Chapter 9

Canonical Forms

1 Nilpotent Operators

If a linear transformation L mapping an n-dimensional complex vector space into itself has n linearly independent eigenvectors then the matrix representing L with respect to the basis of eigenvectors will be a diagonal matrix. In this chapter we turn our attention to the case where L does not have enough linearly independent eigenvectors to span V. In this case we would like to choose an ordered basis of V for which the corresponding matrix representation of L will be as nearly diagonal as possible. To simplify matters in this first section we will restrict ourselves to operators having a single eigenvalue λ of multiplicity n. It will be shown that such an operator can be represented by a bidiagonal matrix whose diagonal elements are all equal to λ and whose superdiagonal elements are all 0's and 1's. To do this we require some preliminary definitions and theorems.

Recall from Section 2 of Chapter 5 that a vector space V is a *direct sum* of subspaces S_1 and S_2 if and only if each $\mathbf{v} \in V$ can be written uniquely in the form $\mathbf{x}_1 + \mathbf{x}_2$ where $\mathbf{x}_1 \in S_1$ and $\mathbf{x}_2 \in S_2$. This direct sum is denoted by $S_1 \oplus S_2$.

Lemma 9.1.1. *Let $B_1 = \{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ and $B_2 = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ be disjoint sets which are bases for subspaces S_1 and S_2, respectively, of a vector space V. Then $V = S_1 \oplus S_2$ if and only if $B = B_1 \cup B_2$ is a basis for V.*

Proof. Exercise □

Definition. Let L be a linear operator mapping a vector space V into itself. A subspace S of V is said to be **invariant** under L if $L(\mathbf{x}) \in S$ for each $\mathbf{x} \in S$.

For example if L has an eigenvalue λ and S_λ is the eigenspace corresponding to λ then S_λ is invariant under L since $L(\mathbf{x}) = \lambda\mathbf{x} \in S_\lambda$ for each $\mathbf{x} \in S_\lambda$.

If S is an invariant subspace of L then the restriction of L to S which we will denote $L_{[S]}$ is a linear operator mapping S into itself.

Lemma 9.1.2. *Let L be a linear operator mapping a vector space V into itself and let S_1 and S_2 be invariant subspaces of L with $S_1 \cap S_2 = \{\mathbf{0}\}$. If $S = S_1 \oplus S_2$*

then S is invariant under L. Furthemore if $A = (a_{ij})$ is the matrix representing $L_{[S_1]}$ with respect to the ordered basis $[\mathbf{x}_1, \ldots, \mathbf{x}_r]$ of S_1 and $B = (b_{ij})$ is the matrix representing $L_{[S_2]}$ with respect to the ordered basis $[\mathbf{y}_1, \ldots, \mathbf{y}_k]$ of S_2 then the matrix C representing $L_{[S]}$ with respect to $[\mathbf{x}_1, \ldots, \mathbf{x}_r,\ \mathbf{y}_1, \ldots, \mathbf{y}_k]$ is given by

(1)
$$\begin{aligned} C &= \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & \cdots & a_{1r} & 0 & & 0 \\ \vdots & & & & & \\ a_{r1} & \cdots & a_{rr} & 0 & & 0 \\ 0 & & 0 & b_{11} & \cdots & b_{1k} \\ \vdots & & & & & \\ 0 & & 0 & b_{k1} & & b_{kk} \end{pmatrix} \end{aligned}$$

Proof. We should first note that since $S_1 \cap S_2 = \{\mathbf{0}\}$ it follows that $\mathbf{x}_1, \ldots, \mathbf{x}_r$, $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are linearly independent and hence form a basis for a subspace S of V. By Lemma 9.1.1, $S = S_1 \oplus S_2$ so that it really does make sense to speak of a direct sum of S_1 and S_2. If $\mathbf{s} \in S$ then there exist $\mathbf{x} \in S_1$ and $\mathbf{y} \in S_2$ such that $\mathbf{s} = \mathbf{x} + \mathbf{y}$. Since $L(\mathbf{x}) \in S_1$ and $L(\mathbf{y}) \in S_2$ it follows that

$$L(\mathbf{s}) = L(\mathbf{x}) + L(\mathbf{y})$$

is an element of $S_1 \oplus S_2 = S$. Therefore S is invariant under L.

Let $\mathbf{s}_i^{(1)} = L(\mathbf{x}_i)$ for $i = 1, \ldots, r.$ and $\mathbf{s}_j^{(2)} = L(\mathbf{y}_j)$ for $j = 1, \ldots, k$. Since each $\mathbf{s}_i^{(1)}$ is in S_1 and each $\mathbf{s}_j^{(2)}$ is in S_2 it follows that

$$\begin{aligned} L_{[S]}(\mathbf{x}_i) &= \mathbf{s}_i^{(1)} + \mathbf{0} \\ &= a_{1i}\mathbf{x}_1 + a_{2i}\mathbf{x}_2 + \cdots + a_{ri}\mathbf{x}_r + 0\mathbf{y}_1 + \cdots + 0\mathbf{y}_k \end{aligned}$$

and hence the ith column of the matrix C representing $L_{[S]}$ will be

$$\mathbf{c}_i = (a_{1i}, a_{2i}, \ldots, a_{ri}, 0, \ldots, 0)^T$$

Similarly

$$\begin{aligned} L_{[S]}(\mathbf{y}_j) &= 0 + \mathbf{s}_j^{(2)} \\ &= 0\mathbf{x}_1 + \cdots + 0\mathbf{x}_r + b_{1j}\mathbf{y} + \cdots + b_{kj}\mathbf{y}_k \end{aligned}$$

and hence $\mathbf{c}_{j+r}$ is given by

$$\mathbf{c}_{j+r} = (0, \ldots, 0, b_{1j}, \ldots, b_{kj})^T$$

Thus the matrix C representing $L_{[S]}$ with respect to $[\mathbf{x}_1, \ldots, \mathbf{x}_r, \mathbf{y}_1, \ldots \mathbf{y}_k]$ will be of the form (9.1). $\square$

It is possible to have a direct sum of more than two matrices. In general if $S_1, S_2, \ldots, S_r$ are subspaces of a vector space V then $V = S_1 \oplus \cdots \oplus S_r$ if and only if each $\mathbf{v} \in V$ can be written uniquely as a sum $\mathbf{s}_1 + \cdots + \mathbf{s}_r$ where $\mathbf{s}_i \in S_i$ for $i = 1, \ldots, r$.

Using mathematical induction one can generalize both of the lemmas to direct sums of more than two subspaces. Thus, if each subspace S_i has a basis B_i and the B_i's are all disjoint, then $V = S_1 \oplus \cdots \oplus S_r$ if and only if $B = B_1 \cup B_2 \cup \cdots \cup B_r$ is a basis for V. If $S_1, \ldots, S_r$ are invariant under a linear transformation L and $S = S_1 \oplus \cdots \oplus S_r$, then S is invariant under L and $L_{[S]}$ can be represented by a block diagonal matrix

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{pmatrix} \tag{2}$$

Let L be a linear operator mapping an n-dimensional vector space V into itself. If V can be expressed as a direct sum of invariant subspaces of L then it is possible to represent L as a block diagonal matrix A of the form (2).

