

## Chapter 8

# Jordan Normal Form

### 8.1 Minimal Polynomials

Recall  $p_A(x) = \det(xI - A)$  is called the characteristic polynomial of the matrix  $A$ .

**Theorem 8.1.1.** *Let  $A \in M_n$ . Then there exists a unique monic polynomial  $q_A(x)$  of minimum degree for which  $q_A(A) = 0$ . If  $p(x)$  is any polynomial such that  $p(A) = 0$ , then  $q_A(x)$  divides  $p(x)$ .*

*Proof.* Since there is a polynomial  $p_A(x)$  for which  $p_A(A) = 0$ , there is one of minimal degree, which we can assume is monic. by the Euclidean algorithm

$$p_A(x) = q_A(x)h(x) + r(x)$$

where  $\deg r(x) < \deg q_A(x)$ . We know

$$p_A(A) = q_A(A)h(A) + r(A).$$

Hence  $r(A) = 0$ , and by the minimality assumption  $r(x) \equiv 0$ . Thus  $q_A$  divided  $p_A(x)$  and also any polynomial for which  $p(A) = 0$ . to establish that  $q_A$  is unique, suppose  $q(x)$  is another monic polynomial of the same degree for which  $q(A) = 0$ . Then

$$r(x) = q(x) - q_A(x)$$

is a polynomial of degree less than  $q_A(x)$  for which  $r(A) = q(A) - q_A(A) = 0$ . This cannot be unless  $r(x) \equiv 0 = 0q = q_A$ .  $\square$

**Definition 8.1.1.** The polynomial  $q_A(x)$  in the theorem above is called the minimal polynomial.

**Corollary 8.1.1.** *If  $A, B \in M_n$  are similar, then they have the same minimal polynomial.*

*Proof.*

$$\begin{aligned} B &= S^{-1}AS \\ q_A(B) &= q_A(S^{-1}AS) = S^{-1}q_A(A)S = q_A(A) = 0. \end{aligned}$$

If there is a minimal polynomial for  $B$  of smaller degree, say  $q_B(x)$ , then  $q_B(A) = 0$  by the same argument. This contradicts the minimality of  $q_A(x)$ .  $\square$

Now that we have a minimum polynomial for any matrix, can we find a matrix with a given polynomial as its minimum polynomial? Can the degree of the polynomial and the size of the matrix match? The answers to both questions are affirmative and presented below in one theorem.

**Theorem 8.1.2.** *For any  $n^{\text{th}}$  degree polynomial*

$$p(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0$$

*there is a matrix  $A \in M_n(\mathbb{C})$  for which it is the minimal polynomial.*

*Proof.* Consider the matrix given by

$$A = \begin{bmatrix} 0 & 0 & \cdots & \cdots & -a_0 \\ 1 & 0 & & & -a_1 \\ 0 & 1 & 0 & \ddots & -a_2 \\ & \ddots & & 0 & \vdots \\ 0 & \cdots & & 0 & 1 & -a_{n-1} \end{bmatrix}.$$

Observe that

$$\begin{aligned} Ae_1 &= e_1 = A^0e_1 \\ Ae_1 &= e_2 = Ae_1 \\ Ae_2 &= e_3 = A^2e_1 \\ &\vdots \\ Ae_{n-1} &= e_n = A^{n-1}e_1 \end{aligned}$$

and

$$Ae_n = -a_{n-1}e_n - a_{n-2}e_{n-1} - \cdots - a_1e_2 - a_0e_1$$

Since  $Ae_n = A^n e_1$ , it follows that

$$p(A)e_1 = A^n e_1 + a_{n-1}A^{n-1}e_1 + a_{n-2}A^{n-2}e_1 + \cdots + a_1Ae_1 + a_0Ie_1 = 0$$

Also

$$p(A)e_k = p(A)A^{k-1}e_1 = A^{k-1}p(A)e_1 = A^{k-1}(0) = 0 \quad k = 2, \dots, n.$$

Hence  $p(A)e_j = 0$  for  $j = 1 \dots n$ . Thus  $p(A) = 0$ . We know also that  $p(x)$  is monic. Suppose now that

$$q(x) = x^m + b_{m-1}x^{m-1} \cdots + b_1x + b_0$$

where  $m < n$  and  $q(A) = 0$ . Then

$$\begin{aligned} q(A)e_1 &= A^m e_1 + b_{m-1}A^{m-1}e_1 + \cdots + b_1Ae_1 + b_0e_1 \\ &= e_{m+1} + b_{m-1}e_m + \cdots + b_1e_2 + b_0e_1 = 0. \end{aligned}$$

But the vectors  $e_{m+1} \dots e_1$  are linear independent from which we conclude that  $q(A) = 0$  is impossible. Thus  $p(x)$  is minimal.  $\square$

**Definition 8.1.2.** For a given monic polynomial  $p(x)$ , the matrix  $A$  constructed above is called the *companion* matrix to  $p$ .

