Jordan normal form

Branko Ćurgus

March 13, 2020 at 00:41

Throughout this note $\mathscr{V}$ is a nontrivial finite dimensional vector space over $\mathbb{C}$. We set $n = \dim \mathscr{V}$. The symbol $\mathbb{N}$ denotes the set of positive integers and $i, j, k, l, m, n, p, q, r \in \mathbb{N}$. For $T \in \mathscr{L}(\mathscr{V})$ by $\operatorname{nul}(T)$ we denote the null-space and by $\operatorname{ran}(T)$ the range of T.

1 Nilpotent operators

An operator $N \in \mathscr{L}(\mathscr{V})$ is *nilpotent* if there exists $q \in \mathbb{N}$ such that $N^q = 0$. If $N^q = 0$ and $N^{q-1} \neq 0$, then q is called *the degree of nilpotency* of N.

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

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

form a basis of $\operatorname{nul}(N)$ *and the vectors*

$$N^{q_k-1} v_k, N^{q_k-2} v_k, \ldots, N^2 v_k, N v_k, v_k, \qquad k \in \{1, \ldots, m\},$$

form a basis of $\mathscr{V}$.

Proof. The proof is by induction on the dimension n. Since in one-dimensional vector space each linear operator is a multiplication by a fixed scalar, the only nilpotent operator for $n = 1$ is the zero operator. So the statement is trivially true for $n = 1$.

Let $n \in \mathbb{N}$ be such that $n > 1$ and assume that the statement is true for any vector space of dimension less than n. It is always a good idea to be specific and state what is being assumed. Let $n \in \mathbb{N}$ be such that $n > 1$. The following implication is our inductive hypothesis:

If $\mathscr{W}$ is a vector space over $\mathbb{C}$ such that $\dim \mathscr{W} < n$ and if $M \in \mathscr{L}(\mathscr{W})$ is a nilpotent operator such that $l = \dim \operatorname{nul}(M)$, then there exist $w_1, \ldots, w_l \in \mathscr{W}$ and positive integers $p_1, \ldots, p_l$ such that the vectors

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

form a basis of $\operatorname{nul}(M)$ and the vectors

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

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

Next we present a proof of the inductive step.

Let $\mathscr{V}$ be a nontrivial finite dimensional vector space over $\mathbb{C}$ with $\dim \mathscr{V} = n$. Let $N \in \mathscr{L}(\mathscr{V})$ be a nilpotent operator.

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

Set $m = \dim \operatorname{nul}(N)$ and $\mathscr{W} = \operatorname{ran}(N)$. Since all powers of an invertible operator are invertible and a power of N is 0, N it is not invertible. Thus $m = \dim \operatorname{nul}(N) \geq 1$. By the famous "nullity-rank" theorem $\dim \mathscr{W} < n$. Since $N \neq 0$, $\dim \mathscr{W} > 0$. It is clear that $\mathscr{W}$ is invariant under N. Set M to be the restriction of N onto $\mathscr{W}$. Then $M \in \mathscr{L}(\mathscr{W})$. Since N is nilpotent, M is nilpotent as well. Clearly, $\operatorname{nul}(M) = \operatorname{nul}(N) \cap \operatorname{ran}(N)$. Set $l = \dim \operatorname{nul}(M)$. The vector space $\mathscr{W}$ and the operator M satisfy all the assumptions of the inductive hypothesis. This allows us to deduce that there exist $w_1, \ldots, w_l \in \mathscr{W}$ and positive integers $p_1, \ldots, p_l$ such that the vectors

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

form a basis of $\operatorname{nul}(M) = \operatorname{nul}(N) \cap \operatorname{ran}(N)$ and the vectors

$$M^{p_j - 1} w_j, \ldots, M w_j, w_j, \qquad j \in \{1, \ldots, l\}, \tag{2}$$

form a basis of $\mathscr{W} = \operatorname{ran}(N)$. Since $w_j \in \operatorname{ran}(N)$, there exist $v_j \in \mathscr{V}$ such that $w_j = N v_j$ for all $j \in \{1, \ldots, l\}$. We know from (1) that the vectors

$$M^{p_1 - 1} w_j = N^{p_j} v_j, \qquad j \in \{1 \ldots, l\}$$

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

$$N^{p_1} v_1, \ldots, N^{p_l} v_l,\ v_{l+1}, \ldots, v_m, \tag{3}$$

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

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

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

of $\mathscr{W} = \operatorname{ran}(N)$ with $\dim \operatorname{ran}(N)$ vectors. To this basis we added the vectors $v_1, \ldots, v_m$ where $m = \dim \operatorname{nul}(N)$. Now we have

$$m + \dim \operatorname{ran}(N) = \dim \operatorname{nul}(N) + \dim \operatorname{ran}(N) = \dim \mathscr{V} = n \tag{4}$$

vectors:

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

For easier writing 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

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

Next we will prove that the vectors in (6) are linearly independent. Let $\alpha_{k,j}\in\mathbb{C}, j\in\{0,\ldots,q_k-1\},\ k\in\{1,\ldots,m\}$ 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

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

Since the vectors in the last double sum are exactly the vectors from (2) which are linearly independent, we conclude that

$$\alpha_{k,0}=\cdots=\alpha_{k,q_k-2}=0\quad\text{for all}\quad k\in\{1,\ldots,l\}.$$

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 (3) which are linearly independent. Thus

$$\alpha_{k,q_k-1}=0\quad\text{for all}\quad k\in\{1,\ldots,m\}.$$

This completes the proof that all the coefficients in (7) must be zero. Thus, the vectors in (6) are linearly independent. Since by (4) there are exactly n vectors in (6) these vectors do form a basis of $\mathscr{V}$. This completes the proof. □

2 A Decomposition of a Vector Space

Lemma 2.1. *Let $\mathscr{V}$ be a vector space over $\mathbb{F}$. Let A and B be commuting linear operators on $\mathscr{V}$. Then $\operatorname{nul}(B)$ and $\operatorname{ran}(B)$ are invariant subspaces for A.*

Proof. This is a very simple exercise. □

Proposition 2.2. *Let $\mathscr{V}$ be a vector space over $\mathbb{F}$. Let $T\in\mathscr{L}(\mathscr{V})$. If λ and μ are distinct eigenvalues of T and j and k are positive integers, then*

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

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

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

We will prove this implication. Let $v \in \mathscr{V}$ be such that $(T-\mu I)^k v = 0_{\mathscr{V}}$ and $v \neq 0_{\mathscr{V}}$. Let $i \in \{1, \ldots, k\}$ be such that $(T-\mu I)^i v = 0_{\mathscr{V}}$ and $(T-\mu I)^{i-1} v \neq 0_{\mathscr{V}}$. Set $w = (T-\mu I)^{i-1} v$. Then w is an eigenvector of T corresponding to μ, that is $Tw = \mu w$ and $w \neq 0$. 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 $w \neq 0_{\mathscr{V}}$ we have that

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

Consequently,

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

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

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

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

Corollary 2.3. *Let $\mathscr{V}$ be a finite dimensional vector space over $\mathbb{F}$. Let $T \in \mathscr{L}(\mathscr{V})$. If λ and μ are distinct eigenvalues of T and j and k are positive integers, then*

$$\operatorname{nul}\big((T-\lambda I)^j\big) \subseteq \operatorname{ran}\big((T-\mu I)^k\big).$$

Proof. Since the operators $(T-\lambda I)^j$ and $(T-\mu I)^k$ commute, by Lemma 2.1, $\operatorname{nul}\big((T-\lambda I)^j\big)$ is invariant under $(T-\mu I)^k$. Denote by S the restriction of $(T-\mu I)^k$ onto $\operatorname{nul}\big((T-\lambda I)^j\big)$. Since clearly,

$$\operatorname{nul}(S) = \operatorname{nul}\big((T-\lambda I)^j\big) \cap \operatorname{nul}\big((T-\mu I)^k\big).$$

Proposition 2.2 implies that S is an injection, and thus bijection. Hence,

$$S\Big(\operatorname{nul}\big((T-\lambda I)^j\big)\Big) = \operatorname{nul}\big((T-\lambda I)^j\big)$$

and consequently

$$\operatorname{nul}\big((T-\lambda I)^j\big) = (T-\mu I)^k \Big(\operatorname{nul}\big((T-\lambda I)^j\big)\Big) \subseteq \operatorname{ran}\big((T-\mu I)^k\big).$$