The simplest such representation occurs in the case that L is diagonalizable. This occurs when the dimensions of the eigenspaces are equal to the multiplicities of the eigenvalues. In this case we can choose A so that each diagonal block A_i is a diagonal matrix and hence the matrix A is also diagonal.

If however there are any eigenvalues for which the dimension of the eigenspace is less than the multiplicity of the eigenvalue, then the subspace $S_{\lambda_1} \oplus \cdots \oplus S_{\lambda_r}$ will have dimension less than n and hence will be a proper subspace of V. In this case we would like to do is somehow enlarge the deficient S_{λ_i}'s and obtain a direct sum representation of V of the form $S_1 \oplus \cdots \oplus S_r$ where each S_i is invariant under L. Furthermore, we would like the corresponding block representation of L to be as close to a diagonal representation as possible. Indeed we will show that it is possible to find invariant subspaces S_i so that each $L_{[S_i]}$ can be represented by a bidiagonal matrix of a certain form.

As a simple example consider the case where the matrix A representing L is a 3×3 matrix with a triple eigenvalue λ and the eigenspace S_λ has dimension 1. In this case we would like to show that L can be represented by a 3×3 matrix

$$J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

If such a representation is possible then A would have to be similar to J, i.e.,

$AX = XJ$ for some nonsingular matrix X. If we let $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ denote the column vectors of X this would say that

$$A(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)J$$

and hence

$$\begin{aligned} A\mathbf{x}_1 &= \lambda \mathbf{x}_1 \\ A\mathbf{x}_2 &= \mathbf{x}_1 + \lambda \mathbf{x}_2 \\ A\mathbf{x}_3 &= x_2 + \lambda x_3 \end{aligned}$$

or equivalently

$$\begin{aligned} (A - \lambda I)\mathbf{x}_1 &= \mathbf{0} \\ (A - \lambda I)\mathbf{x}_2 &= \mathbf{x}_1 \\ (A - \lambda I)\mathbf{x}_3 &= \mathbf{x}_2 \end{aligned}$$

These equations imply that

(3) $$(A - \lambda I)^3 \mathbf{x}_3 = (A - \lambda I)^2 \mathbf{x}_2 = (A - \lambda I)\mathbf{x}_1 = \mathbf{0}$$

Thus if we can find a vector $\mathbf{x}$ such that

(4) $$(A - \lambda I)^3 \mathbf{x} = \mathbf{0} \qquad \text{and} \qquad (A - \lambda I)^2 \mathbf{x} \neq \mathbf{0}$$

then we can set

(5) $$\mathbf{x}_3 = \mathbf{x}, \qquad \mathbf{x}_2 = (A - \lambda I)\mathbf{x} \quad \text{and} \quad \mathbf{x}_1 = (A - \lambda I)^2 \mathbf{x}$$

The equations given in (4) really provide the key to our problem. If we can find a vector $\mathbf{x}$ satisfying (4) then it is not difficult to show that the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ defined in (5) are linearly independent and hence that $X = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ is invertible. Equation (3) implies that

$$(A - \lambda I)^3 \mathbf{x} = \mathbf{0}$$

for all $\mathbf{x} \in R(X)$. Note that

$$(A - \lambda I)^2 \mathbf{x}_1 \neq \mathbf{0}$$

This type of condition plays an important role in the theory we are about to develop. We state this condition for a general linear operator L in the following definition.

Definition. Let L be a linear operator mapping a vector space V into itself. L is said to be *nilpotent of index k* on V if $L^k(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$ and $L^{k-1}(\mathbf{v}_0) \neq \mathbf{0}$ for some $\mathbf{v}_0 \in V$.

Lemma 9.1.3. *Let L be a linear operator mapping a vector space V into itself and let $\mathbf{v} \in V$. If $L^k(\mathbf{v}) = \mathbf{0}$ and $L^{k-1}(\mathbf{v}) \neq \mathbf{0}$ for some integer $k \geq 1$ then the vectors $\mathbf{v}, L(\mathbf{v}), L^2(\mathbf{v}), \ldots, L^{k-1}(\mathbf{v})$ are linearly independent.*

Proof. The proof will be by induction. The result clearly holds in the case $k = 1$ since

$$\mathbf{v} = L^0(\mathbf{v}) \neq \mathbf{0} \qquad \text{and} \qquad L(\mathbf{v}) = \mathbf{0}$$

and hence we have only a single nonzero vector $\mathbf{v}$. (Here L^0 is taken to be the identity operator.) Assume now that we have a value of k such that the result holds for all $j < k$ and suppose we have a vector $\mathbf{v}$ satisfying

$$L^{k-1}(\mathbf{v}) \neq \mathbf{0} \qquad \text{and} \qquad L^k(\mathbf{v}) = \mathbf{0}$$

To show linear independence we consider the equation

$$\alpha_1 \mathbf{v} + \alpha_2 L(\mathbf{v}) + \cdots + \alpha_k L^{k-1}(\mathbf{v}) = \mathbf{0} \tag{6}$$

If we let $\mathbf{w} = L(\mathbf{v})$ and apply L to both sides of (6) we get

$$\alpha_1 L(\mathbf{v}) + \alpha_2 L^2(\mathbf{v}) + \cdots + \alpha_{k-1} L^{k-1}(\mathbf{v}) = \mathbf{0}$$

or

$$\alpha_1 \mathbf{w} + \alpha_2 L(\mathbf{w}) + \cdots + \alpha_{k-1} L^{k-2}(\mathbf{w}) = \mathbf{0}$$

Since

$$L^{k-2}(\mathbf{w}) = L^{k-1}(\mathbf{v}) \neq \mathbf{0} \qquad \text{and} \qquad L^{k-1}(\mathbf{w}) = L^k(\mathbf{v}) = \mathbf{0}$$

then by our induction hypothesis

$$\mathbf{w}, L(\mathbf{w}), \ldots, L^{k-2}(\mathbf{w})$$

are linearly independent and hence

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{k-1} = 0$$

Thus (6) reduces to

$$\alpha_k L^{k-1}(\mathbf{v}) = \mathbf{0}$$

It follows that α_k must also be zero and hence $\mathbf{v}$, $L(\mathbf{v}), \ldots, L^{k-1}(\mathbf{v})$ are linearly independent. □