The transpose of the companion matrix can also be used to generate a linear differential system which has the same characteristic polynomial as a given  $n^{\text{th}}$  order differential equation. Consider the linear differential equation

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1y' + a_0 = 0.$$

This  $n^{\text{th}}$  order ODE can be converted to a first order system as follows:

$$\begin{aligned} u_1 &= y \\ u_2 &= u_1' &= y' \\ u_3 &= u_2' &= y'' \\ &\vdots &\vdots \\ u_n &= u_{n-1}' &= y^{(n-1)} \end{aligned}$$

Then we have

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}' = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & 0 \\ & & 0 & 1 & \\ & & & \ddots & \ddots \\ & 0 & & & \ddots & 1 \\ a_0 & -a_1 & \dots & \dots & -a_{n-1} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

## 8.2 Invariant subspaces

There seems to be no truly simple way to the Jordan normal form. The approach taken here is intended to reveal a number of features of a matrix, interesting in their own right. In particular, we will construct “generalized eigenspaces” that envelop the entire connection of a matrix with its eigenvalues. We have in various ways considered subspaces  $V$  of  $\mathbb{C}_n$  that are invariant under the matrix  $A \in M_n(\mathbb{C})$ . Recall this means that  $AV \subset V$ . For example, eigenvectors can be used to create invariant subspaces. Null spaces, the eigenspace of the zero eigenvalue, are invariant as well. Triangular matrices furnish an easily recognizable sequence of invariant subspaces. Assuming  $T \in M_n(\mathbb{C})$  is upper triangular, it is easy to see that the subspaces generated by the coordinate vectors  $\{e_1, \dots, e_m\}$  for  $m = 1, \dots, n$  are invariant under  $T$ .

We now consider a specific type of invariant subspace that will lead the so-called Jordan normal form of a matrix, the closest matrix similar to  $A$  that resembles a diagonal matrix.

**Definition 8.2.1 (Generalized Eigenspace).** Let  $A \in M_n(\mathbb{C})$  with spectrum  $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ . Define the *generalized eigenspace* pertaining to  $\lambda_i$  by

$$V_{\lambda_i} = \{x \in \mathbb{C}_n \mid (A - \lambda_i I)^n x = 0\}$$

Observe that all the eigenvectors pertaining to  $\lambda_i$  are contained in  $V_{\lambda_i}$ . If the span of the eigenvectors pertaining to  $\lambda_i$  is not equal to  $V_{\lambda_i}$  then there must be a positive power  $p$  and a vector  $x$  such that  $(A - \lambda_i I)^p x = 0$  but that  $y = (A - \lambda_i I)^{p-1} x \neq 0$ . Thus  $y$  is an eigenvector pertaining to  $\lambda_i$ . For this reason we will call  $V_{\lambda_i}$  the space of **generalized eigenvectors** pertaining to  $\lambda_i$ . Our first result, that  $V_{\lambda_i}$  is invariant under  $A$ , is simple to prove, noting that only closure under vector addition and scalar multiplication need be established.

**Theorem 8.2.1.** *Let  $A \in M_n(\mathbb{C})$  with spectrum  $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ . Then for each  $i = 1, \dots, k$ ,  $V_{\lambda_i}$  is an invariant subspace of  $A$ .*

One might question as to whether  $V_{\lambda_i}$  could be enlarged by allowing higher powers than  $n$  in the definition. The negative answer is most simply expressed by evoking the Hamilton-Cayley theorem. We write the characteristic polynomial  $p_A(\lambda) = \prod (\lambda - \lambda_i)^{m_A(\lambda_i)}$ , where  $m_A(\lambda_i)$  is the algebraic multiplicity of  $\lambda_i$ . Since  $p_A(A) = \prod (A - \lambda_i I)^{m_A(\lambda_i)} = 0$ , it is an easy matter to see that we exhaust all of  $\mathbb{C}^n$  with the spaces  $V_{\lambda_i}$ . This is to say that allowing higher powers in the definition will not increase the subspaces  $V_{\lambda_i}$ . Indeed, as we shall see, the power of  $(\lambda - \lambda_i)$  can be decreased to the geometric multiplicity  $m_g(\lambda_i)$  the power for  $\lambda_i$ . For now the general power  $n$  will suffice. One very important result, and an essential first step in deriving the Jordan form, is to establish that any square matrix  $A$  is similar to a block diagonal matrix, with each block carrying a single eigenvalue.

