_1 \subseteq V_{(j+1)}$ spanned by $S_1 = \mathcal{B} \cup \{u_1\}$. Then we pick a vector $u_1$ in $V_{(j+1)} \setminus K_1$ and so on. Thus we get at the $\alpha_{(j+1)}$ stage vectors $u_1, u_2, \cdots, u_{\alpha_{(j+1)}}$ such that

1. $\mathcal{B} \cup \left\{u_1, u_2, \cdots, u_{\alpha_{j+1}}\right\}$ is a basis for $V_{(j+1)}$,

2. $u_1 \notin V_j$, and

3. For $2 \leq r \leq \alpha_{(j+1)}$ we have $u_r \notin K_{(r-1)}$, where $K_{(r-1)}$ is the subspace spanned by $B \cup \left\{u_1, u_2, u_{(r-1)}\right\}$

Using these properties we prove the following:

**Lemma 6.3.1 The vectors $Nu_1, Nu_2, \cdots, Nu_{\alpha_{j+1}}$ are linearly independent vectors in $V_j \setminus V_{(j-1)}$**

Proof:
We have

$$\begin{aligned}
\sum_{r=1}^{\alpha_{(j+1)}} a_r N u_r &= \theta_n \Longrightarrow \\
N\left(\sum_{r=1}^{\alpha_{(j+1)}} a_r u_r\right) &= \theta_n \Longrightarrow \\
\sum_{r=1}^{\alpha_{(j+1)}} a_r u_r &\in \textit{Null Space of } N \Longrightarrow \\
\sum_{r=1}^{\alpha_{(j+1)}} a_r u_r &\in V_1 \Longrightarrow \\
\sum_{r=1}^{\alpha_{(j+1)}} a_r u_r &\in V_j \text{ since } V_1 \subset V_j \Longrightarrow \\
\sum_{r=1}^{\alpha_{(j+1)}} a_r u_r &= \sum_{t=1}^{d_j} b_t v_t
\end{aligned}$$

Supose now $\alpha_{(j+1)} \neq 0$. Then we get from above,

$$u_{(j+1)} = (\alpha_{(j+1)})^{-1}\left\{-\sum_{r=1}^{\alpha_j} a_r u_r + \sum_{t=1}^{d_j} b_t v_t\right\}$$

which is a contradiction to Property 3 above of our choice of the $u_j$ vectors. Hence $\alpha_{(j+1)} = 0$. We therefore get

$$\begin{aligned}
N\left(\sum_{r=1}^{\alpha_{(j+1)}} a_r u_r\right) = \theta_n &\Longrightarrow \alpha_{(j+1)} = 0 \Longrightarrow \\
N\left(\sum_{r=1}^{\alpha_j} a_r u_r\right) = \theta_n &\Longrightarrow \alpha_j = 0 \text{ (by similar argument as before)}
\end{aligned}$$

Continuing this process we get all $\alpha_r$ are zero. Thus we have the lemma. From the above it now follows that the dimension of $V_j$ exceeds that of $V_{(j-1)}$ by at least $\alpha_{(j+1)}$ and hence we get $\alpha_j \geq \alpha_{(j+1)}$, thus proving (6.3.21).

We now look at the differences of the dimensions gained at two successive stages. More precisely we define,

$$\left.\begin{array}{lcl} n_1 & = & \alpha_1 - \alpha_2 \\ n_2 & = & \alpha_2 - \alpha_3 \\ n_3 & = & \alpha_3 - \alpha_2 \\ \cdots & \cdots & \cdots \\ n_j & = & \alpha_j - \alpha_{(j+1)} \\ \cdots & \cdots & \cdots \\ n_{\gamma_N - 1} & = & \alpha_{(\gamma_N - 1)} - \alpha_{\gamma_N} \\ n_{\gamma_N} & = & \alpha_{\gamma_N} - \alpha_{(\gamma_N + 1)} = \alpha_{\gamma_N} \end{array}\right\} \tag{6.3.22}$$

(We observe that $\alpha_{\gamma_N}$ is at least 1 and hence $n_{\gamma_N}$ is at least one). We now state the following theorem (without proof):

**Theorem 6.3.2 With the above notations, there exists an invertible matrix, $P \in \mathcal{F}^{n\times n}$ such that $P^{-1}NP = N_{can}$, where $N_{can}$ is an $n \times n$ matrix which is built as follows: It is made up of diagonal blocks of canonical nilpotent matrices of different sizes such that,**

1. **The blocks are arranged in non increasing order as we go down the diagonal,**

2. **The leading block is $N_{\gamma_N}$, the canonical $\gamma_N \times \gamma_N$ nilpotent matrix, (and all other blocks can be at most of this size),**

3. **For $1 \leq j \leq \gamma_N$, There are $n_j$ blocks of size $j \times j$**

**$N_{can}$ is called the "Canonical Form" of the nilpotent matrix $N$.**

From the above we can verify the following facts:

1. **The total number of blocks is equal to $g_1 (= \alpha_1)$, the geometeric multiplicity of the eigenvalue 0**:
   Reason:
   We have

$$\begin{array}{rcl} \text{Total number of blocks} & = & n_1 + n_2 + +n_{(\gamma_N - 1)} + n_{\gamma_N} \\ & = & (\alpha_1 - \alpha_2) + (\alpha_2 - \alpha_1) + \cdots + \alpha_{\gamma_N - 1} - \alpha_{\gamma_N} + \alpha_{\gamma_N} \\ & = & \alpha_1 \end{array}$$

2. **Block sizes all add up to $n$**
Reason:
We have

$$\begin{aligned}\text{The sum of all the block sizes} &= n_1 + 2n_2 + 3n_3 + \cdots + (\gamma_{_N} - 1)n_{(\gamma_{_N}-1)} + \gamma_{_N}\alpha_{\gamma_{_N}} \\ &= (\alpha_1 - \alpha_2) + 2(\alpha_2 - \alpha_3) + \cdots + \\ &\quad +(\gamma_{_N} - 1)(\alpha_{(\gamma_{_N}-1)} - \alpha_{\gamma_{_N}}) + \gamma_{_N}\alpha_{\gamma_{_N}} \\ &= \alpha_1 + \alpha_2 + \cdots + \alpha_{(\gamma_{_N}-1)} + \alpha_{\gamma_{_N}} \\ &= n\end{aligned}$$

3. **The number of blocks,$\boldsymbol{n_j}$, of size $\boldsymbol{j \times j}$ can also be written as**

$$\begin{aligned}\boldsymbol{n_j} &= \boldsymbol{2d_j - d_{(j-1)} - d_{(j+1)}} && (6.3.23) \\ &= \boldsymbol{2dim.(V_j) - dim.(V_{(j-1)}) - dim.(V_{(j+1)})} && (6.3.24)\end{aligned}$$

Reason:
We have

$$\begin{aligned}n_j &= \alpha_j - \alpha_{(j+1)} \text{ (by (6.3.22))} \\ &= (d_j - d_{(j-1)}) - (d_{(j+1)} - d_j) \\ &= 2d_j - d_{(j-1)} - d_{(j+1)} \\ &= 2dim.(V_j) - dim.(V_{(j-1)}) - dim.(V_{(j+1)})\end{aligned}$$

**Summarizing we see that the canonical form, of a nilpotent matrix $\boldsymbol{N \in \mathcal{F}^{n\times n}}$, with order of nilpotency as $\boldsymbol{\gamma_{_N}}$, has the following structure:**

1. **It is an $\boldsymbol{n \times n}$ matrix, divided into a number of diagonal blocks of canonical nilpotent matrices, arranged in such a way that as we go down the diagonal, the block sizes are nonincreasing,**

2. **The total number of blocks is equal to the dimension of the Null Space of $\boldsymbol{N}$ (= the nullity of $\boldsymbol{N}$=geometric multiplicity of the only eigenvalue 0 of $\boldsymbol{N}$),**

3. **The leading block is of size $\boldsymbol{\gamma_{_N} \times \gamma_{_N}}$, and all the remaining blocks of size at most $\boldsymbol{\gamma_{_N} \times \gamma_{_N}}$, and**

4. **For $1 \leq j \leq \gamma_{N}$, the number of blocks, $n_j$, of size $j \times j$, is given by**

$$\begin{aligned} \boldsymbol{n_j} &= \boldsymbol{2d_j - d_{(j-1)} - d_{(j+1)}} && (6.3.25) \\ &= \boldsymbol{2dim.(V_j) - dim.(V_{(j-1)}) - dim.(V_{(j+1)})} && (6.3.26) \end{aligned}$$

**where**

$$\boldsymbol{V_j} = \boldsymbol{Null\ Space\ of\ N^j} \textbf{ (for } \boldsymbol{1 \leq j \leq \gamma_{N}}\textbf{)} \quad (6.3.27)$$

**Example 6.3.1** Conside the matrix,

$$N = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & -1 & -1 & 1 \\ 0 & -1 & -1 & 0 \end{pmatrix} \quad (6.3.28)$$

We have,

$$N^2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 \end{pmatrix}$$

Further we have

$$N^3 = 0_{4\times 4}$$

Hence $N$ is a nilpotent matrix with order of nilpotency, $\gamma_{N} = 3$. Hnce we have

$$\begin{aligned} c_{N}(\lambda) &= \lambda^4 \\ m_{N}(\lambda) &= \lambda^3 \end{aligned}$$

In the canonical form, the leading block is of size $\gamma_{N} \times \gamma_{N} = 3 \times 3$. Since the matrix $N$ is $4 \times 4$, the next block has to be of size $1 \times 1$. Hence the canonical form of $N$ is given by,

$$N_{can} = \left(\begin{array}{ccc|c} 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & \\ \hline & & & 0 \end{array}\right) \quad (6.3.29)$$