If $L^{k-1}(\mathbf{v}) \neq \mathbf{0}$ and $L^k(\mathbf{v}) = \mathbf{0}$ for some $\mathbf{v} \in V$ then the vectors $\mathbf{v}$, $L(\mathbf{v}), \ldots, L^{k-1}(\mathbf{v})$ form a basis for a subspace which we will denote by $C_L(\mathbf{v})$. The subspace $C_L(\mathbf{v})$ is invariant under L since for each

$$\mathbf{w} = \alpha_1 \mathbf{v} + \alpha_2 L(\mathbf{v}) + \cdots + \alpha_k L^{k-1}(\mathbf{v})$$

in $C_L(\mathbf{v})$ we have

$$L(\mathbf{w}) = \alpha_1 L(\mathbf{v}) + \alpha_2 L^2(\mathbf{v}) + \cdots + \alpha_{k-1} L^{k-1}(\mathbf{v})$$

and hence $L(\mathbf{w})$ is also in $C_L(\mathbf{v})$. We will refer to $C_L(\mathbf{v})$ as the *L-cyclic subspace* generated by $\mathbf{v}$. In particular if L is nilpotent of index k then for each nonzero vector $\mathbf{v}_0 \in V$ there is an integer k_0, $1 \leq k_0 \leq k$ such that $L^{k_0-1}(\mathbf{v}_0) \neq \mathbf{0}$ and $L^{k_0}(\mathbf{v}) = \mathbf{0}$. Thus if L is nilpotent on V then one can associate an L-cyclic subspace $C_L(\mathbf{v})$ with each nonzero vector $\mathbf{v}$ in V. It is easily seen that L-cyclic subspaces are invariant under L.

Let $C_L(\mathbf{v})$ be an L cyclic subspace of V with basis $\{\mathbf{v}, L(\mathbf{v}), \ldots, L^{k-1}(\mathbf{v})\}$. Let

$$\mathbf{y}_i = L^{k-i}(\mathbf{v}) \quad \text{for } i = 1, \ldots, k \qquad (\text{where } L^0 = I)$$

Then

$$[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k] = [L^{k-1}(\mathbf{v}), L^{k-2}(\mathbf{v}), \ldots, \mathbf{v}]$$

is an ordered basis for $C_L(\mathbf{v})$. Since

$$\begin{aligned} L(\mathbf{y}_1) &= 0 \\ L(\mathbf{y}_j) &= \mathbf{y}_{j-1} \qquad \text{for} \qquad j = 2, \ldots, k \end{aligned}$$

it follows that the matrix representing $L_{[C_L(\mathbf{v})]}$ with respect to $[\mathbf{y}_1, \ldots, \mathbf{y}_k]$ is given by

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

Thus, $L_{[C_L(\mathbf{v})]}$ can be represented by a bidiagonal matrix with 0's along the main diagonal and 1's along the superdiagonal.

Lemma 9.1.4. *Let L be a linear operator mapping a vector space V into itself. If L is nilpotent of index k on V and $L^{k-1}(\mathbf{v}_1)$, $L^{k-1}(\mathbf{v}_2), \ldots, L^{k-1}(\mathbf{v}_r)$ are linearly independent, then the kr vectors*

$$\begin{aligned} &\mathbf{v}_1, L(\mathbf{v}_1), \ldots, L^{k-1}(\mathbf{v}_1) \\ &\mathbf{v}_2, L(\mathbf{v}_2), \ldots, L^{k-1}(\mathbf{v}_2) \\ &\vdots \\ &\mathbf{v}_r, L(\mathbf{v}_r), \ldots, L^{k-1}(\mathbf{v}_r) \end{aligned}$$

are linearly independent.

Proof. The proof is by induction on k. If $k = 1$ there is nothing to prove. Assume the result holds for all indices less than k and that L is nilpotent of index k. If

$$\begin{aligned}
&\alpha_{11}\mathbf{v}_1 + \alpha_{12}L(\mathbf{v}_1) + \cdots + \alpha_{1k}L^{k-1}(\mathbf{v}_1) \\
&+ \alpha_{21}\mathbf{v}_2 + \alpha_{22}L(\mathbf{v}_2) + \cdots + \alpha_{2k}L^{k-1}(\mathbf{v}_2) \\
&\vdots \\
&+ \alpha_{r1}\mathbf{v}_r + \alpha_{r2}L(\mathbf{v}_r) + \cdots + \alpha_{rk}L^{k-1}(v_r) \\
&= \mathbf{0}
\end{aligned} \tag{7}$$

then applying L to both sides of (7) we get

$$\begin{aligned}
&\alpha_{11}\mathbf{y}_1 + \alpha_{12}L(\mathbf{y}_1) + \cdots + \alpha_{1,k-1}L^{k-2}(\mathbf{y}_1) \\
&+ \alpha_{21}\mathbf{y}_2 + \alpha_{22}L(\mathbf{y}_2) + \cdots + \alpha_{2,k-1}L^{k-2}(\mathbf{y}_2) \\
&\vdots \\
&+ \alpha_{r1}\mathbf{y}_r + \ \alpha_{r2}L(\mathbf{y}_r) + \cdots + \alpha_{r,k-1}L^{k-2}(\mathbf{y}_r) \\
&= \mathbf{0}
\end{aligned} \tag{8}$$

where $\mathbf{y}_i = L(\mathbf{v}_i)$ for $i = 1, \ldots, r$. Since $L^{k-2}(\mathbf{y}_i) = L^{k-1}(\mathbf{v}_i)$ for each i it follows that $L^{k-2}(\mathbf{y}_1), \ldots, L^{k-2}(\mathbf{y}_n)$ are linearly independent. Let S be the subspace of V spanned by

$$\mathbf{y}_1, L(\mathbf{y}_1), \ldots, L^{k-2}(\mathbf{y}_1), \ldots, \mathbf{y}_r, L(\mathbf{y}_r), \ldots, L^{k-2}(\mathbf{y}_r)$$

Since L is nilpotent of index $k-1$ on S it follows by the induction hypothesis that

$$\begin{aligned}
&\mathbf{y}_1, L(\mathbf{y}_1), \ldots, L^{k-2}(\mathbf{y}_1) \\
&\mathbf{y}_2, L(\mathbf{y}_2), \ldots, L^{k-2}(\mathbf{y}_2) \\
&\ \vdots \\
&\mathbf{y}_r, L(\mathbf{y}_r), \ldots, L^{k-2}(\mathbf{y}_r)
\end{aligned}$$

are linearly independent. Therefore

$$\alpha_{ij} = 0 \text{ for } 1 \leq i \leq r,\ 1 \leq j \leq k-1$$

and consequently (8) reduces to

$$\alpha_{1k}L^{k-1}(\mathbf{v}_1) + \alpha_{2k}L^{k-1}(\mathbf{v}_2) + \cdots + \alpha_{rk}L^{k-1}(\mathbf{v}_r) = 0$$