□

Lemma 2.4. *Let $\mathscr{V}$ be a vector space over $\mathbb{F}$. Let $\mathscr{U}$ and $\mathscr{W}$ be subspaces of $\mathscr{V}$ such that*

$$\mathscr{V} = \mathscr{U} \oplus \mathscr{W}.$$

Let $S \in \mathscr{L}(\mathscr{V})$ be such that $S\mathscr{U} \subseteq \mathscr{U}$ and $S\mathscr{W} \subseteq \mathscr{W}$. If $\operatorname{nul}(S) \cap \mathscr{W} = \{0\}$, then

$$\operatorname{nul}\big((S|_{\mathscr{U}})^j\big) = \operatorname{nul}\big(S^j\big) \qquad \text{for all} \qquad j \in \mathbb{N}. \tag{8}$$

Proof. Assume $\operatorname{nul}(S) \cap \mathscr{W} = \{0\}$. We first prove the equality for $j = 1$. Since $\operatorname{nul}\big(S|_{\mathscr{U}}\big) = \operatorname{nul}\big(S\big) \cap \mathscr{U}$, the inclusion $\operatorname{nul}\big(S|_{\mathscr{U}}\big) \subseteq \operatorname{nul}\big(S\big)$ is clear. Let $v \in \operatorname{nul}(S)$ be arbitrary. Then $v = u + w$ with $u \in \mathscr{U}$ and $w \in \mathscr{W}$. Applying S to this identity we get $0 = Sv = Su + Sw$. Since $Su \in \mathscr{U}$ and $Sw \in \mathscr{W}$, the assumption that the sum of $\mathscr{U}$ and $\mathscr{W}$ is direct yields $Sw = 0$. Since $\operatorname{nul}(S) \cap \mathscr{W} = \{0\}$, we have $w = 0$. Thus, $v \in \mathscr{U}$, and hence $v \in \operatorname{nul}\big(S|_{\mathscr{U}}\big)$.

To prove (8) for arbitrary $j \in \mathbb{N}$ we will first prove that

$$\operatorname{nul}\big(S^j\big) \cap \mathscr{W} = \{0\} \qquad \text{for all} \qquad j \in \mathbb{N}. \tag{9}$$

A simple proof proceeds by mathematical induction. The statement in (9) is true for $j = 1$. Let $j \in \mathbb{N}$ and assume that the statement in (9) is true for j. Now assume that $w \in \mathscr{W}$ and $S^{j+1}w = 0$. Then $Sw \in \mathscr{W}$ and $S^j(Sw) = 0$. By the inductive hypothesis, that is $\operatorname{nul}\big(S^j\big) \cap \mathscr{W} = \{0\}$ we conclude $Sw = 0$. Since $\operatorname{nul}(S) \cap \mathscr{W} = \{0\}$, we deduce that $w = 0$.

Having (9), we can apply the equality proved in the first part of the proof to the operator S^j. ☐

Corollary 2.5. *Let $\mathscr{V}$ be a finite dimensional vector space over $\mathbb{F}$. Let $\mathscr{U}$ and $\mathscr{W}$ be subspaces of $\mathscr{V}$ such that*

$$\mathscr{V} = \mathscr{U} \oplus \mathscr{W}.$$

Let $T \in \mathscr{L}(V)$ be such that $T\mathscr{U} \subseteq \mathscr{U}$ and $T\mathscr{W} \subseteq \mathscr{W}$. Then

$$\sigma\big(T|_{\mathscr{U}}\big) \cup \sigma\big(T|_{\mathscr{W}}\big) = \sigma(T). \tag{10}$$

If $\lambda \in \sigma\big(T\big)$ and $\lambda \notin \sigma\big(T|_{\mathscr{W}}\big)$, then $\lambda \in \sigma\big(T|_{\mathscr{U}}\big)$ and

$$\operatorname{nul}\big((T|_{\mathscr{U}} - \lambda I)^j\big) = \operatorname{nul}\big((T - \lambda I)^j\big) \qquad \textit{for all} \qquad j \in \mathbb{N}. \tag{11}$$

