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 $\lambda$ 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 $\lambda$ 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 $v \in V$ can be written uniquely in the form $x_1 + x_2$ where $x_1 \in S_1$ and $x_2 \in S_2$. This direct sum is denoted by $S_1 \oplus S_2$.

**Lemma 9.1.1.** Let $B_1 = \{x_1, \ldots, x_r\}$ and $B_2 = \{y_1, \ldots, y_k\}$ 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 $L(x) \in S$ for each $x \in S$.

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

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 $S_1 \cap S_2 = \{0\}$. If $S = S_1 \oplus S_2$
then $S$ is invariant under $L$. Furthermore if $A = (a_{ij})$ is the matrix representing $L_{[S_1]}$ with respect to the ordered basis $[x_1, \ldots, x_r]$ of $S_1$ and $B = (b_{ij})$ is the matrix representing $L_{[S_2]}$ with respect to the ordered basis $[y_1, \ldots, y_k]$ of $S_2$ then the matrix $C$ representing $L_{[S]}$ with respect to $[x_1, \ldots, x_r, y_1, \ldots, y_k]$ is given by

\begin{equation}
C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1r} & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{r1} & \cdots & a_{rr} & 0 & 0 \\ 0 & 0 & b_{11} & \cdots & b_{1k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & b_{k1} & b_{kk} & \end{bmatrix}
\end{equation}

**Proof.** We should first note that since $S_1 \cap S_2 = \{0\}$ it follows that $x_1, \ldots, x_r, y_1, \ldots, y_k$ 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 $s \in S$ then there exist $x \in S_1$ and $y \in S_2$ such that $s = x + y$. Since $L(x) \in S_1$ and $L(y) \in S_2$ it follows that

$$L(s) = L(x) + L(y)$$

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

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

$$L_{[S]}(x_i) = s^{(1)}_i + 0 = a_{11}x_1 + a_{21}x_2 + \cdots + a_{ri}x_r + 0y_1 + \cdots + 0y_k$$

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

$$c_i = (a_{11}, a_{21}, \ldots, a_{ri}, 0, \ldots, 0)^T$$

Similarly

$$L_{[S]}(y_j) = 0 + s^{(2)}_j = 0x_1 + \cdots + 0x_r + b_{1j}y_1 + \cdots + b_{kj}y_k$$

and hence $c_{j+r}$ is given by

$$c_{j+r} = (0, \ldots, 0, b_{1j}, \ldots, b_{kj})^T$$
Thus the matrix $C$ representing $L_{[S]}$ with respect to $[x_1, \ldots, x_r, y_1, \ldots, y_k]$ will be of the form (9.1).

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 $v \in V$ can be written uniquely as a sum $s_1 + \cdots + s_r$ where $s_i \in S_i$ 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

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

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 $S_{\lambda_1} \oplus \cdots \oplus S_{\lambda_k}$ 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 $S_{\lambda_i}$’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 $3 \times 3$ matrix with a triple eigenvalue $\lambda$ and the eigenspace $S_{\lambda}$ has dimension 1. In this case we would like to show that $L$ can be represented by a $3 \times 3$ 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 \( x_1, x_2, x_3 \) denote the column vectors of \( X \) this would say that

\[
A(x_1, x_2, x_3) = (x_1, x_2, x_3)J
\]

and hence

\[
Ax_1 = \lambda x_1 \\
x_2 = x_1 + \lambda x_2 \\
x_3 = x_2 + \lambda x_3
\]

or equivalently

\[
(A - \lambda I)x_1 = 0 \\
(A - \lambda I)x_2 = x_1 \\
(A - \lambda I)x_3 = x_2
\]

These equations imply that

\[
(A - \lambda I)^2 x_3 = (A - \lambda I)^2 x_2 = (A - \lambda I)x_1 = 0
\]

Thus if we can find a vector \( x \) such that

\[
(A - \lambda I)^2 x = 0 \quad \text{and} \quad (A - \lambda I)^2 x \neq 0
\]

then we can set

\[
x_3 = x, \quad x_2 = (A - \lambda I)x \quad \text{and} \quad x_1 = (A - \lambda I)^2 x
\]