Since $L^{k-1}(\mathbf{v}_1), \ldots, L^{k-1}(\mathbf{v}_r)$ are linearly independent it follows that

$$\alpha_{1k} = \alpha_{2k} = \cdots = \alpha_{rk} = 0$$

and hence

$$\begin{aligned}
&\mathbf{v}_1, L(\mathbf{v}_1), \ldots, L^{k-1}(\mathbf{v}_1) \\
&\mathbf{v}_2, L(\mathbf{v}_2), \ldots, L^{k-1}(\mathbf{v}_2) \\
&\ \vdots \\
&\mathbf{v}_r, L(\mathbf{v}_r), \ldots, L^{k-1}(\mathbf{v}_r)
\end{aligned}$$

are linearly independent. □

Theorem 9.1.5. *Let L be a linear operator mapping an n-dimensional vector space V into itself. If L is nilpotent of index k on V then V can be decomposed into a direct sum of L-cyclic subspaces.*

Proof. The proof will be by induction on k. If $k = 1$ then L is the zero operator on V. Thus if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is any basis of V then $C_L(\mathbf{v}_i)$ is the one-dimensional subspace spanned by $\mathbf{v}_i$ for each i and hence

$$V = C_L(\mathbf{v}_1) \oplus \cdots \oplus C_L(\mathbf{v}_n).$$

Suppose now that we have an integer $k > 1$ such that the result holds for all indices less than k and L is nilpotent of index k. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a basis for $\ker(L^{k-1})$. This basis can be extended to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{y}_1, \ldots, \mathbf{y}_r\}$ of V (where $r = n - m$).

Since $\mathbf{y}_i \notin \ker(L^{k-1})$ it follows that $L^{k-1}(\mathbf{y}_i) \neq 0$. Let

$$B_1 = \left\{\mathbf{y}_1, L(\mathbf{y}_1), \ldots, L^{k-1}(\mathbf{y}_1), \ldots, \mathbf{y}_r, L(\mathbf{y}_r), \ldots, L^{k-1}(\mathbf{y}_r)\right\}$$

We claim B_1 is a basis for a subspace S_1 of V. By Lemma 9.1.4 it suffices to show that $L^{k-1}(\mathbf{y}_1)$, $L^{k-1}(\mathbf{y}_2), \ldots, L^{k-1}(\mathbf{y}_r)$ are linearly independent. If

$$\alpha_1 L^{k-1}(\mathbf{y}_1) + \alpha_2 L^{k-1}(\mathbf{y}_2) + \cdots + \alpha_r L^{k-1}(\mathbf{y}_r) = 0$$

then

$$L^{k-1}(\alpha_1 \mathbf{y}_1 + \cdots + \alpha_r \mathbf{y}_r) = 0$$

and hence $\alpha_1 \mathbf{y}_1 + \cdots + \alpha_r \mathbf{y}_r \in \ker(L^{k-1})$. But then $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$ otherwise $\mathbf{v}_1, \ldots, \mathbf{v}_m$, $\mathbf{y}_1, \ldots, \mathbf{y}_r$ would be dependent. Thus $L^{k-1}(\mathbf{y}_1), \ldots$, $L^{k-1}(\mathbf{y}_r)$ are linearly independent and hence B_1 is a basis for a subspace S_1 of V. It follows from Lemma 9.1.1 that

$$S_1 = C_L(\mathbf{y}_1) \oplus \cdots \oplus C_L(\mathbf{y}_r)$$

If $S_1 \neq V$ extend B_1 to a basis B for V. Let B_2 be the set of additional basis elements (i.e., $B = B_1 \cup B_2$ and $B_1 \cap B_2 = \emptyset$). B_2 is a basis for a subspace S_2 of V and by Lemma 9.1.1 $V = S_1 \oplus S_2$. By construction S_2 is a subspace of $\ker(L^{k-1})$. (If $\mathbf{s} \in S_2$ then it must be of the form $\mathbf{s} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m + 0\mathbf{y}_1 + \cdots + 0\mathbf{y}_r$). Thus L is nilpotent of index $k_1 < k$ on S_2. By the induction hypothesis S_2 can be written as a direct sum of L-cyclic subspaces and since $V = S_1 \oplus S_2$ it follows that V is a direct sum of L-cyclic subspaces. □

Corollary 9.1.6. *If L is a linear operator mapping an n-dimensional vector space V into itself and L is nilpotent of index k on V then L can be represented* *by a matrix of the form*

$$A = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{pmatrix}$$

where each J_i *is a* $k_i \times k_i$ *bidiagonal matrix* ($1 \le k_i \le k$ *and* $\sum_{i=1}^{s} k_i = n$) *with 0's along the main diagonal and 1's along the superdiagonal.*

Proof. By Theorem 9.1.5 we can write

$$V = C_L(\mathbf{v}_1) \oplus \cdots \oplus C_L(\mathbf{v}_s)$$

If $C_L(\mathbf{v}_i)$ has dimension k_i then the matrix representing $L_{[C_L(\mathbf{v}_i)]}$ with respect to $[L^{k_i-1}(\mathbf{v}_i), \ldots, \mathbf{v}_i]$ will be

$$J_i = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$$

The conclusion follows from Lemma 9.1.2. □

It follows from Corollary 9.1.6 that if L is nilpotent on an n-dimensional vector space V then all of its eigenvalues are 0. Conversely if all of the eigenvalues of L are 0 then it follows from Theorem 6.4.3 that L can be represented by a triangular matrix T whose diagonal elements are all 0. Thus for some k, T^k will be the zero matrix and hence L^k will be the zero operator. Thus if L is a linear operator mapping an n-dimensional vector space V into itself then L is nilpotent if and only if all of its eigenvalues are 0.

Corollary 9.1.7. *Let* L *be a linear operator mapping an* n*-dimensional vector space* V *into itself. If* L *has only one distinct eigenvalue* λ *then* L *can be represented by a matrix* A *of the form*

$$A = \begin{pmatrix} J_1(\lambda) & & & \\ & J_2(\lambda) & & \\ & & \ddots & \\ & & & J_s(\lambda) \end{pmatrix} \tag{9}$$

where each $J_i(\lambda)$ is a bidiagonal matrix of the form

$$J_i(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix} \tag{10}$$

Proof. Let $\mathcal{I}$ denote the identity operator V. The eigenvalues of the operator $L - \lambda\mathcal{I}$ are all 0 and hence $L - \lambda\mathcal{I}$ is nilpotent. It follows from Corollary 9.1.6 that with respect to some ordered basis $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ of V the operator $L - \lambda\mathcal{I}$ can be represented by a matrix of the form

$$J = \begin{pmatrix} J_1(0) & & & \\ & J_2(0) & & \\ & & \ddots & \\ & & & J_s(0) \end{pmatrix} \quad \text{where} \quad J_i(0) = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$$

The matrix representing $\lambda\mathcal{I}$ with respect to $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is simply λI. Since $L = (L - \lambda\mathcal{I}) + \lambda\mathcal{I}$ it follows that the matrix representing L with respect to $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is

$$J + \lambda I = \begin{pmatrix} J_1(\lambda) & & & \\ & J_2(\lambda) & & \\ & & \ddots & \\ & & & J_s(\lambda) \end{pmatrix}$$