**Theorem 8.2.2.** *Let  $A \in M_n(\mathbb{C})$  with spectrum  $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$  and with invariant subspaces  $V_{\lambda_i}$ ,  $i = 1, 2, \dots, k$ . Then (i) The spaces  $V_{\lambda_i}$ ,  $j = 1, \dots, k$  are mutually linearly independent. (ii)  $\bigoplus_{i=1}^k V_{\lambda_i} = \mathbb{C}^n$  (alternatively  $\mathbb{C}^n = \mathfrak{S}(V_{\lambda_1}, \dots, V_{\lambda_k})$ ) (iii)  $\dim V_{\lambda_i} = m_A(\lambda_i)$ . (iv)  $A$  is similar to a block diagonal matrix with  $k$  blocks  $A_1, \dots, A_k$ . Moreover,  $\sigma(A_i) = \{\lambda_i\}$  and  $\dim A_i = m_A(\lambda_i)$ .*

*Proof.* (i) It should be clear that the subspaces  $V_{\lambda_i}$  are linearly independent of each other. For if there is a vector  $x$  in both  $V_{\lambda_i}$  and  $V_{\lambda_j}$  then there is a vector for some integer  $q$ , it must be true that  $(A - \lambda_j I)^{q-1} x \neq 0$  but  $(A - \lambda_j I)^q x = 0$ . This means that  $y = (A - \lambda_j I)^{q-1} x$  is an eigenvector pertaining to  $\lambda_j$ . Since  $(A - \lambda_i I)^n x = 0$  we must also have that

$$\begin{aligned} (A - \lambda_j I)^{q-1} (A - \lambda_i I)^n x &= (A - \lambda_i I)^n (A - \lambda_j I)^{q-1} x \\ &= (A - \lambda_i I)^n y = 0 \\ &= \sum_{k=0}^n \binom{n}{k} (-\lambda_i)^{n-k} A^k y \\ &= \sum_{k=0}^n \binom{n}{k} (-\lambda_i)^{n-k} \lambda_j^k y \\ &= (\lambda_j - \lambda_i)^n y = 0 \end{aligned}$$

This is impossible unless  $\lambda_j = \lambda_i$ . (ii) The key part of the proof is to *block* diagonalize  $A$  with respect to these invariant subspaces. To that end, let  $S$

be the matrix with columns generated from bases of the individual  $V_{\lambda_i}$  taken in the order of the indices. Supposing there are more linearly independent vectors in  $\mathbb{C}_n$  other than those already selected, fill out the matrix  $S$  with vectors linearly independent to the subspaces  $V_{\lambda_i}$ ,  $i = 1, \dots, k$ . Now define  $\tilde{A} = S^{-1}AS$ . We conclude by the invariance of the subspaces and their mutual linear independence that  $\tilde{A}$  has the following block structure.

$$\tilde{A} = S^{-1}AS = \begin{bmatrix} A_1 & 0 & \cdots & 0 & * \\ 0 & A_2 & & 0 & * \\ \vdots & & \ddots & & \vdots \\ & & & A_k & * \\ 0 & \cdots & & 0 & B \end{bmatrix}$$

It follows that

$$p_A(\lambda) = p_{\tilde{A}}(\lambda) = \left( \prod p_{A_i}(\lambda) \right) p_B(\lambda)$$

Any root  $r$  of  $p_B(\lambda)$  must be an eigenvalue of  $A$ , say  $\lambda_j$ , and there must be an eigenvector  $x$  pertaining to  $\lambda_j$ . Moreover, due to the block structure we can assume that  $x = [0, \dots, 0, x]^T$ , where there are  $k$  blocked zeros of the sizes of the  $A_i$  respectively. Then it is easy to see that  $ASx = \lambda_j Sx$ , and this implies that  $Sx \in V_{\lambda_j}$ . Thus there is another vector in  $V_{\lambda_j}$ , which contradicts its definition. Therefore  $B$  is null, or what is the same thing,  $\bigoplus_{i=1}^k V_{\lambda_i} = \mathbb{C}_n$ .  
 (iii) Let  $d_i = \dim V_{\lambda_i}$ . From (ii) know  $\sum d_i = n$ . Suppose that  $\lambda_i \in \sigma(A_j)$ . Then there is another eigenvector  $x$  pertaining to  $\lambda_i$  and for which  $A_j x = \lambda_i x$ . Moreover, this vector has the form  $x = [0, \dots, 0, x, 0, \dots, 0]^T$ , analogous to the argument above. By construction  $Sx \notin V_{\lambda_i}$ , but  $ASx = \lambda_i Sx$ , and this contradicts the definition of  $V_{\lambda_i}$ . We thus have that  $p_{A_i}(\lambda) = (\lambda - \lambda_i)^{d_i}$ . Since  $p_A(\lambda) = \prod p_{A_i}(\lambda) = \prod (\lambda - \lambda_i)^{m_A(\lambda_i)}$ , it follows that  $d_i = m_A(\lambda_i)$ .  
 (iv) Putting (ii), and (iii) together gives the block diagonal structure as required.  $\square$

On account of the mutual linear independence of the invariant subspaces  $V_{\lambda_i}$  and the fact that they exhaust  $\mathbb{C}_n$  the following corollary is immediate.

