# Jordan normal form

Branko Ćurgus

December 6, 2011 at 23:45

Throughout this note $\mathcal{V}$ is a finite dimensional vector space over $\mathbb{C}$. The symbol $\mathbb{N}$ denotes the set of positive integers and $i, j, k, l, m, n, p, q, r \in \mathbb{N}$.

## 1 Nilpotent operators

**Theorem 1.1.** *Let $\mathcal{V}$ be a nontrivial finite dimensional vector space over $\mathbb{C}$ with $n = \dim \mathcal{V}$. Let $N \in \mathcal{L}(\mathcal{V})$ be a nilpotent operator such that $m = \dim \mathcal{N}(N)$. Then there exist vectors $v_1, \dots, v_m \in \mathcal{V}$ and positive integers $q_1, \dots, q_m$ such that*

$$v_k \notin \mathcal{R}(N) \quad \textit{for all} \qquad k \in \{1, \dots, m\},$$

*the vectors*

$$N^{q_1-1}v_1, \dots, N^{q_m-1}v_m$$

*form a basis of $\in \mathcal{N}(N)$ and the vectors*

$$v_k, Nv_k, \dots, N^{q_k-1}v_k, \qquad k \in \{1, \dots, m\},$$

*form a basis of $\mathcal{V}$.*

*Proof.* First notice that if $N = 0$, then $\mathcal{N}(N) = \mathcal{V}$ and the theorem is trivially true. In this case $m = n$ and any basis $v_1, \dots, v_n$ of $\mathcal{V}$ with positive integers $q_1 = \cdots = q_n = 1$ satisfies the requirement of the theorem. From now on we assume that $N \neq 0$.

The proof is by induction on the dimension $n$. The statement is trivially true for $n = 1$. Let $n \in \mathbb{N}$ and assume that the statement is true for any vector space of dimension less or equal to $n$. It is always good to be specific and state what is being assumed. The following implication is our inductive hypothesis:

If $\mathcal{W}$ is a vector space over $\mathbb{C}$ such that $\dim \mathcal{W} \leq n$ and if $M \in \mathcal{L}(\mathcal{W})$ is a nilpotent operator such that $l = \dim \mathcal{N}(M)$, then there exist $w_1, \dots, w_l \in \mathcal{W}$ and positive integers $p_1, \dots, p_l$ such that

$$w_j \notin \mathcal{R}(M) \quad \text{for all} \qquad j \in \{1, \dots, l\},$$

the vectors

$$M^{p_1-1}w_1, \dots, M^{p_l-1}w_l$$

form a basis of $\mathcal{N}(M)$ and the vectors

$$w_j, Mw_j, \dots, M^{p_j-1}w_j, \qquad j \in \{1, \dots, l\},$$

form a basis of $\mathcal{W}$.

Next we present a proof of the inductive step.

Let $\mathcal{V}$ be a nontrivial finite dimensional vector space over $\mathbb{C}$ with $\dim \mathcal{V} = n+1$. Let $N \in \mathcal{L}(\mathcal{V})$ be a nilpotent operator such that $m = \dim \mathcal{N}(N)$. Set $\mathcal{W} = \mathcal{R}(N)$. Since $N$ is nilpotent it is not invertible. Thus $m = \dim \mathcal{N}(N) \geq 1$. By the famous "rank-nullity" theorem $\dim \mathcal{W} < n+1$. Since $N \neq 0$, $\dim \mathcal{W} > 0$. Clearly $N\mathcal{W} \subseteq \mathcal{W}$. Denote by $M$ the restriction $N|_{\mathcal{W}}$ of $N$ to $\mathcal{W}$. Then $M \in \mathcal{L}(\mathcal{W})$. Since $N$ is nilpotent, $M$ is nilpotent as well. Clearly, $\mathcal{N}(M) = \mathcal{N}(N) \cap \mathcal{R}(N)$. Set $l = \dim \mathcal{N}(M)$. The vector space $\mathcal{W}$ and the operator $M$ satisfy all the assumptions of the inductive hypothesis. This allows us to deduce that there exist $w_1, \ldots, w_l \in \mathcal{W}$ and positive integers $p_1, \ldots, p_l$ such that

$$w_j \notin \mathcal{R}(M) \quad \text{for all} \qquad j \in \{1, \ldots, l\}, \tag{1}$$