The equations given in (4) really provide the key to our problem. If we can find a vector \( x \) satisfying (4) then it is not difficult to show that the vectors \( x_1, x_2, \) and \( x_3 \) defined in (5) are linearly independent and hence that \( X = (x_1, x_2, x_3) \) is invertible. Equation (3) implies that

\[
(A - \lambda I)^3 x = 0
\]

for all \( x \in R(X) \). Note that

\[
(A - \lambda I)^2 x_1 \neq 0
\]

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 \( L^k(v) = 0 \) for all \( v \in V \) and \( L^{k-1}(v_0) \neq 0 \) for some \( v_0 \in V \).
Lemma 9.1.3. Let $L$ be a linear operator mapping a vector space $V$ into itself and let $v \in V$. If $L^k(v) = 0$ and $L^{k-1}(v) \neq 0$ for some integer $k \geq 1$ then the vectors $v, L(v), L^2(v), \ldots, L^{k-1}(v)$ are linearly independent.

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

$v = L^0(v) \neq 0$ and $L(v) = 0$

and hence we have only a single nonzero vector $v$. (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 $v$ satisfying

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

To show linear independence we consider the equation

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

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

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

or

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

Since

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

then by our induction hypothesis

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

are linearly independent and hence

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

Thus (6) reduces to

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

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

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

$w = \alpha_1 v + \alpha_2 L(v) + \cdots + \alpha_k L^{k-1}(v)$
in $C_L(v)$ we have

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

and hence $L(w)$ is also in $C_L(v)$. We will refer to $C_L(v)$ as the $L$-cyclic subspace generated by $v$. In particular if $L$ is nilpotent of index $k$ then for each nonzero vector $v_0 \in V$ there is an integer $k_0$, $1 \leq k_0 \leq k$ such that $L^{k_0-1}(v_0) \neq 0$ and $L^{k_0}(v) = 0$. Thus if $L$ is nilpotent on $V$ then one can associate an $L$-cyclic subspace $C_L(v)$ with each nonzero vector $v$ in $V$. It is easily seen that $L$-cyclic subspaces are invariant under $L$.

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

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

Then

$$[y_1, y_2, \ldots, y_k] = [L^{k-1}(v), L^{k-2}(v), \ldots, v]$$

is an ordered basis for $C_L(v)$. Since

$$L(y_1) = 0$$

$$L(y_j) = y_{j-1} \quad \text{for} \quad j = 2, \ldots, k$$

it follows that the matrix representing $L_{[C_L(v)]}$ with respect to $[y_1, \ldots, y_k]$ 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, $L_{[C_L(v)]}$ 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 $L^{k-1}(v_1), L^{k-1}(v_2), \ldots, L^{k-1}(v_r)$ are linearly independent, then the $kr$ vectors

$$v_1, L(v_1), \ldots, L^{k-1}(v_1)$$
$$v_2, L(v_2), \ldots, L^{k-1}(v_2)$$
$$\vdots$$
$$v_r, L(v_r), \ldots, L^{k-1}(v_r)$$

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

\[
\begin{align*}
\alpha_{11}v_1 + \alpha_{12}L(v_1) + \cdots + \alpha_{1k}L^{k-1}(v_1) \\
+ \alpha_{21}v_2 + \alpha_{22}L(v_2) + \cdots + \alpha_{2k}L^{k-1}(v_2) \\
\vdots \\
+ \alpha_{r1}v_r + \alpha_{r2}L(v_r) + \cdots + \alpha_{rk}L^{k-1}(v_r) \\
= 0
\end{align*}
\]

(7)

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

\[
\begin{align*}
\alpha_{11}y_1 + \alpha_{12}L(y_1) + \cdots + \alpha_{1,k-1}L^{k-2}(y_1) \\
+ \alpha_{21}y_2 + \alpha_{22}L(y_2) + \cdots + \alpha_{2,k-1}L^{k-2}(y_2) \\
\vdots \\
+ \alpha_{r1}y_r + \alpha_{r2}L(y_r) + \cdots + \alpha_{r,k-1}L^{k-2}(y_r) \\
= 0
\end{align*}
\]