Proof. The inclusion $\subseteq$ in (10) is clear. To prove $\supseteq$, let $\lambda \in \sigma(T)$ and let $v \neq 0$ be such that $Tv = \lambda v$. Let $v = u + w$, with $u \in \mathscr{U}$ and $w \in \mathscr{W}$. Since $v \neq 0$ we have $u \neq 0$ or $w \neq 0$. Applying T to both sides of $v = u+w$ and using the fact that v is an eigenvalue corresponding to λ we get $Tu + Tw = Tv = \lambda v = \lambda u + \lambda w$. Consequently, $(Tu - \lambda u) + (Tw - \lambda w) = 0$. Since the sum $\mathscr{V} = \mathscr{U} \oplus \mathscr{W}$ is direct and $Tu - \lambda u \in \mathscr{U}$ and $Tw - \lambda w \in \mathscr{W}$ we conclude $Tw - \lambda w = 0$ and $Tu - \lambda u = 0$. Since $u \neq 0$ or $w \neq 0$, we have $\lambda \in \sigma\big(T|_{\mathscr{U}}\big)$ or $\lambda \in \sigma\big(T|_{\mathscr{W}}\big)$.

Assume $\lambda \in \sigma\big(T\big)$ and $\lambda \notin \sigma\big(T|_{\mathscr{W}}\big)$. Then $\operatorname{nul}(T - \lambda I) \cap \mathscr{W} = \{0\}$. Lemma 2.4 applies to the operator $T - \lambda I$ and yields (11). Since $\lambda \in \sigma\big(T\big)$, $\operatorname{nul}(T - \lambda I) \neq \{0\}$. Now (11) with $j = 1$ yields $\lambda \in \sigma\big(T|_{\mathscr{U}}\big)$. ☐

Theorem 2.6. *Let $\mathscr{V}$ be a finite dimensional vector space over $\mathbb{C}$, $n = \dim \mathscr{V}$ and let $T \in \mathscr{L}(\mathscr{V})$. Let $\lambda_1, \ldots, \lambda_k$, be all the distinct eigenvalues of T. Set*

$$\mathscr{W}_j = \operatorname{nul}\big((T - \lambda_j I)^n\big) \quad \textit{and} \quad n_j = \dim \mathscr{W}_j, \quad j \in \{1, \ldots, k\}.$$

Then

(a) *Each of the subspaces $\mathscr{W}_1, \ldots, \mathscr{W}_k$, is invariant under T.*
(b) *$\mathscr{V} = \mathscr{W}_1 \oplus \cdots \oplus \mathscr{W}_k$.*
(c) *Set $T_j = T|_{\mathscr{W}_j}$ and $N_j = T_j - \lambda_j I$, $j \in \{1, \ldots, k\}$. Then $N_j^{n_j} = 0$, that is, N_j is a nilpotent operator on $\mathscr{W}_j$.*

Proof. (a) Since T commutes with each of the operators $(T-\lambda_j I)^d$, $j \in \{1,\dots,k\}$ Lemma 2.1 implies that each subspace $\mathscr{W}_1,\dots,\mathscr{W}_k$, is an invariant subspace of T.

To prove (b) we proceed by mathematical induction on the number k of distinct eigenvalues of T. We first prove the base step. Assume that λ is the only eigenvalue of T. Let $\mathscr{B} = \{v_1,\dots,v_n\}$ be a basis of $\mathscr{V}$ such that the matrix $\mathsf{M}_{\mathscr{B}}^{\mathscr{B}}(T)$ is upper triangular. Then, as we proved earlier all the diagonal entries of $\mathsf{M}_{\mathscr{B}}^{\mathscr{B}}(T)$ equal to λ. From the definition of $\mathsf{M}_{\mathscr{B}}^{\mathscr{B}}(T)$ it follows that

$$(T-\lambda I)\bigl(\operatorname{span}\{v_1,\dots,v_j\}\bigr) \subseteq \bigl(\operatorname{span}\{v_1,\dots,v_j-1\}\bigr) \quad \text{for all} \quad j \in \{2,\dots,n\}.$$

Therefore

$$\begin{aligned}
(T-\lambda I)^n(\mathscr{V}) &= (T-\lambda I)^{n-1}(T-\lambda I)\bigl(\operatorname{span}\{v_1,\dots,v_n\}\bigr)\\
&\subseteq (T-\lambda I)^{n-1}\bigl(\operatorname{span}\{v_1,\dots,v_{n-1}\}\bigr)\\
&\ \vdots\\
&\subseteq (T-\lambda I)(T-\lambda I)\bigl(\operatorname{span}\{v_1,v_2\}\bigr)\\
&\subseteq (T-\lambda I)\bigl(\operatorname{span}\{v_1\}\bigr)\\
&= \{0_{\mathscr{V}}\}.
\end{aligned}$$

Thus $\mathscr{V} = \operatorname{nul}\bigl((T-\lambda I)^n\bigr)$. This completes the proof of the base case.