We can easily compute,

$$\begin{aligned}
V_1 = Null\ Space\ of\ N &= \left\{ x = \begin{pmatrix} \alpha \\ \beta \\ -\beta \\ -\alpha \end{pmatrix} : \alpha, \beta \in \mathcal{F} \right\} \\
V_2 = Null\ Space\ of\ N^2 &= \left\{ x = \begin{pmatrix} \gamma \\ \ell \\ m \\ -\gamma \end{pmatrix} : \ell, m, \gamma \in \mathcal{F} \right\} \\
V_3 = Null\ Space\ of\ N^3 &= \mathcal{F}^4
\end{aligned}$$

Hence we get,

$$\begin{aligned}
d_1 = dim.(V_1) &= 2 \\
d_2 = dim.(V_2) &= 3 \\
d_3 = dim.(V_3) &= 4
\end{aligned}$$

Hence we have,

$$\begin{aligned}
\alpha_1 = d_1 &= 2 \\
\alpha_2 = d_2 - d_1 &= 1 \\
\alpha_3 = d_3 - d_2 &= 1
\end{aligned}$$

This gives us,

$$n_1 = \alpha_1 - \alpha_2 = 2 - 1 = 1 \tag{6.3.30}$$
$$n_2 = \alpha_2 - \alpha - 3 = 1 - 1 = 0 \tag{6.3.31}$$
$$n_3 = \alpha_3 = 1$$

Thus we have,

$$\begin{aligned}
Number\ of\ blocks\ of size\ 3 \times 3 &= n_3 = 1 \\
Number\ of\ blocks\ of size\ 2 \times 2 &= n_2 = 0 \\
Number\ of\ blocks\ of size\ 1 \times 1 &= n_1 = 1
\end{aligned}$$

Thus there is one $3 \times 3$ block and one $1 \times 1$ block and there is no $2 \times 2$ block, as seen in (6.3.29)

**Example 6.3.2** Consider the matrix,

$$N = \begin{pmatrix} 1 & 1 & -2 & 0 & 1 & -1 \\ 3 & 1 & 5 & 1 & -1 & 3 \\ -2 & -1 & 0 & 0 & -1 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ -5 & -3 & -1 & -1 & -1 & -1 \\ -3 & -2 & -1 & -1 & 0 & -1 \end{pmatrix}$$

We can compute $N^2$ to get

$$N^2 = \begin{pmatrix} 6 & 3 & 3 & 1 & 1 & 2 \\ -6 & -3 & -3 & -1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -6 & -3 & -3 & -1 & -1 & -2 \\ -6 & -3 & -3 & -1 & -1 & -2 \end{pmatrix}$$

Finally we have

$$N^3 = 0_{6\times 6}$$

Thus $N$ is a nilpotent matrix with order of nilpotency, $\gamma_N = 3$. Hence we get

$$\begin{aligned} c_N(\lambda) &= \lambda^6 \\ m_N(\lambda) &= \lambda^3 \end{aligned}$$

The cnaonical form will have the leading block as a $3 \times 3$ block. Since the matrix is $6 \times 6$ the remaining $3 \times 3$ diagonal block may be divided into either a single $3 \times 3$ block; or as one $2 \times 2$ block and one $1 \times 1$; or a three $1 \times 1$ blocks. We shall determine which of these holds for this matrix. We can compute,

$$\begin{aligned} V_1 &= \textit{Null Space of } N \\ &= \left\{ \alpha \begin{pmatrix} 2 \\ -4 \\ -1 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} -4 \\ 5 \\ 2 \\ 0 \\ 3 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 1 \\ -2 \\ -2 \\ 0 \\ 0 \\ 3 \end{pmatrix} : \alpha, \beta, \gamma \in \mathcal{F} \right\} \end{aligned}$$

$$
\begin{aligned}
V_2 &= Null\ Space\ of\ N^2\\
&= \left\{ \alpha \begin{pmatrix} 1\\0\\0\\0\\-6\\0 \end{pmatrix} + \beta \begin{pmatrix} 0\\1\\0\\0\\-3\\0 \end{pmatrix} + \gamma \begin{pmatrix} 0\\0\\1\\0\\-3\\0 \end{pmatrix} + +\delta \begin{pmatrix} 0\\0\\0\\1\\-1\\0 \end{pmatrix} + \epsilon \begin{pmatrix} 0\\0\\0\\0\\-2\\1 \end{pmatrix} : \alpha, \beta, \gamma, \delta, \epsilon \in \mathcal{F} \right\}\\
V_3 &= Null\ Space\ of\ N^3\\
&= \mathcal{F}^6
\end{aligned}
$$

Hence we get

$$
\begin{aligned}
d_1 = dim.(V_1) &= 3\\
d_2 = dim.(V_2) &= 5\\
d_3 = dim.(V_3) &= 6
\end{aligned}
$$