(8)

where \( y_i = L(v_i) \) for \( i = 1, \ldots, r \). Since \( L^{k-2}(y_i) = L^{k-1}(v_i) \) for each \( i \) it follows that \( L^{k-2}(y_1), \ldots, L^{k-2}(y_n) \) are linearly independent. Let \( S \) be the subspace of \( V \) spanned by

\[
y_1, L(y_1), \ldots, L^{k-2}(y_1), \ldots, y_r, L(y_r), \ldots, L^{k-2}(y_r)
\]

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

\[
y_1, L(y_1), \ldots, L^{k-2}(y_1) \\
y_2, L(y_2), \ldots, L^{k-2}(y_2) \\
\vdots \\
y_r, L(y_r), \ldots, L^{k-2}(y_r)
\]

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

\[
\alpha_{1k}L^{k-1}(v_1) + \alpha_{2k}L^{k-1}(v_2) + \cdots + \alpha_{rk}L^{k-1}(v_r) = 0
\]

Since \( L^{k-1}(v_1), \ldots, L^{k-1}(v_r) \) are linearly independent it follows that

\[
\alpha_{1k} = \alpha_{2k} = \cdots = \alpha_{rk} = 0
\]

and hence

\[
v_1, L(v_1), \ldots, L^{k-1}(v_1) \\
v_2, L(v_2), \ldots, L^{k-1}(v_2) \\
\vdots \\
v_r, L(v_r), \ldots, L^{k-1}(v_r)
\]
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 $\{v_1, \ldots, v_n\}$ is any basis of $V$ then $C_L(v_i)$ is the one-dimensional subspace spanned by $v_i$ for each $i$ and hence

$$V = C_L(v_1) \oplus \cdots \oplus C_L(v_n).$$

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 $\{v_1, \ldots, v_m\}$ be a basis for $\ker(L^{k-1})$. This basis can be extended to a basis $\{v_1, \ldots, v_m, y_1, \ldots, y_r\}$ of $V$ (where $r = n - m$).

Since $y_i \not\in \ker(L^{k-1})$ it follows that $L^{k-1}(y_i) \neq 0$. Let

$$B_1 = \{y_1, L(y_1), \ldots, L^{k-1}(y_1), \ldots, y_r, L(y_r), \ldots, L^{k-1}(y_r)\}$$

We claim $B_1$ is a basis for a subspace $S_1$ of $V$. By Lemma 9.1.4 it suffices to show that $L^{k-1}(y_1), L^{k-1}(y_2), \ldots, L^{k-1}(y_r)$ are linearly independent. If

$$\alpha_1L^{k-1}(y_1) + \alpha_2L^{k-1}(y_2) + \cdots + \alpha_rL^{k-1}(y_r) = 0$$

then

$$L^{k-1}(\alpha_1 y_1 + \cdots + \alpha_r y_r) = 0$$

and hence $\alpha_1 y_1 + \cdots + \alpha_r y_r \in \ker(L^{k-1})$. But then $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$ otherwise $v_1, \ldots, v_m, y_1, \ldots, y_r$ would be dependent. Thus $L^{k-1}(y_1), \ldots, L^{k-1}(y_r)$ 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

$$S_1 = C_L(y_1) \oplus \cdots \oplus C_L(y_r)$$

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 $s \in S_2$ then it must be of the form $s = \alpha_1 v_1 + \cdots + \alpha_m v_m + 0y_1 + \cdots + 0y_r$.) 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 ($1 \leq k_i \leq k$ 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

$$V = C_L(v_1) \oplus \cdots \oplus C_L(v_s)$$

If $C_L(v_i)$ has dimension $k_i$ then the matrix representing $L_{[C_L(v_i)]}$ with respect to $[L^{k_{i-1}}(v_i), \ldots, v_i]$ 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 $\lambda$ then $L$ can be represented by a matrix $A$ of the form