A matrix of the form (10) is said to be a *simple Jordan matrix*. Thus a simple Jordan matrix is a bidiagonal matrix whose diagonal elements all have the same value λ and whose superdiagonal elements are all 1. □

Example. Let

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

We can think of A as representing an operator from R^5 into R^5. Since $\lambda = 1$ is the only eigenvalue, A is similar to a block diagonal matrix whose diagonal blocks are simple Jordan matrices with 1's along both the diagonal and the superdiagonal. The eigenspace corresponding to $\lambda = 1$ is spanned by the vectors

$\mathbf{x} = (1,0,0,0,0)^T$ and $\mathbf{y} = (0,0,-1,0,1)^T$. Thus the bidiagonal matrix will consist of two simple Jordan blocks, $J_1(1)$ and $J_2(1)$. If we order the blocks so that the first block is the largest then the only possibilities for the block diagonal matrix are:

$$\left(\begin{array}{ccc|cc} 1 & 1 & 0 & & \\ 0 & 1 & 1 & & \\ 0 & 0 & 1 & & \\ \hline & & & 1 & 1 \\ & & & 0 & 1 \end{array}\right) \qquad \text{or} \qquad \left(\begin{array}{cccc|c} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

To determine which of these forms is correct one must compute powers of $A-I$.

$$A - I = \begin{pmatrix} 0 & 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad (A-I)^2 = \begin{pmatrix} 0 & 0 & 2 & 5 & 2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$(A-I)^3 = \begin{pmatrix} 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad (A-I)^4 = O$$

Thus $A-I$ is nilpotent of index 4. The systems

$$(A-I)^k\mathbf{s} = \mathbf{x} \qquad \text{and} \qquad (A-I)^j\mathbf{s} = \mathbf{y}$$

are clearly inconsistent if k and j are greater than 3. We determine the maximum k and maximum j for which these systems are consistent. For $k = 3$ the system

$$(A-I)^3\mathbf{s} = \mathbf{x}$$

is consistent and will have infinitely many solutions. We pick one of these solutions

$$\mathbf{x}_1 = \left(0,0,0,\tfrac{1}{2},0\right)^T$$

To generate the rest of the cyclic subspace we compute

$$\begin{aligned} \mathbf{x}_2 &= (A-I)\mathbf{x}_1 = \left(\tfrac{1}{2},1,\tfrac{1}{2},0,0\right)^T \\ \mathbf{x}_3 &= (A-I)\mathbf{x}_2 = (A-I)^2\mathbf{x}_1 = \left(\tfrac{5}{2},\tfrac{1}{2},0,0,0\right)^T \end{aligned}$$

With respect to the ordered basis $[\mathbf{x},\mathbf{x}_3,\mathbf{x}_2,\mathbf{x}_1]$ the matrix representing the operator A on this subspace will be of the form

$$J_1(1) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The systems

$$(A - I)^j \mathbf{s} = \mathbf{y}$$

are inconsistent for all positive integers j. Thus the cyclic subspace containing $\mathbf{y}$ has dimension 1. It follows that the matrix representing A with respect to $[\mathbf{x}, \mathbf{x}_3, \mathbf{x}_2, \mathbf{x}_1, \mathbf{y}]$ is

$$J = \begin{pmatrix} J_1(1) & \\ & J_2(1) \end{pmatrix} = \left(\begin{array}{cccc|c} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

The reader may verify that if Y is the matrix whose columns are $\mathbf{x}$, $\mathbf{x}_3$, $\mathbf{x}_2$, $\mathbf{x}_1$, $\mathbf{y}$, respectively, then

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

In the next section we will show that a matrix A with distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ is similar to a matrix J of the form

$$J = \begin{pmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_m \end{pmatrix}$$

where each B_i is of the form (9) with diagonal elements equal to λ_i, i.e.,

$$B_i = \begin{pmatrix} J_1(\lambda_i) & & & \\ & J_2(\lambda_i) & & \\ & & \ddots & \\ & & & J_s(\lambda_i) \end{pmatrix}$$

where the $J_k(\lambda_i)$'s are simple Jordan matrices. We say that J is the *Jordan canonical form* of A. The Jordan canonical form is unique except for a reordering of the blocks.

Exercises

1. Let L be a linear operator on a vector space V of dimension 5 and let A be any matrix representing L. If L is nilpotent of index 3 then what are the possible Jordan canonical forms of A?
2. Let A be a 4×4 matrix whose only eigenvalue is $\lambda = 2$. What are the possible Jordan canonical forms of A?

3. Let L be a linear operator on a vector space V of dimension 6 and let A be a matrix representing L. If L has only one distinct eigenvalue λ and the eigenspace S_λ has dimension 3 then what are the possible Jordan canonical forms of A?

4. For each of the following find a matrix S such that $S^{-1}AS$ is a simple Jordan matrix.

$$\text{(a)}\quad A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & -1 & 2 \end{pmatrix} \qquad \text{(b)}\quad A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

5. In each of the following find a matrix S such that $S^{-1}AS$ is the Jordan canonical form of A.

$$\text{(a)}\quad A = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 3 & -1 & 0 \end{pmatrix} \qquad \text{(b)}\quad A = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

6. Let S_1 and S_2 be subspaces of a vector space V. Prove that $V = S_1 \oplus S_2$ if and only if $V = S_1 + S_2$ and $S_1 \cap S_2 = \{0\}$.

7. Prove Lemma 9.1.1.

8. Let L be a linear operator mapping a vector space V into itself. Show that $\ker(L)$ and $R(L)$ are invariant subspaces of V under L.

9. Let L be a linear operator on a vector space V. Let $S_k[\mathbf{v}]$ denote the subspace spanned by $\mathbf{v}$, $L(\mathbf{v}), \ldots, L^{k-1}(\mathbf{v})$. Show that $S_k[\mathbf{v}]$ is invariant under L if and only if $L^k(\mathbf{v}) \in S_k[\mathbf{v}]$.

10. Let L be a linear operator on a vector space V and let S be a subspace of V. Let $\mathcal{I}$ represent the identity operator and let λ be a scalar. Show that L is invariant on S if and only if $L - \lambda\mathcal{I}$ is invariant on S.

11. Let S be the subspace of $C[a, b]$ spanned by x, xe^x, and $xe^x + x^2e^x$. Let D be the differentiation operator on S.

    (a) Find a matrix A representing D with respect to $[e^x, xe^x, xe^x + x^2e^x]$.

    (b) Determine the Jordan canonical form of A and the corresponding basis of S.

12. Let D denote the linear operator on P_n defined by $D(p) = p'$ for all $p \in P_n$. Show that D is nilpotent and can be represented by a simple Jordan matrix.

2 The Jordan Canonical Form

In this section we will show that any linear operator L on an n-dimensional vector space V can be represented by a block diagonal matrix whose diagonal blocks are simple Jordan matrices. We will apply this result to solving systems of linear differential equations of the form $Y' = AY$ where A is defective.