the vectors

$$M^{p_1-1}w_1, \ldots, M^{p_l-1}w_l \tag{2}$$

form a basis of $\mathcal{N}(M)$ the vectors

$$w_j, Mw_j, \ldots, M^{p_j-1}w_j, \qquad j \in \{1, \ldots, l\}, \tag{3}$$

form a basis of $\mathcal{W} = \mathcal{R}(N)$. Since $w_j \in \mathcal{R}(N)$, there exist $v_j \in \mathcal{V}$ such that $w_j = Nv_j$ for all $j \in \{1, \ldots, l\}$. Since by (1), $w_j \notin \mathcal{R}(M)$, we have $v_j \notin \mathcal{R}(N)$ for all $j \in \{1, \ldots, l\}$. We know that vectors in (2), that is,

$$M^{p_1-1}w_1 = N^{p_1}v_1, \ldots, M^{p_l-1}w_l = N^{p_l}v_l,$$

form a basis of $\mathcal{N}(M) = \mathcal{N}(N) \cap \mathcal{R}(N)$. Recall that $m = \dim \mathcal{N}(N)$, $l \leq m$, and let $v_{l+1}, \ldots, v_m$ be such that

$$N^{p_1}v_1, \ldots, N^{p_l}v_l, \; v_{l+1}, \ldots, v_m, \tag{4}$$

form a basis of $\mathcal{N}(N)$. (It is possible that $l = m$. In this case we already have a basis of $\mathcal{N}(N)$ and the last step can be skipped.)

Now let us review the stage: We started with the basis

$$w_j = Nv_j, \; Mw_j = N^2v_j, \ldots, M^{p_j-1}w_j = N^{p_j}v_j, \qquad j \in \{1, \ldots, l\},$$

of $\mathcal{W} = \mathcal{R}(N)$ which has exactly $\dim \mathcal{R}(N)$ vectors. Then we added the vectors $v_1, \ldots, v_m$. Now we have $m + \dim \mathcal{R}(N) = \dim \mathcal{N}(N) + \dim \mathcal{R}(N) = \dim \mathcal{V}$ vectors:

$$v_j, \; Nv_j, \; N^2v_j, \ldots, N^{p_j}v_j, \qquad j \in \{1, \ldots, l\}, \qquad v_{l+1}, \ldots, v_m. \tag{5}$$

For easier record keeping set

$$q_k = \begin{cases} p_k + 1 & \text{if} \quad k \in \{1, \ldots, l\} \\ 1 & \text{if} \quad k \in \{l+1, \ldots, m\}. \end{cases}$$

Then (5) can be rewritten as

$$v_k, \; Nv_k, \; N^2v_k, \ldots, N^{q_k-1}v_k, \qquad k \in \{1, \ldots, m\}. \tag{6}$$

Next we will prove that the vectors in (6) are linearly independent. Let

$$\alpha_{k,j} \in \mathbb{C}, \quad j \in \{0, \ldots, q_k - 1\}, \quad k \in \{1, \ldots, m\}$$

be such that

$$\sum_{k=1}^{m} \sum_{j=0}^{q_k-1} \alpha_{k,j} N^j v_k = 0. \tag{7}$$

Applying $N$ to the last equality yields

$$\sum_{k=1}^{l}\sum_{j=0}^{q_k-1}\alpha_{k,j}N^{j+1}v_k=\sum_{k=1}^{l}\sum_{j=0}^{q_k-2}\alpha_{k,j}N^{j+1}v_k=\sum_{k=1}^{l}\sum_{j=0}^{p_k-1}\alpha_{k,j}M^{j}w_k=0.$$

Since the vectors in the last double sum are linearly independent (they are the vectors from (3)) we have

$$\alpha_{k,0}=\cdots=\alpha_{k,q_k-2}=0,\qquad k\in\{1,\ldots,l\}.$$

Substituting these values in (7) we get

$$\sum_{k=1}^{m}\alpha_{k,q_k-1}N^{q_k-1}v_k=0.$$

But, beautifully, the vectors in the last sum are exactly the vectors in (4) which are linearly independent. Thus

$$\alpha_{k,q_k-1}=0,\qquad k\in\{1,\ldots,m\}.$$

This completes the proof that all the coefficients in (7) must be zero. Thus, the vectors in (6) are linearly independent. Since there are exactly $n+1$ vectors in (6) they form a basis of $\mathcal{V}$. This completes the proof. □