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

(9)
where each $J_i(\lambda)$ is a bidiagonal matrix of the form

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

Proof. Let $I$ denote the identity operator on $V$. The eigenvalues of the operator $L - \lambda I$ are all 0 and hence $L - \lambda I$ is nilpotent. It follows from Corollary 9.1.6 that with respect to some ordered basis $[v_1, \ldots, v_n]$ of $V$ the operator $L - \lambda I$ can be represented by a matrix of the form

$$J = \begin{pmatrix}
J_1(0) \\
J_2(0) \\
& \ddots \\
J_s(0)
\end{pmatrix}$$

where

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

The matrix representing $\lambda I$ with respect to $[v_1, \ldots, v_n]$ is simply $\lambda I$. Since $L = (L - \lambda I) + \lambda I$ it follows that the matrix representing $L$ with respect to $[v_1, \ldots, v_n]$ 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 $\lambda$ 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 $\mathbb{R}^5$ into $\mathbb{R}^5$. 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:

\[
\begin{pmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{pmatrix}
\text{ or }
\begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

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

\[
A - I = \begin{pmatrix}
0 & 2 & 1 & 1 & 1 \\
0 & 0 & 1 & 2 & 1 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\quad (A - I)^2 = \begin{pmatrix}
0 & 0 & 2 & 5 & 2 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

\[
(A - I)^3 = \begin{pmatrix}
0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\quad (A - I)^4 = O
\]

Thus \( A - I \) is nilpotent of index 4. The systems

\[
(A - I)^k \mathbf{s} = \mathbf{x} \quad \text{and} \quad (A - I)^j \mathbf{s} = \mathbf{y}
\]

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

\[
\mathbf{x}_1 = (0, 0, 0, \frac{1}{2}, 0)^T
\]

To generate the rest of the cyclic subspace we compute

\[
\mathbf{x}_2 = (A - I)\mathbf{x}_1 = (\frac{1}{2}, 1, \frac{1}{2}, 0, 0)^T
\]

\[
\mathbf{x}_3 = (A - I)\mathbf{x}_2 = (A - I)^2\mathbf{x}_1 = (\frac{5}{2}, \frac{1}{2}, 0, 0, 0)^T
\]

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
\[(A - I)^j s = y\]
are inconsistent for all positive integers \(j\). Thus the cyclic subspace containing \(y\) has dimension 1. It follows that the matrix representing \(A\) with respect to \([x, x_3, x_2, x_1, y]\) is
\[
J = \begin{pmatrix}
J_1(1) \\
J_2(1)
\end{pmatrix}
= \begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

The reader may verify that if \(Y\) is the matrix whose columns are \(x, x_3, x_2, x_1, y\), 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
\[
J = \begin{pmatrix}
B_1 & B_2 & \cdots & B_m
\end{pmatrix}
\]
where each \(B_i\) is of the form (9) with diagonal elements equal to \(\lambda_i\), 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 \(\times\) 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 \( \lambda \) and the eigenspace \( S_\lambda \) 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.

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

(b) \[
A = \begin{pmatrix}
1 & 2 & 0 & 0 \\
0 & 1 & 2 & 0 \\
0 & 0 & 1 & 2 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

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

(a) \[
A = \begin{pmatrix}
-1 & 1 & 0 & 0 \\
-1 & 1 & 0 & 0 \\
-2 & 2 & 0 & 0 \\
0 & 3 & -1 & 0
\end{pmatrix}
\]

(b) \[
A = \begin{pmatrix}
0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

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 \( S_1 \cap S_2 = \{0\} \).