Let us begin by considering the case where L has more than one distinct eigenvalue. We wish to show that if L has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ then V can be decomposed into a direct sum of invariant subspaces $S_1, \ldots, S_k$ such that $L - \lambda_i I$ is nilpotent on S_i for each $i = 1, \ldots, k$. To do this we must first prove the following lemma and theorem.

Lemma 9.2.1. *If L is a linear operator mapping an n-dimensional vector space V into itself then there exists a positive integer k_0 such that $\ker(L^{k_0}) = \ker(L^{k_0+k})$ for all $k > 0$.*

Proof. If $i < j$ then clearly $\ker(L^i)$ is a subspace of $\ker(L^j)$. We claim that if $\ker(L^i) = \ker(L^{i+1})$ for some i then $\ker(L^i) = \ker(L^{i+k})$ for all $k \geq 1$. We will prove this by induction on k. In the case $k = 1$, there is nothing to prove. Assume for some $k > 1$ the result holds all indices less than k. If $\mathbf{v} \in \ker(L^{i+k})$ then

$$0 = L^{i+k}(\mathbf{v}) = L^{i+k-1}(L(\mathbf{v}))$$

Thus $L(\mathbf{v}) \in \ker(L^{i+k-1})$. By the induction hypothesis $\ker(L^{i+k-1}) = \ker(L^i)$. Therefore $L(\mathbf{v}) \in \ker(L^i)$ and hence $\mathbf{v} \in \ker(L^{i+1})$. Since $\ker(L^{i+1}) = \ker(L^i)$ it follows that $\mathbf{v} \in \ker(L^i)$ and hence $\ker(L^i) = \ker(L^{i+k})$. Thus if $\ker(L^{i+1}) = \ker(L^i)$ for some i then

$$\ker(L^i) = \ker(L^{i+1}) = \ker(L^{i+1}) = \ldots$$

Since V is finite dimensional, the dimension of $\ker(L^k)$ cannot keep increasing as k increases. Thus for some k_0 we must have $\dim(\ker(L^{k_0})) = \dim(\ker(L^{k_0+1}))$ and hence $\ker(L^{k_0})$ and $\ker(L^{k_0+1})$ must be equal. It follows then that

$$\ker(L^{k_0}) = \ker(L^{k_0+1}) = \ker(L^{k_0+2}) = \ldots$$

□

Theorem 9.2.2. *If L is a linear transformation on an n-dimensional vector space V then there exist invariant subspaces X and Y such that $V = X \oplus Y$, L is nilpotent on X, and $L_{[Y]}$ is invertible.*

Proof. Choose k_0 to be the smallest positive integer such that $\ker(L^{k_0}) = \ker(L^{k_0+1})$. It follows from Lemma 9.2.1 that $\ker(L^{k_0}) = \ker(L^{k_0+j})$ for all $j \geq 1$. Let $X = \ker(L^{k_0})$. Clearly X is invariant under L for if $\mathbf{x} \in X$ then $L(\mathbf{x}) \in \ker(L^{k_0-1})$ which is a proper subspace of $\ker(L^{k_0})$. Let $Y = R(L^{k_0})$. If $\mathbf{w} \in X \cap Y$ then $\mathbf{w} = L^{k_0}(\mathbf{v})$ for some $\mathbf{v}$ and hence

$$\mathbf{0} = L^{k_0}(\mathbf{w}) = L^{k_0}(L^{k_0}(\mathbf{v})) = L^{2k_0}(\mathbf{v})$$

Thus $\mathbf{v} \in \ker(L^{2k_0}) = \ker(L^{k_0})$ and hence

$$\mathbf{w} = L^{k_0}(\mathbf{v}) = \mathbf{0}$$

Therefore $X \cap Y = \{\mathbf{0}\}$. We claim $V = X \oplus Y$. Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ be a basis for X and let $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n-r}\}$ be a basis for Y. By Lemma 9.2.1 it suffices to show that $\mathbf{x}_1, \ldots, \mathbf{x}_r, \mathbf{y}_1, \ldots \mathbf{y}_{n-r}$ are linearly independent and hence form a basis for V. If

(1) $$\alpha_1\mathbf{x}_1 + \cdots + \alpha_r\mathbf{x}_r + \beta_1\mathbf{y}_1 + \cdots + \beta_{n-r}\mathbf{y}_{n-r} = 0$$

then applying L^{k_0} to both sides gives

$$\beta_1 L^{k_0}(\mathbf{y}_1) + \cdots + \beta_{n-r}L^{k_0}(\mathbf{y}_{n-r}) = \mathbf{0}$$

or

$$L^{k_0}(\beta_1\mathbf{y}_1 + \cdots + \beta_{n-r}\mathbf{y}_{n-r}) = 0$$

Therefore $\beta_1\mathbf{y}_1 + \cdots + \beta_{n-r}\mathbf{y}_{n-r} \in X \cap Y$ and hence

$$\beta_1\mathbf{y}_1 + \cdots + \beta_{n-r}\mathbf{y}_{n-r} = \mathbf{0}$$

Since the $\mathbf{y}_i$'s are linearly independent it follows that

$$\beta_1 = \beta_2 = \cdots = \beta_{n-r} = 0$$

and hence (1) simplifies to

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_r\mathbf{x}_r = \mathbf{0}$$

Since the $\mathbf{x}_i$'s are linearly independent it follows that

$$\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$$

Thus, $\mathbf{x}_1, \ldots, \mathbf{x}_r, \mathbf{y}_1, \ldots, \mathbf{y}_{n-r}$ are linearly independent and therefore $V = X \oplus Y$. L is invariant and nilpotent on X. We claim that L is invariant and invertible on Y. Let $\mathbf{y} \in Y$, then $\mathbf{y} = L^{k_0}(\mathbf{v})$ for some $\mathbf{v} \in V$. Thus,

$$L(\mathbf{y}) = L(L^{k_0}(\mathbf{v})) = L^{k_0+1}(\mathbf{v}) = L^{k_0}(L(\mathbf{v}))$$

Therefore $L(\mathbf{y}) \in Y$ and hence Y is invariant under L. To prove $L_{[Y]}$ is invertible it suffices to show that

$$\ker(L_{[Y]}) = Y \cap \ker(L) = \{\mathbf{0}\}$$

This, however, follows immediately since $\ker(L) \subset X$ and $X \cap Y = \{\mathbf{0}\}$. □

We are now ready to prove the main result of this section.

Theorem 9.2.3. *Let L be a linear operator mapping a finite dimensional vector space V into itself. If $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of L then V can be decomposed into a direct sum*

$$X_1 \oplus X_2 \oplus \cdots \oplus X_k$$

such that $L - \lambda \mathcal{I}$ is nilpotent on X_i and the dimension of X_i equals the multiplicity of λ_i.