**Remark 1.2.** In this remark we will establish a connection between the lengths $q_1,\ldots,q_m$ and the numbers

$$m_j=\dim\mathcal{N}(N^j),\quad j\in\{1,\ldots,d\}.$$

Here $d\in\{1,\ldots,n\}$ is the degree of nilpotency of $N$, that is the smallest positive integer such that $T^d=0$. Then

$$0=m_0<m_1=m<m_2<\cdots<m_d=n=m_{d+1},$$

where, for convenience, we define $m_0=0$ and $m_{d+1}=n$. It follows from the previous theorem that

$$0<m_{i+1}-m_i\leq m_i-m_{i-1},\qquad i\in\{1,\ldots,d-1\}.$$

We can always assume that the lengths $q_1,\ldots,q_m\in\{1,\ldots,d\}$ from the previous theorem are in nonincreasing order. That is,

$$d=q_1\geq\cdots\geq q_m\geq 1.$$

Then the formula for $q_k$ is

$$q_k=\max\bigl\{j\in\{1,\ldots,d\}\,:\,m_j-m_{j-1}\geq k\bigr\},\qquad k\in\{1,\ldots,m\}.$$

Conversely, the numbers $m_1-m_0\geq m_2-m_1\geq\cdots\geq m_d-m_{d-1}\geq 1$ can be determined from $q_1,\ldots,q_m$ by

$$m_j-m_{j-1}=\max\bigl\{k\in\{1,\ldots,m\}\,:\,q_k\geq j\bigr\},\qquad j\in\{1,\ldots,d\}.$$

## 2 More about the upper triangular matrix representations

In class we proved the following theorem.

**Theorem 2.1.** *Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$. For $T \in \mathcal{L}(\mathcal{V})$ there exists a basis $\mathcal{B}$ for $\mathcal{V}$ such that the matrix $\mathsf{M}_{\mathcal{B}}(T)$ is upper triangular.*

Our next goal is to understand which complex numbers are on the diagonal of a triangular matrix $\mathsf{M}_{\mathcal{B}}(T)$.

**Theorem 2.2.** *Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$ and $n = \dim \mathcal{V}$. Let $T \in \mathcal{L}(\mathcal{V})$, let $\mathcal{B} = \{v_, \ldots, v_n\}$ be a basis for $\mathcal{V}$ such that the matrix $\mathsf{M}_{\mathcal{B}}(T)$ is upper triangular, that is*

$$\mathsf{M}_{\mathcal{B}}(T) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}. \tag{8}$$

*Then each eigenvalue $\lambda$ of $T$ appears among $\{a_{11}, a_{22}, \ldots, a_{nn}\}$ at least*

$$\dim \mathcal{N}\bigl((T - \lambda I)^n\bigr)$$

*times.*

*Proof.* We shall prove the theorem for $\lambda = 0$. The general case follows by considering the operator $T - \lambda I$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$. Assume that the set $\bigl\{i \in \{1, \ldots, n\} : a_{ii} \neq 0\bigr\}$ has exactly $r$ elements. Let

$$\bigl\{i \in \{1, \ldots, n\} : a_{ii} \neq 0\bigr\} = \bigl\{k_1, \ldots, k_r\bigr\}, \quad \text{where} \quad k_1 < \cdots < k_r.$$

In other words, the diagonal entries $a_{k_1 k_1}, \ldots, a_{k_r k_r}$ are nonzero and all other diagonal entries are 0. Since the matrix $\mathsf{M}_{\mathcal{B}}(T)$ is upper triangular, the vectors

$$\mathsf{C}_{\mathcal{B}}\bigl(Tv_{k_j}\bigr) = \begin{bmatrix} a_{1k_j} \\ \vdots \\ a_{k_j k_j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad j \in \{1, \ldots, r\},$$

are linearly independent. As the mapping $\mathsf{C}_{\mathcal{B}} : \mathcal{V} \to \mathbb{C}^n$ is an isomorphism, the vectors $Tv_{k_j}$, $j \in \{1, \ldots, r\}$, are linearly independent. Consequently, $\dim \mathcal{R}(T) \geq r$. Hence $\dim \mathcal{N}(T) = n - \dim \mathcal{R}(T) \leq n - r$. Since there are exactly $d - r$ zero entries on the diagonal of $\mathsf{M}_{\mathcal{B}}(T)$, we see that there are at least $\dim \mathcal{N}(T)$ zero entries on the diagonal of $\mathsf{M}_{\mathcal{B}}(T)$. Applying this result to the operator $T^n$ we conclude that there are at least $\dim \mathcal{N}(T^n)$ zero entries on the diagonal of $\mathsf{M}_{\mathcal{B}}(T^n)$. But, the diagonal entries of $\mathsf{M}_{\mathcal{B}}(T^n)$ are $a_{11}^n, \ldots, a_{nn}^n$ and the number of zeros among $a_{11}^n, \ldots, a_{nn}^n$ is identical to the number of zeros among $a_{11}, \ldots, a_{nn}$. Hence, there are at least $\dim \mathcal{N}(T^n)$ zero entries on the diagonal of $\mathsf{M}_{\mathcal{B}}(T)$. $\square$