Consequently we have,

$$
\begin{aligned}
\alpha_1 &= d_1 = 3\\
\alpha_2 &= d_2 - d_2 = 2\\
\alpha_3 &= d_3 - d_2 = 1
\end{aligned}
$$

Hence we get,

$$
\begin{aligned}
n_3 &= \alpha_3 = 1\\
n_2 &= \alpha_2 - \alpha_3 = 2 - 1 = 1\\
n_1 &= \alpha_1 - \alpha_2 = 3 - 2 = 1
\end{aligned}
$$

Thus there is one $3 \times 3$ block, one $2 \times 2$ block and one $1 \times 1$ block. Hence the canonical form of $N$ is given by,

$$
N_{can} = \left( \begin{array}{ccc|cc|c}
0 & 1 & 0 & & & \\
0 & 0 & 1 & & & \\
0 & 0 & 0 & & & \\ \hline
 & & & 0 & 1 & \\
 & & & 0 & 0 & \\ \hline
 & & & & & 0
\end{array} \right)
$$

## 6.4 Jordan Canonical Form

We shall now briefly look at the Jordan cannonical form of a general matrix $A \in \mathcal{F}^{n\times n}$ whose characteristic polynomial and minimal polynomial are given as usual by

$$c_{_A}(\lambda) = (\lambda-\lambda_1)^{a_1}(\lambda-\lambda_2)^{a_2}\cdots(\lambda-\lambda_k)^{a_k} \tag{6.4.1}$$

$$m_{_A}(\lambda) = (\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k} \tag{6.4.2}$$

In Theorem 6.2.3 we have seen that any such $A$ can be written as the sum $D+N$ of a diagonalizable matrix $D$ and a nilpotent matrix $N$ and that this decomposition is unique if $D$ and $N$ commute. The simplest such sum is one where $D$ is a diagonal matrix and $N$ is a canonical nilpotent matrix. The simplest digonal $D$ is the diagonal matrix all of whose diagonal entries are the same, say $a$. Thus the simplest such general matrix $D+N$ we can think of is a matrix whose diagonal entries are all $a$ and the entries to the right of the diagonal are 1 and all other entries are zero. Such a matrix is called a "canonical Jordan matrix". We have

**Definition 6.4.1 The $\boldsymbol{m\times m}$ matrix given below is called the "CANONICAL $\boldsymbol{m\times m}$ JORDAN MATRIX with diagonal $\boldsymbol{a}$" and is denoted by $\boldsymbol{J_m(a)}$:**

$$\boldsymbol{J_m(a)} = \begin{pmatrix} \mathbf{a} & \mathbf{1} & & & \\ & \mathbf{a} & \mathbf{1} & & \\ & & \ddots & \ddots & \\ & & & \mathbf{a} & \mathbf{1} \\ & & & & \mathbf{a} \end{pmatrix}_{\boldsymbol{m\times m}}$$

It turns out that all matrices $A \in \mathcal{F}^{n\times n}$ whose characteristic polynomial and minimal polynomial are given by (6.4.1) and (6.4.2) can be built by such canonical Jordan matrices as follows:
We can find an invertible matrix $P \in \mathcal{F}^{n\times n}$ such that

$$P^{-1}AP = J$$

where $J$ has the structure which we describe in detail below:
$J \in \mathcal{F}n\times n$ is divided into $k$ diagonal blocks $A^{(1)}, A^{(2)}, \cdots, A^{(j)}, \cdots, A^{(k)}$,

one block corresponding to each of the $k$ eignevalues $\lambda_1, \lambda_2, \cdots \lambda_j, \cdots, \lambda_k$ respectively. Thus we have

$$P^{-1}AP \quad = \left(\begin{array}{c|c|c|c|c|c}
\mathrm{A}^{(1} & & & & & \\ \hline
 & \mathrm{A}^{(2)} & & & & \\ \hline
 & & \ddots & & & \\ \hline
 & & & \mathrm{A}^{(j)} & & \\ \hline
 & & & & \ddots & \\ \hline
 & & & & & \mathrm{A}^{(k)}
\end{array}\right)$$

How is each of these blocks structured?
We shall describe, in general, how $A^{(j)}$ is structured.
$A^{(j)}$ is further subdivided into diagonal subblocks , each subblock being a canonical Jordan matrix with diagonal $\lambda_j$ . What remains to be answered are the following questions:

1. What is the size of $A^{(j)}$?,
2. How many subblocks are there in $A^{(j)}$?,
3. How are these subblocks arranged?, and
4. What are the sizes of these blocks?