Proof. Let $L_1 = L - \lambda_1 \mathcal{I}$. By Theorem 9.2.2 there exist subspaces X_1 and Y_1 which are invariant under L_1 such that $V = X_1 \oplus Y_1$, L_1 is nilpotent on X_1 and $L_{1[Y]}$ is invertible. It follows that X_1 and Y_1 are also invariant under L. By Corollary 9.1.2, $L_{[X_1]}$ can be represented by a block diagonal matrix A_1 where diagonal blocks are simple Jordan matrices whose diagonal elements all equal λ_1. Thus

$$\det(A_1 - \lambda I) = (\lambda_1 - \lambda)^{m_1}$$

where m_1 is the dimension of X_1. Let B_1 be a matrix representing $L_{[Y_1]}$. Since L_1 is invertible on Y_1 it follows that λ_1 is not an eigenvalue of B_1. Thus

$$\det(B_1 - \lambda I) = q(\lambda)$$

where $q(\lambda_1) \neq 0$. It follows from Lemma 9.1.2 that the operator L on V can be represented by the matrix

$$A = \begin{pmatrix} A_1 & \\ & B_1 \end{pmatrix}$$

Thus if each eigenvalue λ_i of L has multiplicity r_i then

$$\begin{aligned} (\lambda_1 - \lambda)^{r_1}(\lambda_2 - \lambda)^{r_2} \cdots (\lambda_k - \lambda)^{r_k} &= \det(A - \lambda I) \\ &= \det(A_1 - \lambda I)\,\det(B_1 - \lambda I) \\ &= (\lambda_1 - \lambda)^{m_1} q(\lambda) \end{aligned}$$

Therefore $r_1 = m_1$ and

$$q(\lambda) = (\lambda_2 - \lambda)^{r_2} \cdots (\lambda_k - \lambda)^{r_k}$$

If we consider the operator $L_2 = L - \lambda_2 \mathcal{I}$ on the vector space Y_1 then we can decompose Y_1 into a direct sum $X_2 \oplus Y_2$ such that X_2 and Y_2 are invariant under L, L_2 is nilpotent on X_2 and $L_{[Y_2]}$ is invertible. Indeed we can continue this process of decomposing Y_i into a direct sum $X_{i+1} \oplus Y_{i+1}$ until we obtain a direct sum of the form

$$V = X_1 \oplus X_2 \oplus \cdots \oplus X_{k-1} \oplus Y_{k-1}$$

The vector space Y_{k-1} will be of dimension r_k with a single eigenvalue λ_k. Thus, if we set $X_k = Y_{k-1}$ then $L - \lambda_k \mathcal{I}$ will be nilpotent on X_k and we will have the desired decomposition of V. ☐

It follows from Theorem 9.2.3 that each operator L mapping an n-dimensional vector space V into itself can be represented by a block diagonal matrix of the form

$$J = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}$$

where each A_i is an $r_i \times r_i$ block diagonal matrix (r_i = multiplicity of λ_i) whose blocks consist of simple Jordan matrices with λ_i's along the main diagonal.

If A is an $n \times n$ matrix then A represents the operator L_A with respect to the standard basis on R^n where L_A is defined by

$$L_A(\mathbf{x}) = A\mathbf{x} \qquad \text{for each } \mathbf{x} \in R^n$$

By the preceding remarks L_A can be represented by a matrix J of the form just described. It follows that A is similar to J. Thus each $n \times n$ matrix A with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ is similar to a matrix J of the form

$$(2) \qquad J = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}$$

where A_i is an $r_i \times r_i$ matrix (r_i = multiplicity of λ_i) of the form

$$(3) \qquad A_i = \begin{pmatrix} J_1(\lambda_i) & & & \\ & J_2(\lambda_i) & & \\ & & \ddots & \\ & & & J_s(\lambda_i) \end{pmatrix}$$

with the $J(\lambda_i)$'s being simple Jordan matrices. The matrix J defined by (2) and (3) is called the *Jordan canonical form* of A. The Jordan canonical form of a matrix is unique except for a reordering of the simple Jordan blocks along the diagonal.

Example Find the Jordan canonical form of the matrix

$$A = \begin{pmatrix} -3 & 1 & 0 & 1 & 1 \\ -3 & 1 & 0 & 1 & 1 \\ -4 & 1 & 0 & 2 & 1 \\ -3 & 1 & 0 & 1 & 1 \\ -4 & 1 & 0 & 1 & 2 \end{pmatrix}$$

Solution: The characteristic polynomial of A is

$$|A - \lambda I| = \lambda^4(1 - \lambda)$$

The eigenspace corresponding to $\lambda = 1$ is spanned by $\mathbf{x}_1 = (1, 1, 1, 1, 2)^T$ and the eigenspace corresponding to $\lambda = 0$ is spanned by $\mathbf{x}_2 = (1, 1, 0, 1, 1)^T$ and $\mathbf{x}_3 = (0, 0, 1, 0, 0)^T$. Thus the Jordan canonical form of A then will consist of three simple Jordan blocks. Except for a reordering of the blocks there are only two possibilities:

$$\left(\begin{array}{c|c|cccc} 1 & & & & \\ \hline & 0 & & & \\ \hline & & 0 & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{array}\right) \qquad \text{or} \qquad \left(\begin{array}{c|cc|cc} 1 & & & & \\ \hline & 0 & 1 & & \\ & & 0 & & \\ \hline & & & 0 & 1 \\ & & & & 0 \end{array}\right)$$

To determine which of these forms is correct we compute $(A - 0I)^2 = A^2$.

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

Next we consider the systems

$$A^2\mathbf{x} = \mathbf{x}_i$$

for $i = 2, 3$. Since these systems turn out to be inconsistent, the Jordan canonical form of A cannot have any 3×3 simple Jordan blocks and consequently it must be of the form

$$J = X^{-1}AX = \left(\begin{array}{c|cc|cc} 1 & & & & \\ \hline & 0 & 1 & & \\ & & 0 & & \\ \hline & & & 0 & 1 \\ & & & & 0 \end{array}\right)$$

To find X we must solve

$$A\mathbf{x} = \mathbf{x}_i$$

for $i = 2, 3$. The system, $A\mathbf{x} = \mathbf{x}_2$, has infinitely many solutions. We need choose only one of these say $\mathbf{x}_4 = (1, 3, 0, 0, 1)^T$. Similarly $A\mathbf{x} = \mathbf{x}_3$ has infinitely many solutions one of which is $\mathbf{x}_5 = (1, 0, 0, 2, 1)^T$. Let

$$X = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 3 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 2 \\ 2 & 1 & 1 & 0 & 1 \end{pmatrix}$$