**Theorem 2.3.** *Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$ and $n = \dim \mathcal{V}$. Let $T \in \mathcal{L}(\mathcal{V})$ and let $\mathcal{B} = \{v_, \ldots, v_n\}$ be a basis for $\mathcal{V}$ such that the matrix $\mathsf{M}_{\mathcal{B}}(T)$ is upper triangular with the elements on the main diagonal being $a_{11}, \ldots, a_{nn}$, see (8). Let*

$$p(z) = \bigl(z - a_{11}\bigr)\bigl(z - a_{22}\bigr) \cdots \bigl(z - a_{nn}\bigr). \tag{9}$$

*Then $p(T) = 0$.*

*Proof.* For $k \in \{1, 2, \ldots, n\}$, the matrix $\mathsf{M}_{\mathcal{B}}(T - a_{kk}I)$ is upper triangular and its entry in the $k$-th column and the $k$-th row is 0. Therefore,

$$(T - a_{11}I)\bigl(\operatorname{span}\{v_1\}\bigr) = \{0_{\mathcal{V}}\} \tag{10}$$

and, for $k \in \{2, \ldots, n\}$,

$$(T - a_{kk}I)\bigl(\operatorname{span}\{v_1, \ldots, v_k\}\bigr) \subseteq \operatorname{span}\{v_1, \ldots, v_{k-1}\}. \tag{11}$$

The inclusions (11) and (10) imply

$$\begin{aligned}
p(T)(\mathcal{V}) &= (T - a_{11}I)(T - a_{22}I) \cdots (T - a_{nn}I)(\mathcal{V}) \\
&= (T - a_{11}I) \cdots (T - a_{nn}I)\bigl(\operatorname{span}\{v_1, \ldots, v_n\}\bigr) \\
&\subseteq (T - a_{11}I) \cdots (T - a_{(n-1)(n-1)}I)\bigl(\operatorname{span}\{v_1, \ldots, v_{n-1}\}\bigr) \\
&\;\vdots \\
&\subseteq (T - a_{11}I)(T - a_{22}I)\bigl(\operatorname{span}\{v_1, v_2\}\bigr) \\
&\subseteq (T - a_{11}I)\bigl(\operatorname{span}\{v_1\}\bigr) \\
&= \{0_{\mathcal{V}}\}.
\end{aligned}$$

Thus $p(T) = 0$. The theorem is proved. □

## 3 A decomposition of a vector space

**Lemma 3.1.** *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$. Let $A$ and $B$ be linear mappings on $\mathcal{V}$. If $A$ and $B$ commute, then $\mathcal{N}(B)$ is an invariant subspace for $A$.*

*Proof.* Let $v$ be in $\mathcal{N}(B)$. Then $0_{\mathcal{V}} = Bv = ABv = BAv$. Therefore $Av$ belongs to $\mathcal{N}(B)$. □

**Lemma 3.2.** *Let $\mathcal{V}$ be a finite dimensional vector space over a field $\mathbb{F}$. Let $A$ and $B$ be linear operators on $\mathcal{V}$. Assume that $A$ and $B$ commute and that $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0_{\mathcal{V}}\}$. Then $\mathcal{N}(AB) = \mathcal{N}(A) \oplus \mathcal{N}(B)$.*

*Proof.* By Lemma 3.1 $\mathcal{N}(B)$ is an invariant subspace of $A$. Denote by $C$ the restriction of $A$ to $\mathcal{N}(B)$, that is $Cw = Aw$ for all $w$ in $\mathcal{N}(B)$. Then

$$\mathcal{N}(C) = \{w \in \mathcal{N}(B) : Cw = 0_{\mathcal{V}}\} = \mathcal{N}(A) \cap \mathcal{N}(B) = \{0_{\mathcal{V}}\}.$$

It follows that $C$ is a bijection of $\mathcal{N}(B)$ onto itself. Since $\mathcal{N}(B)$ is finite dimensional, $C$ is onto. Therefore, for every $v$ in $\mathcal{N}(B)$ there exists $u$ in $\mathcal{N}(B)$ such that $v = Cu = Au$. Let $w$ be arbitrary element of $\mathcal{N}(AB)$. Since $\mathcal{N}(AB) = \mathcal{N}(BA)$ we have $Aw \in \mathcal{N}(B)$. Hence, there exists $u$ in $\mathcal{N}(B)$ such that $Aw = Au$. Consequently, $w - u \in \mathcal{N}(A)$. Thus $w = (w - u) + u$, where $u \in \mathcal{N}(B)$ and $w - u \in \mathcal{N}(A)$. This proves that $\mathcal{N}(AB) \subseteq \mathcal{N}(A) \oplus \mathcal{N}(B)$. The converse inclusion is straightforward. □