We now proceed to answer these questions.
**Question 1:**
**What is the size of $\boldsymbol{A^{(j)}}$?**
$A^{(j)}$ will be an $a_j \times a_j$ matrix
**Question 2:**
**How many subblocks are there in $\boldsymbol{A^{(j)}}$?**
The number of subblocks in $A^{(j)}$ is equal to the geometric multiplicity $g_j$ of the eigenvalue $\lambda_j$.
Note:
Hence, we observe that if $g_j = a_j$ then each subblock is just $1 \times 1$ and therefore $A^{(j)}$ will just be an $a_j \times a_j$ diagonal matrix all of whose diagonal entries are $\lambda_j$
**Question 3:**
**How are these subblocks arranged?**
These subblocks are arranged in such a way that as we go down the diagonal

in $A^{(j)}$, the sizes of the subblocks are nonincreasing
It remains to answer the fourth question.
**Question 4:**
**What are the sizes of these blocks?**
The structure is analogous to what we did for nilpotent matrices. We define the follwing subspaces

$$V_r^{(j)} \quad = \quad Null\ Space\ of\ (A - \lambda_j I_{n\times n})^r \text{ for } 1 \le r \le r_j \tag{6.4.3}$$

We then define the dimensions of these subspaces,

$$d_r^{(j)} \quad = \quad dim.(V_r^{(j)}),\ 1 \le j \le r_j \tag{6.4.4}$$

Then we have

$$g_j = d_1^{(j)} < d_2^{(j)} < \cdots < d_{r_j-1}^{(j)} < d_{r_j}^{(j)} = a_j \tag{6.4.5}$$

We then define

$$\alpha_r^{(j)} \quad = \quad d_r^{(j)} - d_{r-1}^{(j)} \text{ for } 2 \le r \le r_j \tag{6.4.6}$$

These $\alpha_r^{(j)}$ give us the extra dimension we get when we move up from $V_{(r-1)}^{(j)}$ to $V_r^{(j)}$. We shall define $\alpha_1^{(j)} = g_j$ and $\alpha_{r_j+1}^{(j)} = 0$. Then we can write

$$\left.\begin{array}{ll}
d_1^{(j)} = & \alpha_1^{(j)} \text{ (the geometric multiplicity of the eigenvalue 0)} \\
d_2^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} \\
d_3^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} + \alpha_3^{(j)} \\
\cdots & \cdots\cdots \\
d_{(r-1)}^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} + \alpha_3^{(j)} + \cdots + \alpha_{(r-1)}^{(j)} \\
d_r^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} + \alpha_3^{(j)} + \cdots + \alpha_{(r-1)}^{(j)} + \alpha_r^{(j)} \\
d_{(r+1)}^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} + \alpha_3^{(j)} + \cdots + \alpha_{(r-1)}^{(j)} + \alpha_r^{(j)} + \alpha_{(r+1)}^{(j)} \\
\cdots & \cdots\cdots\cdots\cdots\cdots \\
d_{(r_j-1)}^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} + \alpha_3^{(j)} + \cdots + \alpha_{(r-1)}^{(j)} + \alpha_r^{(j)} + \alpha_{(r+1)}^{(j)} + \cdots + \alpha_{(r_j-1)}^{(j)} \\
d_{r_j}^{(j)} = & d_1^{(j)} + \alpha_2^{(j)} + \alpha_3^{(j)} + \cdots + \alpha_{(r-1)}^{(j)} + \alpha_r^{(j)} + \alpha_{(r+1)}^{(j)} + \cdots + \alpha_{(r_j-1)}^{(j)} + \alpha_{r_j}^{(j)} = a_j
\end{array}\right\} \tag{6.4.7}$$

These $\alpha_r^{(j)}$ have a nice structure as follows:

$$\alpha_1^{(j)} \ge \alpha_2^{(j)} \ge \cdots \ge \alpha_r^{(j)} \ge \alpha_{(r+1)}^{(j)} \ge \cdots \ge \alpha_{(r_j-1)}^{(j)} \ge \alpha_{r_j}^{(j)} \tag{6.4.8}$$

This means that the extra dimension that we acquire from moving up from $V_{(r-1)}^{(j)}$ to $V_r^{(j)}$ is at least as much as the extra dimension that we acquire in the next stage of moving up from $V_r^{(j)}$ to $V_{(r+1)}^{(j)}$.
We now look at the differences of the dimensions gained at two successive stages. More precisely we define,

$$\left.\begin{array}{lcl} n_1^{(j)} & = & \alpha_1^{(j)} - \alpha_2^{(j)} \\ n_2^{(j)} & = & \alpha_2^{(j)} - \alpha_3^{(j)} \\ n_3^{(j)} & = & \alpha_3^{(j)} - \alpha_2^{(j)} \\ \cdots & \cdots & \cdots \\ n_r^{(j)} & = & \alpha_r^{(j)} - \alpha_{(r+1)}^{(j)} \\ \cdots & \cdots & \cdots \\ n_{r_j-1}^{(j)} & = & \alpha_{(r_j-1)}^{(j)} - \alpha_{r_j}^{(j)} \\ n_{r_j}^{(j)} & = & \alpha_{r_j}^{(j)} - \alpha_{(r_j+1)}^{(j)} = \alpha_{r_j} \end{array}\right\} \tag{6.4.9}$$