Now we prove the inductive step. Let $k \in \mathbb{N}$ and assume that the statement is true for an operator with k distinct eigenvalues. Let T be an operator with $k+1$ distinct eigenvalues $\lambda_1,\dots,\lambda_k,\lambda_{k+1}$. For convenience we set $\lambda_{k+1} = \lambda$. Then, by assumption $\lambda \neq \lambda_j$ for all $j \in \{1,\dots,k\}$. We set

$$\mathscr{U} = \operatorname{ran}\bigl((T-\lambda I)^n\bigr) \qquad \text{and} \qquad \mathscr{W} = \operatorname{nul}\bigl((T-\lambda I)^n\bigr).$$

Since T and $(T-\lambda I)^n$ commute, Lemma 2.1 implies that both $\mathscr{U}$ and $\mathscr{W}$ are invariant under T.

Next we prove that

$$\operatorname{ran}\bigl((T-\lambda I)^n\bigr) \cap \operatorname{nul}\bigl((T-\lambda I)^n\bigr) = \mathscr{U} \cap \mathscr{W} = \{0\}. \tag{12}$$

(Prove this as an exercise.)

By the Nullity-Rank theorem

$$\mathscr{V} = \mathscr{U} \oplus \mathscr{W}. \tag{13}$$

(Provide details as an exercise.)

By Corollary 2.3

$$\operatorname{nul}\bigl((T-\lambda_j I)^n\bigr) \subseteq \mathscr{U} \quad \text{for all} \quad j \in \{1,\dots,k\}. \tag{14}$$

Let $m = \dim \mathscr{U}$. Denote by S the restriction of T onto $\mathscr{U}$. The inclusion in (14) implies that $\lambda_1,\dots,\lambda_k$ are eigenvalues of S. Similarly, (12) implies that λ is not an eigenvalue of S. Now Corollary 2.5 yields

$$\sigma(S) = \{\lambda_1,\dots,\lambda_k\}.$$

The second claim of Corollary 2.5 implies

$$\operatorname{nul}\big((T-\lambda_j I)^n\big) = \operatorname{nul}\big((S-\lambda_j I)^n\big).$$

Since $n > m = \dim \mathscr{U}$ we have

$$\operatorname{nul}\big((S-\lambda_j I)^m\big) = \operatorname{nul}\big((S-\lambda_j I)^{m+1}\big) = \cdots = \operatorname{nul}\big((S-\lambda_j I)^n\big).$$

Therefore,

$$\operatorname{nul}\big((T-\lambda_j I)^n\big) = \operatorname{nul}\big((S-\lambda_j I)^m\big). \tag{15}$$

The inductive hypothesis applies to S. Therefore

$$\mathscr{U} = \operatorname{ran}\big((T-\lambda I)^n\big) = \bigoplus_{j=1}^{k} \operatorname{nul}\big((S-\lambda_j I)^m\big). \tag{16}$$

Now (16), (15), and (13) yield

$$\mathscr{V} = \bigoplus_{j=1}^{k+1} \operatorname{nul}\big((T-\lambda_j I)^n\big).$$

Now we prove (c). Lemma 2.1 implies that $\mathscr{W}_j$ is an invariant subspace of $T-\lambda_j I$. Denote by N_j the restriction of $T-\lambda_j I$ to its invariant subspace $\mathscr{W}_j$ and by T_j the restriction of T to $\mathscr{W}_j$. Then, $T_j = \lambda_j I + N_j$ and the operator N_j is nilpotent. $\square$

Definition 2.7. Let $k \in \{1,\dots,n\}$ be such that $\lambda_1,\dots,\lambda_k$ are all the distinct eigenvalues of T. Set

$$n_j = \dim \operatorname{nul}\big((T-\lambda_j)^n\big), \qquad j \in \{1,\dots,k\}.$$

The number n_j is called the *algebraic multiplicity* of the eigenvalue λ_j. The polynomial

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

is called the *characteristic polynomial* of T.