**Proposition 3.3.** *Let $\mathcal{V}$ be a finite dimensional vector space over a field $\mathbb{F}$. Let $q \in \mathbb{N}, q > 1$, and let $A_1, A_2, \dots, A_q$, be linear operators on $\mathcal{V}$. Assume that*

$$A_j A_k = A_k A_j \qquad \text{and} \qquad \mathcal{N}(A_j) \cap \mathcal{N}(A_k) = \{0_{\mathcal{V}}\}, \quad j \neq k, \quad j,k \in \{1, \dots, q\}. \tag{12}$$

*Then*

$$\mathcal{N}\big(A_1 A_2 \cdots A_q\big) = \bigoplus_{j=1}^{q} \mathcal{N}(A_j).$$

*Proof.* The proof is by mathematical induction. We already proved the proposition for two operators. The inductive hypothesis is that the proposition is true for $q-1$ operators. To prove the inductive step assume (12). By the inductive hypothesis

$$\mathcal{N}\big(A_1 A_2 \cdots A_{q-1}\big) = \bigoplus_{j=1}^{q-1} \mathcal{N}(A_j). \tag{13}$$

Set $A = A_1 A_2 \cdots A_{q-1}$ and $B = A_q$. By repeated application of the first equality in (12) it follows that $AB = BA$. To apply Lemma 3.2 we need to verify $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0_{\mathcal{V}}\}$. By the inductive hypothesis the null space of $A$ is given by (13). Thus we need to prove

$$\left( \bigoplus_{j=1}^{q-1} \mathcal{N}(A_j) \right) \bigcap \mathcal{N}(A_q) = \{0_{\mathcal{V}}\}.$$

Let $v$ be in the above intersection. Then there exist $v_j \in \mathcal{N}(A_j), j \in \{1, \dots, q-1\}$ such that

$$v = v_1 + \cdots + v_{q-1} \qquad \text{and} \qquad A_q v = 0_{\mathcal{V}}.$$

The last two equalities imply

$$0_{\mathcal{V}} = A_q v_1 + \cdots + A_q v_{q-1}.$$

Since by (12) $A_j$ and $A_q$ commute, Lemma 3.1 implies that $\mathcal{N}(A_k)$ is invariant under $A_q$. That is, $A_q v_j \in \mathcal{N}(A_j)$ for all $j \in \{1, \dots, q-1\}$. This and the fact that the sum in (13) is direct yield $A_q v_j = 0_{\mathcal{V}}$ for all $j \in \{1, \dots, q-1\}$. By the second relation in (12) we get

$$v_j \in \mathcal{N}(A_j) \cap \mathcal{N}(A_k) = \{0_{\mathcal{V}}\} \quad \text{for all} \quad j \in \{1, \dots, q-1\}.$$

This proves that $v = 0_{\mathcal{V}}$. Now Lemma 3.2 yields $\mathcal{N}(AB) = \mathcal{N}(A) \oplus \mathcal{N}(B)$. Together with (13), this implies the claim of the proposition. □

**Proposition 3.4.** *Let $\mathcal{V}$ be a finite dimensional vector space over a field $\mathbb{F}$. Let $T \in \mathcal{L}(\mathcal{V})$. If $\lambda$ and $\mu$ are distinct eigenvalues of $T$ and $j$ and $k$ are natural numbers, then*

$$\mathcal{N}\big((T - \lambda I)^j\big) \bigcap \mathcal{N}\big((T - \mu I)^k\big) = \{0_{\mathcal{V}}\}.$$

*Proof.* The set equality in the proposition is equivalent to the implication

$$v \in \mathcal{N}\big((T - \mu I)^k\big) \setminus \{0_{\mathcal{V}}\} \quad \Rightarrow \quad v \notin \mathcal{N}\big((T - \lambda I)^j\big).$$

We will prove the last implication. Let $v \in \mathcal{V}$ be such that $(T - \mu I)^k v = 0_{\mathcal{V}}$ and $v \neq 0_{\mathcal{V}}$. Let $i \in \{1, \dots, k\}$ be such that $(T - \mu I)^i v = 0_{\mathcal{V}}$ and $(T - \mu I)^{i-1} v \neq 0_{\mathcal{V}}$. Set $w := (T - \mu I)^{i-1} v$. Then