(We observe that $\alpha_{r_j}^{(j)}$ is at least 1 and hence $n_{r_j}^{(j)}$ is at least one). We now state the following theorem (without proof):

**Theorem 6.4.1 With the above notations, there exists an invertible matrix, $P \in \mathcal{F}^{n\times n}$ such that $P^{-1}AP = A_{can}$, where $A_{can}$ is an $n \times n$ matrix which is built as follows: It is made up of diagonal blocks of matrices $A^{(1)}, A^{(2)}, \cdots, A^{(j)}, \cdots, A^{(k)}$. For each $j$, $(1 \le j \le k)$, the $A^{(j)}$ is divided into a number of subblocks of different sizes such that,**

1. **The subblocks are arranged in non increasing order as we go down the diagonal,**

2. **The leading subblock is $J_{r_j}(\lambda_j)$, the canonical $r_j \times r_j$ Jordan matrix with diagonal , (and all other blocks can be at most of this size),**

3. **For $1 \le r \le r_j$, there are $n_r^{(j)}$ blocks of size $r \times r$**

4. **The total number of subblocks in $A^j$ is equal to $g_j$, the geometric multiplicity of $\lambda_j$**

**$A_{can}$ is called the "Jordan Canonical Form" of the matrix $A$.**

Analogous to the nilpotent matrices we can easily see that

$$\begin{aligned} n_r^{(j)} &= \textit{number of } r \times r \textit{ subblocks in } A^{(j)} \\ &= 2dim.(V_r^{(j)}) - dim.(V_{(r-1)}^{(j)}) - dim.(V_{(r+1)}^{(j)}) \end{aligned} \tag{6.4.10}$$

<u>Reason</u>:
We have

$$\begin{aligned} n_r^{(j)} &= \alpha_r^{(j)} - \alpha_{(r+1)}^{(j)} \\ &= (d_r^{(j)} - d_{(r-1)}^{(j)}) - (d_{(r+1)}^{(j)} - d_r^{(j)}) \\ &= 2d_r^{(j)} - d_{(r-1)}^{(j)} - d_{(r+1)}^{(j)} \\ &= 2dim.(V_r^{(j)}) - dim.(V_{(r-1)}^{(j)}) - dim.(V_{(r+1)}^{(j)}) \end{aligned}$$

**Example 6.4.1** Suppose $A \in \mathcal{F}^{14\times14}$ has characteristic polynomial and minimal polynomial given by,

$$\begin{aligned} c_{_A}(\lambda) &= (\lambda-2)^5(\lambda+4)^3(\lambda-6)^6 \\ m_{_A}(\lambda) &= (\lambda-2)^3(\lambda+4)^2(\lambda-6)^3 \end{aligned}$$

Then the distinct eigenvalues are $\lambda_1 = 2$, $\lambda_2 = -4$ and $\lambda_3 = 6$, with algebraic multiplicities given by

$$a_1 = 5,\ a_2 = 3, a_3 = 6$$

Since there are three distinct eigenvalues there will be three major blocks $A^{(1)}$, $A^{(2)}$ and $A^{(3)}$ corresponding to the three distinct eigenvalues $\lambda_1 = 2, \lambda_2 = -4$ and $\lambda_3 = 6$, We have

$$A_{can} = \left(\begin{array}{c|c|c} \mathrm{A}^{(1)} & & \\ \hline & \mathrm{A}^{(2)} & \\ \hline & & \mathrm{A}^{(3)} \end{array}\right)$$

The algebraic multiplicities of the eigenvalues give us the fact that $A^{(1)}$ is $5\times5$, $A^{(2)}$ is $3\times3$ and $A^{(3)}$ is $6\times6$. So we have

$$A_{can} = \left(\begin{array}{c|c|c} (\mathrm{A}^{(1)})_{5\times5} & & \\ \hline & (\mathrm{A}^{(2)})_{3\times3} & \\ \hline & & (\mathrm{A}^{(3)})_{6\times6} \end{array}\right) \tag{6.4.11}$$

The minimal polynomial gives us

$$r_1 = 3,\ r_2 = 2 \text{ and } r_3 = 3$$

This means that,

1. The leading subblock in $A^{(1)}$ is

$$J_3(\lambda_1) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

2. The leading subblock in $A^{(2)}$ is

$$J_2(\lambda_2) = \begin{pmatrix} -4 & 1 \\ 0 & -4 \end{pmatrix}$$

3. The leading subblock in $A^{(3)}$ is

$$J_3(\lambda_3) = \begin{pmatrix} 6 & 1 & 0 \\ 0 & 6 & 1 \\ 0 & 0 & 6 \end{pmatrix}$$