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

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

10. Let \( L \) be a linear operator on a vector space \( V \) and let \( S \) be a subspace of \( V \). Let \( I \) represent the identity operator and let \( \lambda \) be a scalar. Show that \( L \) is invariant on \( S \) if and only if \( L - \lambda I \) 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 \( \ker(L^i) \) 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

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

\[\square\]

**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

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

14
Thus \( v \in \ker(L^{2k_0}) = \ker(L^{k_0}) \) and hence
\[
w = L^{k_0}(v) = 0
\]

Therefore \( X \cap Y = \{0\} \). We claim \( V = X \oplus Y \). Let \( \{x_1, \ldots, x_r\} \) be a basis for \( X \) and let \( \{y_1, \ldots, y_{n-r}\} \) be a basis for \( Y \). By Lemma 9.2.1 it suffices to show that \( x_1, \ldots, x_r, y_1, \ldots, y_{n-r} \) are linearly independent and hence form a basis for \( V \). If
\[
(1) \quad \alpha_1 x_1 + \cdots + \alpha_r x_r + \beta_1 y_1 + \cdots + \beta_{n-r} y_{n-r} = 0
\]
then applying \( L^{k_0} \) to both sides gives
\[
\beta_1 L^{k_0}(y_1) + \cdots + \beta_{n-r} L^{k_0}(y_{n-r}) = 0
\]
or
\[
L^{k_0}(\beta_1 y_1 + \cdots + \beta_{n-r} y_{n-r}) = 0
\]
Therefore \( \beta_1 y_1 + \cdots + \beta_{n-r} y_{n-r} \in X \cap Y \) and hence
\[
\beta_1 y_1 + \cdots + \beta_{n-r} y_{n-r} = 0
\]
Since the \( y_i \)'s are linearly independent it follows that
\[
\beta_1 = \beta_2 = \cdots = \beta_{n-r} = 0
\]
and hence (1) simplifies to
\[
\alpha_1 x_1 + \cdots + \alpha_r x_r = 0
\]
Since the \( x_i \)'s are linearly independent it follows that
\[
\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0
\]
Thus, \( x_1, \ldots, x_r, y_1, \ldots, y_{n-r} \) 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 \( y \in Y \), then \( y = L^{k_0}(v) \) for some \( v \in V \). Thus,
\[
L(y) = L(L^{k_0}(v)) = L^{k_0+1}(v) = L^{k_0}(L(v))
\]
Therefore \( L(y) \in Y \) and hence \( Y \) is invariant under \( L \). To prove \( L_{|Y} \) is invertible it suffices to show that
\[
\ker(L_{|Y}) = Y \cap \ker(L) = \{0\}
\]
This, however, follows immediately since \( \ker(L) \subset X \) and \( X \cap Y = \{0\} \). \( \square \)

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 $L - \lambda I$ is nilpotent on $X_i$ and the dimension of $X_i$ equals the multiplicity of $\lambda_i$.

Proof. Let $L_1 = L - \lambda_1 I$. 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 $L_1|_{Y_1}$ 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 $\lambda_1$. 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 $\lambda_1$ is not an eigenvalue of $B_1$. Thus

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

where $q(\lambda) \neq 0$. 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 $\lambda_i$ of $L$ has multiplicity $r_i$, then

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

Therefore $r_1 = m_1$ and

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

If we consider the operator $L_2 = L - \lambda_2 I$ 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 $L_2|_{Y_2}$ 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 \( \lambda_k \). Thus, if we set \( X_k = Y_{k-1} \) then \( L - \lambda_k I \) 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 \( \lambda_i \)) whose blocks consist of simple Jordan matrices with \( \lambda_i \)'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

\[
L_A(x) = Ax \quad \text{for each} \ x \in R^n
\]

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

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

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

\[
A_i = \begin{pmatrix}
J_1(\lambda_i) & & \\
& J_2(\lambda_i) & \\
& & \ddots & \\
& & & J_s(\lambda_i)
\end{pmatrix}
\]