The reader may verify that $X^{-1}AX = J$. □

One of the main applications of the Jordan canonical form is in solving systems of linear differential equations which have defective coefficient matrices. Given such a system

$$\mathbf{Y}'(t) = A\mathbf{Y}(t)$$

we can simplify it by using the Jordan canonical form of A. Indeed if $A = XJX^{-1}$ then

$$\mathbf{Y}' = (XJX^{-1})\mathbf{Y}$$

Thus if we set $\mathbf{Z} = X^{-1}\mathbf{Y}$ then $\mathbf{Y}' = X\mathbf{Z}'$ and the system simplifies to

$$X\mathbf{Z}' = XJ\mathbf{Z}$$

Multiplying by X^{-1} we get

(4) $$\mathbf{Z}' = J\mathbf{Z}$$

Because of the structure of J this new system is much easier to solve. Indeed solving (4) will only involve solving a number of smaller systems each of the form

$$\begin{aligned} z_1' &= \lambda z_1 + z_2 \\ z_2' &= \lambda z_2 + z_3 \\ &\vdots \\ z_{k-1}' &= \lambda z_{k-1} + z_k \\ z_k' &= \lambda z_k \end{aligned}$$

These equations can be solved one at a time starting with the last. The solution to the last equation is clearly

$$z_k = ce^{\lambda t}$$

The solution to any equation of the form

$$z'(t) - \lambda z(t) = u(t)$$

is given by

$$z(t) = e^{\lambda t} \int e^{-\lambda t} u(t)dt$$

Thus we can solve

$$z_{k-1}' - \lambda z_{k-1} = z_k$$

for z_{k-1} and then solve

$$z_{k-2}' - \lambda z_{k-2} = z_{k-1}$$

for z_{k-2}, etc.

Example. Solve the initial value problem

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \\ y_4' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 2 \\ 1 & 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$
$$y_1(0) = y_2(0) = y_3(0) = 0, \; y_4(0) = 2$$

Solution: The coefficient matrix A has two distinct eigenvalues $\lambda_1 = 0$ and $\lambda_2 = 2$ each of multiplicity 2. The corresponding eigenspaces are both dimension 1. Using the methods of this section A can be factored into a product XJX^{-1} where

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

The choice of X is not unique. The reader may verify that the one we have calculated

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

does the job. The system

$$X' = JX$$

can be broken up into two systems

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} x_3' &= 2x_3 + x_4 \\ x_4' &= 2x_4 \end{aligned}$$

The first system is not difficult to solve.

$$\begin{aligned} x_1 &= c_1 t + c_2 \\ x_2 &= c_1 \qquad\qquad (c_1 \text{ and } c_2 \text{ are constants}) \end{aligned}$$

To solve the second system we solve first

$$x_4' = 2x_4$$

getting

$$x_4 = c_3 e^{2t}$$

Thus

$$x_3' - 2x_3 = c_3 e^{2t}$$

and hence

$$x_3 = e^{2t} \int e^{-2t}(c_3 e^{2t})dt = e^{2t}(c_3 t + c_4)$$

Finally we have

$$Y = JX = \begin{pmatrix} (c_1 t + c_2) + c_1 - (c_3 t + c_4)e^{2t} + c_3 e^{2t} \\ (c_1 t + c_2) + c_1 + (c_3 t + c_4)e^{2t} - c_3 e^{2t} \\ -(c_1 t + c_2) + (c_3 t + c_4)e^{2t} \\ (c_1 t + c_2) + (c_3 t + c_4)e^{2t} \end{pmatrix}$$

If we set $t = 0$ and use the initial conditions to solve for the c_i's we get

$$c_1 = -1, \; c_2 = c_3 = c_4 = 1$$

Thus the solution to the initial value problem is

$$\begin{aligned} y_1 &= -t - te^{2t} \\ y_2 &= -t + te^{2t} \\ y_3 &= -1 + t + (1 + t)e^{2t} \\ y_4 &= 1 - t + (1 + t)e^{2t} \end{aligned}$$

□

The Jordan canonical form not only provides a nice representation of an operator but it allows us to solve systems of the form $\mathbf{Y}' = A\mathbf{Y}$ even when the coefficient matrix is defective. From a theoretical point of view its importance cannot be questioned. As far as practical applications go, however, it is generally not very useful.

If $n \geq 5$ it is usually necessary to calculate the eigenvalues of A by some numerical method. The calculated λ_i's are only approximations to the actual eigenvalues. Thus we could have calculated values λ_1' and λ_2' which are unequal while actually $\lambda_1 = \lambda_2$. So in practice it may be difficult to determine the correct multiplicity eigenvalues. Furthermore, in order to solve $\mathbf{Y}' = A\mathbf{Y}$ we need to find the similarity matrix X such that $A = XJX^{-1}$. However, when A has multiple eigenvalues the matrix X may be very sensitive to perturbations and in practice one is not guaranteed that the entries of the computed similarity matrix will have any digits of accuracy whatsoever. A recommended alternative is to compute the matrix exponential e^A and use it to solve the system $\mathbf{Y}' = A\mathbf{Y}$.

Exercises

1. Let A be a 4×4 matrix whose only eigenvalue is $\lambda = 2$. What are the possible Jordan canonical forms for A?
2. Let A be a 5×5 matrix. If $A^2 \neq 0$ and $A^3 = 0$, what are the possible Jordan canonical forms for A?

3. Find the Jordan canonical form J for each of the following matrices and determine a matrix X such that $X^{-1}AX = J$.

(a) $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & -1 & 2 \end{pmatrix}$

(b) $A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$

(c) $A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

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

(e) $A = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$

4. Let L be a linear operator on a finite dimensional vector space V.

(a) Show that $R(L^i) \subset R(L^j)$ whenever $i > j$.

(b) If for some k_0, $R(L^{k_0}) = R(L^{k_0+1})$ then $R(L^{k_0}) = R(L^{k_0+k})$ for all $k \geq 1$.

5. Let L be as in Exercise 4.

(a) Show that there is a smallest positive integer k_0 such that $R(L^{k_0}) = R(L^{k_0+1})$.

(b) Let k_1 be the smallest positive integer such that $\ker(L^{k_1}) = \ker(L^{k_1+1})$. Show that $k_1 = k_0$.

6. Solve the initial value problem

$$\begin{aligned} y_1' &= \qquad\qquad y_3 \\ y_2' &= y_1 - y_2 + 2y_3 \\ y_3' &= y_1 - y_2 + y_3 \\ y_1(0) &= 0, \ y_2(0) = 0, \ y_3(0) = -1 \end{aligned}$$

7. Suppose

$$X^{-1}AX = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} = J$$

If $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are the column vectors of X define

$$\begin{aligned} \mathbf{z}_1 &= a\mathbf{x}_1 \\ \mathbf{z}_2 &= a\mathbf{x}_2 + b\mathbf{x}_1 \\ \mathbf{z}_3 &= a\mathbf{x}_3 + b\mathbf{x}_2 + c\mathbf{x}_3 \end{aligned}$$

where a, b, and c are scalars and $a \neq 0$.

(a) If $Z = (z_1 \ z_2 \ z_3)$ show that

$$AZ = ZJ$$

(b) Let

$$B = \begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{pmatrix}$$

Show that $BJB^{-1} = X^{-1}AX = J$.