To determine the remaining subblocks we need to look at the relevant subspaces $V_r^{(j)} = Null\ Space\ of\ (A - \lambda_j I_{14\times14})^r$. Suppose we are given the following data:

$$\begin{aligned} d_1^{(1)} = dim.(V_1^{(1)}) &= Null\ Space\ of\ (A - 2I_{14\times14}) = 3 \\ d_2^{(1)} = dim.(V_2^{(1)}) &= Null\ Space\ of\ (A - 2I_{14\times14})^2 = 4 \\ d_3^{(1)} = dim.(V_3^{(1)}) &= Null\ Space\ of\ (A - 2I_{14\times14})^3 = 5 \end{aligned}$$

$$\begin{aligned} d_1^{(2)} = dim.(V_1^{(2)}) &= Null\ Space\ of\ (A + 4I_{14\times14}) = 2 \\ d_2^{(2)} = dim.(V_2^{(2)}) &= Null\ Space\ of\ (A + 4I_{14\times14})^2 = 3 \end{aligned}$$

$$\begin{aligned} d_1^{(3)} = dim.(V_1^{(3)}) &= Null\ Space\ of\ (A - 6I_{14\times14}) = 3 \\ d_2^{(3)} = dim.(V_2^{(3)}) &= Null\ Space\ of\ (A - 6I_{14\times14})^2 = 5 \\ d_3^{(3)} = dim.(V_3^{(3)}) &= Null\ Space\ of\ (A - 6I_{14\times14})^3 = 6 \end{aligned}$$

Thus we have,

$$\begin{aligned}\alpha_1^{(1)} &= d_1^{(1)} = 3\\ \alpha_2^{(1)} &= d_2^{(1)} - d_1^{(1)} = 1\\ \alpha_3^{(1)} &= d_3^{(1)} - d_2^{(1)} = 1\end{aligned}$$

Hence we get

$$\begin{aligned}n_3^{(1)} &= \alpha_3^{(1)} = 1\\ n_2^{(1)} &= \alpha_2^{(1)} - \alpha_3^{(1)} = 0\\ n_1^{(1)} &= \alpha_1^{(1)} - \alpha_2^{(1)} = 2\end{aligned}$$

Thus in $A^{(1)}$, there is one subblock of $3\times 3$, (because $n_3^{(1)} = 1$), zero sublocks of size $2 \times 2$, (because $n_2^{(1)} = 0$) and two subblocks of size $1 \times 1$, (because $n_1^{(1)} = 2$). Thus we get,

$$A^{(1)} = \left(\begin{array}{ccc|c|c} 2 & 1 & 0 & & \\ 0 & 2 & 1 & & \\ 0 & 0 & 2 & & \\ \hline & & & 0 & \\ \hline & & & & 0 \end{array}\right) \tag{6.4.12}$$

Similarly we get

$$\begin{aligned}\alpha_1^{(2)} &= d_1^{(2)} = 2\\ \alpha_2^{(2)} &= d_2^{(2)} - d_1^{(2)} = 1\end{aligned}$$

Hence we get

$$\begin{aligned}n_2^{(2)} &= \alpha_2^{(2)} = 1\\ n_1^{(1)} &= \alpha_1^{(2)} - \alpha_2^{(2)} = 1\end{aligned}$$

Thus in $A^{(2)}$, there is one subblock of $2 \times 2$, (because $n_2^{(2)} = 1$), and one subblock of size $1 \times 1$, (because $n_1^{(2)} = 1$). Thus we get,

$$A^{(2)} = \left(\begin{array}{rr|c} -4 & 1 & \\ 0 & -4 & \\ \hline & & 0 \end{array}\right) \tag{6.4.13}$$

For the third eigenvalue we get,

$$\begin{array}{rcl} \alpha_1^{(3)} & = & d_1^{(3)} = 3 \\ \alpha_2^{(2)} & = & d_2^{(3)} - d_1^{(3)} = 2 \\ \alpha_3^{(3)} & = & d_3^{(3)} - d_2^{(3)} = 1 \end{array}$$

Hence we get

$$\begin{array}{rcl} n_3^{(3)} & = & \alpha_3^{(3)} = 1 \\ n_2^{(3)} & = & \alpha_2^{(3)} - \alpha_3^{(3)} = 1 \\ n_1^{(3)} & = & \alpha_1^{(3)} - \alpha_2^{(3)} = 1 \end{array}$$

Thus in $A^{(3)}$, there is one subblock of $3 \times 3$, (because $n_3^{(3)} = 1$), one sublocks of size $2 \times 2$, (because $n_2^{(3)} = 1$) and one subblock of size $1 \times 1$, (because $n_1^{(3)} = 1$). Thus we get,

$$A^{(3)} = \left(\begin{array}{ccc|cc|c} 6 & 1 & 0 & & & \\ 0 & 6 & 1 & & & \\ 0 & 0 & 6 & & & \\ \hline & & & 6 & 1 & \\ & & & 0 & 6 & \\ \hline & & & & & 0 \end{array}\right) \tag{6.4.14}$$

Thus the canonical form of $A$ is given by (6.4.11), where $A^{(1)}$,$A^{(2)}$ and $A^{(3)}$ are given by (6.4.12),(6.4.13) and (6.4.14).