$w$ is an eigenvector of $T$ corresponding to $\mu$: $Tw = \mu w$. Then (as we must have proven before), for an arbitrary polynomial $p \in \mathbb{C}[z]$ we have $p(T)w = p(\mu)w$. In particular

$$(T - \lambda I)^l w = (\mu - \lambda)^l w \quad \text{for all} \quad l \in \mathbb{N}.$$

Since $\mu - \lambda \neq 0$ and $w \neq 0_{\mathcal{V}}$ we have that

$$(T - \lambda I)^l w \neq 0_{\mathcal{V}} \quad \text{for all} \quad l \in \mathbb{N}.$$

Consequently,

$$(T - \lambda I)^l (T - \mu I)^{i-1} v \neq 0_{\mathcal{V}} \quad \text{for all} \quad l \in \mathbb{N}.$$

Since the operators $(T - \lambda I)^l$ and $(T - \mu I)^{i-1}$ commute we have

$$(T - \mu I)^{i-1} (T - \lambda I)^l v \neq 0_{\mathcal{V}} \quad \text{for all} \quad l \in \mathbb{N}.$$

Therefore $(T - \lambda I)^l v \neq 0_{\mathcal{V}}$ for all $l \in \mathbb{N}$. Hence $v \notin \mathcal{N}\big((T - \lambda I)^j\big)$. This proves the proposition. ☐

**Theorem 3.5.** *Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$ and $T \in \mathcal{L}(\mathcal{V})$. We make the following assumptions:*

(i) *$\mathcal{B}$ is a basis of $\mathcal{V}$ for which $\mathsf{M}_{\mathcal{B}}(T)$ is upper triangular.*

(ii) *$\lambda_1, \ldots, \lambda_q$, are all the distinct eigenvalues of $T$.*

(iii) *For $k \in \{1, \ldots, q\}$ denote by $m_k \in \{1, \ldots, n\}$ the number of times the eigenvalue $\lambda_k$ appears on the diagonal of $\mathsf{M}_{\mathcal{B}}(T)$.*

(iv) *For $k \in \{1, \ldots, q\}$ set $\mathcal{W}_k := \mathcal{N}\big((T - \lambda_k I)^{m_k}\big)$.*

*Then*

(a) *Each of the subspaces $\mathcal{W}_1, \ldots, \mathcal{W}_q$, is invariant subspace of $T$.*

(b) *$\mathcal{V} = \mathcal{W}_1 \oplus \cdots \oplus \mathcal{W}_q$.*

(c) *$\dim \mathcal{W}_k = m_k$ and $\mathcal{W}_k = \mathcal{N}\big((T - \lambda_k)^n\big)$ for all $k \in \{1, \ldots, q\}$.*

(d) *For $k \in \{1, \ldots, q\}$ set $T_k = T|_{\mathcal{W}_k}$ and $N_k = T_k - \lambda_k I$. Then $N_k^{m_k} = 0$, that is, $N_k$ is a nilpotent mapping on $\mathcal{W}_k$.*

*Proof.* (a) Since the mapping $T$ commutes with each of the mappings $(T - \lambda_k I)^n$, Lemma 3.1 implies that each subspace $\mathcal{W}_1, \ldots, \mathcal{W}_q$, is an invariant subspace of $T$.

(b) By Theorem 2.3 we have

$$p(T) = (T - \lambda_1 I)^{m_1} \cdots (T - \lambda_q I)^{m_q} = 0.$$

Notice that the mappings $(T - \lambda_1 I)^{m_1}, \ldots, (T - \lambda_q I)^{m_q}$ satisfy the assumptions of Proposition 3.3. Consequently, $\mathcal{V} = \mathcal{N}\left(p(T)\right)$ is the direct sum of the subspaces $\mathcal{W}_1, \ldots \mathcal{W}_q$. This proves (b).

(c) Since clearly $m_k \leq n$, we have that

$$\mathcal{W}_k \subseteq \mathcal{N}\big((T - \lambda_k)^n\big). \tag{14}$$

By Theorem 2.2, $\dim \mathcal{N}\big((T-\lambda_k)^n\big) \le m_k$, and hence

$$\dim \mathcal{W}_k \le \dim \mathcal{N}\big((T-\lambda_k)^n\big) \le m_k. \tag{15}$$

Since

$$n \;=\; \sum_{k=1}^{q} \dim \mathcal{W}_k \;\le\; \sum_{k=1}^{q} m_k \;=\; n,$$

the inequalities in (15) are in fact equalities. That is

$$\dim \mathcal{W}_k = \dim \mathcal{N}\big((T-\lambda_k)^n\big) = m_k. \tag{16}$$

This and (14) imply $\mathcal{W}_k = \mathcal{N}\big((T-\lambda_k)^n\big)$.