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 \( x_1 = (1, 1, 1, 1, 2)^T \) and the eigenspace corresponding to \( \lambda = 0 \) is spanned by \( x_2 = (1, 1, 0, 1, 1)^T \) and \( x_3 = (0, 0, 1, 0, 0)^T \). 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:
\[
\begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 1 \\
0 & 1
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\]
To determine which of these forms is correct we compute \((A - 0I)^2 = A^2\).
\[
A^2 = \begin{pmatrix}
-1 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 0 & 1 \\
-2 & 0 & 0 & 0 & 2
\end{pmatrix}
\]
Next we consider the systems \( A^2x = x_i \) for \( i = 2, 3 \). Since these systems turn out to be inconsistent, the Jordan canonical form of \( A \) cannot have any \( 3 \times 3 \) simple Jordan blocks and consequently it must be of the form
\[
J = X^{-1}AX = \begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 0
\end{pmatrix}
\]
To find \( X \) we must solve \( Ax = x_i \) for \( i = 2, 3 \). The system, \( Ax = x_2 \), has infinitely many solutions. We need choose only one of these say \( x_4 = (1, 3, 0, 0, 1)^T \). Similarly \( Ax = x_3 \) has infinitely many solutions one of which is \( x_5 = (1, 0, 0, 2, 1)^T \). Let
\[
X = \begin{pmatrix}
x_1 & x_2 & x_3 & x_4 & x_5
\end{pmatrix} = \begin{pmatrix}
1 & 1 & 1 & 0 & 1 \\
1 & 1 & 3 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 2 \\
2 & 1 & 1 & 0 & 1
\end{pmatrix}
\]
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

$$Y'(t) = AY(t)$$

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

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

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

$$XZ' = XJZ$$

Multiplying by $X^{-1}$ we get

(4) \[ Z' = JZ \]

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

$$z'_1 = \lambda z_1 + z_2$$

$$z'_2 = \lambda z_2 + z_3$$

$$\vdots$$

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

$$z'_k = \lambda z_k$$

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

\[
\begin{pmatrix}
  y'_1 \\
  y'_2 \\
  y'_3 \\
  y'_4
\end{pmatrix} = \begin{pmatrix}
  1 & 0 & 0 & -1 \\
  0 & 1 & 1 & 0 \\
  0 & -1 & 1 & 2 \\
  1 & 0 & 2 & 1
\end{pmatrix} \begin{pmatrix}
  y_1 \\
  y_2 \\
  y_3 \\
  y_4
\end{pmatrix}
\]

\[y_1(0) = y_2(0) = y_3(0) = 0, \quad y_4(0) = 2\]

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

\[
x'_1 = x_2 \\
x'_2 = 0 \\
x'_3 = 2x_3 + x_4 \\
x'_4 = 2x_4
\]

The first system is not difficult to solve.

\[
x_1 = c_1t + c_2 \\
x_2 = c_1 \\
(c_1 \text{ and } c_2 \text{ are constants})
\]

To solve the second system we solve first

\[x'_4 = 2x_4\]

getting

\[x_4 = c_3e^{2t}\]

Thus

\[x'_3 - 2x_3 = c_3e^{2t}\]
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
\[ y_1 = -t - te^{2t} \]
\[ y_2 = -t + te^{2t} \]
\[ y_3 = -1 + t + (1 + t)e^{2t} \]
\[ y_4 = 1 - t + (1 + t)e^{2t} \]

The Jordan canonical form not only provides a nice representation of an operator but it allows us to solve systems of the form \( Y' = AY \) 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 \( \lambda_i \)'s are only approximations to the actual eigenvalues. Thus we could have calculated values \( \lambda_1' \) and \( \lambda_2' \) 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 \( Y' = AY \) 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 \( Y' = AY \).

**Exercises**

1. Let \( A \) be a \( 4 \times 4 \) matrix whose only eigenvalue is \( \lambda = 2 \). What are the possible Jordan canonical forms for \( A \)?
2. Let \( A \) be a \( 5 \times 5 \) 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) $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

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

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{align*}
y'_1 &= 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{align*}

7. Suppose
\[ X^{-1}AX = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} = J \]

If \( x_1, x_2, \) and \( x_3 \) are the column vectors of \( X \) define
\begin{align*}
z_1 &= ax_1 \\
z_2 &= ax_2 + bx_1 \\
z_3 &= ax_3 + bx_2 + cx_3
\end{align*}

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 \( BJB^{-1} = X^{-1}AX = J \).