## 6.5 Functions of a General Matrix

Just as we defined $f(A)$ for diagonalizable matrices, for any function $f(\lambda)$ analytic in a disc containng the eigenvalues of $A$, we can also define $f(A)$ for a function whose characteristic polynomial and minimal polynomial are as in (6.4.1) and (6.4.2). We first observe that, if $J_m(a)$ is a canonical Jordan matrix, then

$$(J_m(a))^2 = \begin{pmatrix} \mathrm{a}^2 & 2\mathrm{a} & 1 & & & & \\ & \mathrm{a}^2 & 2\mathrm{a} & 1 & & & \\ & & \mathrm{a}^2 & 2\mathrm{a} & 1 & & \\ & & & \ddots & \ddots & & \\ & & & & \mathrm{a}^2 & 2\mathrm{a} & 1 \\ & & & & & \mathrm{a}^2 & 2\mathrm{a} \\ & & & & & & \mathrm{a}^2 \end{pmatrix}_{m\times m}$$

By induction we can then show that, $(J_m(a))^r$ is the $m \times m$ matrix whose first row is

$$(a^r \; ra^{(r-1)} \; \frac{r(r-1)}{2!}a^{(r-2)} \; \frac{r(r-1)(r-2)}{3!}a^{(r-3)} \; \cdots \; rC_j a^{(r-j)} \cdots)$$

The next row is obtained by shifting the above to the right and so on. Thus the $i$th row of $(J_m(a))^r$ will have all entries upto the diagonal entry as zero and then start off with $a^r$ and proceed as above. We have the $i$th row of $(J_m(a))^r$ is give by

$$0 \; 0 \; \cdots \; \underbrace{a^r}_{i-th\ entry} \; rC_1 a^{(r-1)} \; rC_2 a^{(r-2)} \; \cdots$$

Using this we can show that if $p(\lambda)$ is any poynomial in $\mathcal{F}[\lambda]$ then $p(J_m(a))$ is the matrix whose $i$th row is of the form

$$0 \; 0 \; \cdots \; \underbrace{p(a)}_{i-th\ entry} \; rC_1 p'(a) \; rC_2 p''(a) \; \cdots$$

Finally, using the fact that any function $f(\lambda)$ analytic in a disc containing $a$ can be approximated by a sequence of polynomials, we see that we can define $f(J_m(a))$ as the $m \times m$ matrix whose $i$th row is of the form

$$0 \; 0 \; \cdots \; \underbrace{f(a)}_{i-th\ entry} \; rC_1 f'(a) \; rC_2 f''(a) \; \cdots$$

.

Since the Jordan canonical form of any matrix $A$ consists of diagonal subblocks of canonical Jordan matrices we define the Jordan canonical form of $f(A)$ by taking $f(subblock))$ for each of the subblocks, provided $f$ is analytic in a disc containing all the eigenvalues of $A$. We note that if $\mu_1, \mu_2, \cdots, \mu_n$ are the eigenvalues of $A$, written including their multiplicities, the eigenvalues of $f(A)$ are $f(\mu_1), f(\mu_2), \cdots, f(\mu_n)$.

**Example 6.5.1** For the matrix of Example 6.4.1 we had the eigenvalues $2, -4$ and $6$. Hence if $f$ is analytic in a disc $D_R$ of radius $R$ centred at the origin, where $R > 6$, then we can define $f(A)$ as follows:
We have

$$f(A^{(1)}) = \left(\begin{array}{lll|c|c} f(2) & f'(2) & f''(2) & & \\ 0 & f(2) & f'(2) & & \\ 0 & 0 & f(2) & & \\ \hline & & & f(2) & \\ \hline & & & & f(2) \end{array}\right) \tag{6.5.1}$$

$$f(A^{(2)}) = \left(\begin{array}{rr|c} f(-4) & f'(-4) & \\ 0 & f(-4) & \\ \hline & & f(-4) \end{array}\right) \tag{6.5.2}$$

$$f(A^{(3)}) = \left(\begin{array}{lll|rr|c} f(6) & f'(6) & f''(6) & & & \\ 0 & f(6) & f'(6) & & & \\ 0 & 0 & f(6) & & & \\ \hline & & & f(6) & f'(6) & \\ & & & 0 & f(6) & \\ \hline & & & & & f(6) \end{array}\right) \tag{6.5.3}$$

Then the canonical form of $f(A)$ is given by

$$f(A)_{can} = \left(\begin{array}{c|c|c} f(A^{(1)}) & & \\ \hline & f(A^{(2)}) & \\ \hline & & f(A^{(3)}) \end{array}\right) \tag{6.5.4}$$

where $f(A^{(1)})$,$f(A^{(2)})$ and $f(A^{(3)})$ are as in (6.5.1),(6.5.2) and (6.5.3).