(d) Clearly, $\mathcal{W}_k$ is also an invariant subspace of $T-\lambda_k I$. Denote by $N_k$ the restriction of $T-\lambda_k I$ to its invariant subspace $\mathcal{W}_k$ and by $T_k$ the restriction of $T$ to $\mathcal{W}_k$. Then, $T_k = \lambda_k I + N_k$ and the mapping $N_k$ is nilpotent. □

**Definition 3.6.** Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$ and $T \in \mathcal{L}(\mathcal{V})$. Let $1 \le q \le n$ and let $\lambda_1, \ldots, \lambda_q$ be all the distinct eigenvalues of $T$. Set

$$n_k = \dim \mathcal{N}\big((T-\lambda_k)^n\big), \qquad k \in \{1, \ldots, q\}.$$

The number $n_k$ is called the *algebraic multiplicity* of the eigenvalue $\lambda_k$. The polynomial

$$p(z) = \big(z-\lambda_1\big)^{n_1} \cdots \big(z-\lambda_q\big)^{n_q} \tag{17}$$

is called the *characteristic polynomial* of $T$.

**Theorem 3.7** (Hamilton-Cayley)**.** *Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$ and $T \in \mathcal{L}(\mathcal{V})$. Let $p$ be a characteristic polynomial of $T$. Then $p(T) = 0$.*

*Proof.* We use the notation of Theorem 3.5 and Definition 3.6. By Theorem 3.5 (c) we have $n_k = m_k$ for all $k = 1, \ldots, q$. Therefore the polynomials defined in (9) and (17) are identical. Now the theorem follows from Theorem 2.3. □

# 4 The Jordan Normal Form

Let $T$ be a linear operator on a vector space $\mathcal{V}$ over $\mathbb{C}$. Let $\lambda$ be an eigenvalue of $T$ and $l \in \mathbb{N}$. A sequence of nonzero vectors

$$v_1, \ldots, v_l \tag{18}$$

such that

$$Tv_1 = \lambda v_1, \quad \text{and} \quad v_l \notin \mathcal{R}(T-\lambda I), \tag{19}$$

and, if $l > 1$,

$$Tv_j = \lambda v_j + v_{j-1}, \quad j \in \{2, \ldots, l\} \tag{20}$$

is called a *Jordan chain of $T$ corresponding to the eigenvalue $\lambda$*. The number $l$ is the *length of the Jordan chain.* The vector $v_l$ is called the *lead vector of the Jordan chain.*

The lead vector of a Jordan chain satisfies

$$v_l \notin \mathcal{R}(T-\lambda I)$$

and all the other vectors of the corresponding Jordan chain can be expressed in terms of the lead vector:

$$v_{l-j} = (T - \lambda I)^j v_l, \qquad j \in \{0, 1, \ldots, l-1\}.$$

Notice that $(T - \lambda I)^l v_l = 0_{\mathcal{V}}$ since $v_1$ is an eigenvector of $T$.

A sequence

$$(T - \lambda I)^{l-1} v,\ (T - \lambda I)^{l-2} v,\ (T - \lambda I) v,\ \ldots,\ v, \tag{21}$$

is a Jordan chain, provided that $(T - \lambda I)^{l-1} v \neq 0_{\mathcal{V}}$, $(T - \lambda I)^l v = 0_{\mathcal{V}}$ and $v \notin \mathcal{R}(T - \lambda I)$.

Let $\mathcal{W}$ be a subspace of $\mathcal{V}$ spanned by a Jordan chain (18) of $T$. The first equality in (19) and (20) imply that $\mathcal{W}$ is an invariant subspace of $T$. If we denote by $S$ the restriction of $T$ to $\mathcal{W}$, then the matrix representation of $S$ with respect to the basis $\{v_1, \ldots, v_l\}$ is

$$\mathsf{M}_{\mathcal{B}}(S) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}. \tag{22}$$

A matrix of this form is called a *Jordan block* corresponding to the eigenvalue $\lambda$. In words: a Jordan block corresponding to the eigenvalue $\lambda$ is a square matrix with all elements on the main diagonal equal to $\lambda$ and all elements on the superdiagonal equal to 1.

A basis for $\mathcal{V}$ which consists of Jordan chains of $T$ is called a *Jordan basis* for $\mathcal{V}$ with respect to $T$.