3 The Jordan Normal Form

Let T be an operator on a vector space $\mathscr{V}$ over $\mathbb{C}$. Let λ be an eigenvalue of T and let v be such that $(T-\lambda I)^l v = 0_{\mathscr{V}}$ and $(T-\lambda I)^{l-1} v \neq 0_{\mathscr{V}}$. Then the system of vectors

$$(T-\lambda I)^{l-1}v,\ (T-\lambda I)^{l-2}v, \dots, (T-\lambda I)v,\ v, \tag{18}$$

is called a *Jordan chain* of T corresponding to the eigenvalue λ. The vectors in (18) are called *generalized eigenvectors* (or *root vectors*) corresponding to the eigenvalue λ.

Let $\mathscr{W}$ be a subspace of $\mathscr{V}$ generated by a Jordan chain

$$v_j = (T-\lambda I)^{l-j} v, \qquad j \in \{1,\dots,l\},$$

of T. Note that the vector $v_1 = (T - \lambda I)^{l-1} v$ is an eigenvector of T corresponding to the eigenvalue λ. Therefore $T v_1 = \lambda v_1$. We also have

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

It follows that $\mathscr{W}$ is an invariant subspace of T. If we denote by A the restriction of T to $\mathscr{W}$, then the matrix representation of A with respect to the basis $\{v_1, \ldots, v_l\}$ is

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}. \tag{19}$$

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

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

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

$$\left[\begin{array}{cc} \boxed{\begin{array}{cccc} \lambda_1 & 1 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & \lambda_1 \end{array}} & \begin{array}{cccc} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{array} \\ \begin{array}{cccc} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{array} & \boxed{\begin{array}{cccc} \lambda_2 & 1 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots\ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & \lambda_2 \end{array}} \\ & & \ddots \\ & & & \ddots \\ & & & & \ddots \end{array}\right]. \tag{20}$$

In the above matrix λ_1 and λ_2 are not necessarily distinct eigenvalues. A matrix of the form (20) 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 M outside of the main diagonal and the superdiagonal are 0,

(ii) all elements on the main diagonal of M are eigenvalues of T,

(iii) all elements on the superdiagonal of M are either 1 or 0, and,

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

Theorem 3.1. *Let $\mathscr{V}$ be a vector space over $\mathbb{C}$ and let T be a linear operator on $\mathscr{V}$. Then $\mathscr{V}$ has a Jordan basis with respect to T.*

Proof. We use the notation and the results of Theorem 2.6. Let $j \in \{1, \ldots, k\}$. It is important to notice that each Jordan chain of the nilpotent operator N_j is a Jordan chain of T which corresponds to the eigenvalue λ_j. Since N_j is a nilpotent operator in $\mathscr{L}(\mathscr{W}_j)$, by Theorem 1.1 there exists a basis $\mathscr{B}_j = \{v_{j,1}, \ldots, v_{j,n_j}\}$ for $\mathscr{W}_j$ which consists of Jordan chains of N_j. Consequently, $\mathscr{B}_j$ consists of Jordan chains of T. Since $\mathscr{V}$ is a direct sum of $\mathscr{W}_1, \ldots, \mathscr{W}_k$, the union $\mathscr{B} = \mathscr{B}_1 \cup \cdots \cup \mathscr{B}_k$, that is,

$$\mathscr{B} = \bigl\{v_{1,1}, \ldots, v_{1,n_1}, v_{2,1}, \ldots, v_{2,n_2}, \ldots, v_{k,1}, \ldots, v_{k,n_k}\bigr\}$$

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

The matrix $\mathsf{M}_{\mathscr{B}}^{\mathscr{B}}(T)$ is a block diagonal with the blocks $\mathsf{M}_{\mathscr{B}_j}^{\mathscr{B}_j}(T_j)$, $j \in \{1, \ldots, k\}$, on the diagonal and with zeros every where else:

$$\mathsf{M}_{\mathscr{B}}^{\mathscr{B}}(T) = \begin{bmatrix} \mathsf{M}_{\mathscr{B}_1}^{\mathscr{B}_1}(T_1) & 0 & \cdots & 0 \\ 0 & \mathsf{M}_{\mathscr{B}_2}^{\mathscr{B}_2}(T_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathsf{M}_{\mathscr{B}_k}^{\mathscr{B}_k}(T_k) \end{bmatrix}.$$

Since $T_j = \lambda_j I + N_j$, we have

$$\mathsf{M}_{\mathscr{B}_j}(T_j) = \lambda_j \mathsf{I} + \mathsf{M}_{\mathscr{B}_j}(N_j).$$

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