If a basis $\mathcal{B}$ for $\mathcal{V}$ is a Jordan basis with respect to $T$ then the matrix $\mathsf{M}_{\mathcal{B}}(T)$ has Jordan blocks of different sizes on the diagonal and all other elements of $\mathsf{M}_{\mathcal{B}}(T)$ are zeros. Each eigenvalue of $T$ is represented in $\mathsf{M}_{\mathcal{B}}(T)$ by one or more Jordan blocks:

$$\begin{bmatrix} \boxed{\begin{matrix} \lambda_1 & 1 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & \lambda_1 \end{matrix}} & \begin{matrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{matrix} & \\ \begin{matrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{matrix} & \boxed{\begin{matrix} \lambda_2 & 1 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & \lambda_2 \end{matrix}} & \\ & & \ddots \\ & & & \ddots \\ & & & & \ddots \end{bmatrix}. \tag{23}$$

In the above matrix $\lambda_1$ and $\lambda_2$ are not necessarily distinct eigenvalues. A matrix of the form (23) is called the *Jordan normal form* for $T$. More precisely, a square matrix $\mathsf{M} = \left[a_{j,k}\right]$ is a *Jordan normal form* for $T$ if:

(i) all elements of $\mathsf{M}$ outside of the main diagonal and the superdiagonal are 0,

(ii) all elements on the main diagonal of $\mathsf{M}$ are eigenvalues of $T$,

(iii) all elements on the superdiagonal of $\mathsf{M}$ are either 1 or 0, and,

(iv) if $a_{j-1,j-1} \neq a_{j,j}$, with $j \in \{2, \dots, n\}$, then $a_{j-1,j} = 0$.

**Theorem 4.1.** *Let $\mathcal{V}$ be a finite dimensional vector space over $\mathbb{C}$ and let $T \in \mathcal{L}(\mathcal{V})$. Then $\mathcal{V}$ has a Jordan basis with respect to $T$.*

*Proof.* We use the notation and the results of Theorem 3.5. Let $k \in \{1, \dots, q\}$. It is important to notice that each Jordan chain of the nilpotent operator $N_k$ is a Jordan chain of $T$ which corresponds to the eigenvalue $\lambda_k$. Since $N_k$ is a nilpotent mapping in $\mathcal{L}(\mathcal{W}_k)$, by Theorem 1.1 there exists a basis $\mathcal{B}_k = \{v_{k,1}, \dots, v_{k,m_k}\}$ for $\mathcal{W}_k$ which consists of Jordan chains of $N_k$. Consequently, $\mathcal{B}_k$ consists of Jordan chains of $T$. Since $\mathcal{V}$ is a direct sum of $\mathcal{W}_1, \dots, \mathcal{W}_q$, the union $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_q$, that is,

$$\mathcal{B} = \big\{v_{1,1}, \dots, v_{1,m_1}, v_{2,1}, \dots, v_{2,m_2}, \dots, v_{q,1}, \dots, v_{q,m_q}\big\}$$

is a basis for $\mathcal{V}$. This basis consists of Jordan chains of $T$.

The matrix $\mathsf{M}_{\mathcal{B}}(T)$ is a block diagonal with the blocks $\mathsf{M}_{\mathcal{B}_k}(T_k)$, $k = 1, \dots, q$, on the diagonal and with zeros every where else:

$$\mathsf{M}_{\mathcal{B}}(T) = \begin{bmatrix} \mathsf{M}_{\mathcal{B}_1}(T_1) & 0 & \cdots & 0 \\ 0 & \mathsf{M}_{\mathcal{B}_2}(T_2) & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & \mathsf{M}_{\mathcal{B}_q}(T_q) \end{bmatrix}.$$

Since $T_k = \lambda_k I + N_k$, we have

$$\mathsf{M}_{\mathcal{B}_k}(T_k) = \lambda_k \mathsf{I} + \mathsf{M}_{\mathcal{B}_k}(N_k).$$

Thus all the elements on the main diagonal of $\mathsf{M}_{\mathcal{B}_k}(T_k)$ equal $\lambda_k$ and all the elements of superdiagonal of $\mathsf{M}_{\mathcal{B}_k}(T_k)$ are either 1 or 0. If there are exactly $h_k$ Jordan chains in the basis $\mathcal{B}_k$, then 0 appears exactly $h_k - 1$ times on the superdiagonal of $\mathsf{M}_{\mathcal{B}_k}(T_k)$. Therefore $\mathsf{M}_{\mathcal{B}}(T)$ is a Jordan normal form for $T$. $\square$