**Corollary 8.2.1.** *Let  $A \in M_n(\mathbb{C})$  with spectrum  $\sigma(A) = \lambda_1, \dots, \lambda_k$  and with generalized eigenspaces  $V_{\lambda_i}$ ,  $i = 1, 2, \dots, k$ . Then each  $x \in \mathbb{C}_n$  has a unique representation  $x = \sum_{i=1}^k x_i$  where  $x_i \in V_{\lambda_i}$ .*

Another interesting result which reveals how the matrix works as a linear transformation is to decompose the it into components with respect to the

generalized eigenspaces. In particular, viewing the block diagonal form

$$\tilde{A} = S^{-1}AS = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & & 0 \\ \vdots & & \ddots & \\ & & & A_k \end{bmatrix}$$

the space  $\mathbb{C}_n$  can be split into a direct sum of subspaces  $E_1, \dots, E_k$  based on coordinate blocks. This is accomplished in such that any vector  $y \in \mathbb{C}_n$  can be written uniquely as  $y = \sum_{i=1}^k y_i$  where the  $y_i \in E_i$ . (Keep in mind that each  $y_i \in \mathbb{C}_n$ ; its coordinates are zero outside the coordinate block pertaining  $E_i$ .) Then  $\tilde{A}y = \tilde{A} \sum_{i=1}^k y_i = \sum_{i=1}^k \tilde{A}y_i = \sum_{i=1}^k A_i y_i$ . This provides a computational tool — when this block diagonal form is known. Note that the blocks correspond directly to the invariant subspaces by  $SE_i = V_{\lambda_i}$ . We can use these invariant subspaces to get at the minimal polynomial. For each  $i = 1, \dots, k$  define

$$m_i = \min_j \{(A - \lambda_i I)^j x = 0 \mid x \in V_{\lambda_i}\}$$

**Theorem 8.2.3.** *Let  $A \in M_n(\mathbb{C})$  with spectrum  $\sigma(A) = \lambda_1, \dots, \lambda_k$  and with invariant subspaces  $V_{\lambda_i}$ ,  $i = 1, 2, \dots, k$ . Then the minimal polynomial of  $A$  is given by*

$$q(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{m_i}$$

*Proof.* Certainly we see that for any vector  $x \in V_{\lambda_j}$

$$q(A)x = \left( \prod_{i=1}^k (A - \lambda_i I)^{m_i} \right) x = 0$$

Hence, the minimal polynomial  $q_A(\lambda)$  divides  $q(A)$ . To see that indeed they are in fact equal, suppose that the minimal polynomial has the form

$$q_A(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{\hat{m}_i}$$

where  $\hat{m}_i \leq m_i$ , for  $i = 1, \dots, k$  and in particular  $\hat{m}_j < m_j$ . By construction there must exist a vector  $x \in V_{\lambda_j}$  such that  $(A - \lambda_j I)^{m_j} x = 0$  but  $y =$

$(A - \lambda_j I)^{m_j - 1} x \neq 0$ . Then if

$$\begin{aligned} q(A)x &= \left( \prod_{i=1}^k (A - \lambda_i I)^{\hat{m}_i} \right) x \\ &= \left( \prod_{\substack{i=1 \\ i \neq j}}^k (A - \lambda_i I)^{\hat{m}_i} \right) y \\ &= 0 \end{aligned}$$

This cannot be because the contrary implies that there is another vector in one of the invariant subspaces  $V_{\lambda_k}$ .  $\square$

Just one more step is needed before the Jordan normal form can be derived. For a given  $V_{\lambda_i}$  we can interpret the spaces in a hierarchical viewpoint. We know that  $V_{\lambda_i}$  contains all the eigenvectors pertaining to  $\lambda_i$ . Call these eigenvectors the *first order generalized eigenvectors*. If the span of these is not equal to  $V_{\lambda_i}$ , then there must be a vector  $x \in V_{\lambda_i}$  for which  $y = (A - \lambda_i I)^2 x = 0$  but  $(A - \lambda_i I)x \neq 0$ . That is to say  $y$  is an eigenvector of  $A$  pertaining to  $\lambda_i$ . Call such vectors *second order generalized eigenvectors*. In general we call an  $x \in V_{\lambda_i}$  a generalized eigenvector of order  $j$  if  $y = (A - \lambda_i I)^j x = 0$  but  $(A - \lambda_i I)^{j-1} x \neq 0$ . In light of our previous discussion  $V_{\lambda_i}$  contains generalized eigenvectors of order up to but not greater than  $m_{\lambda_i}$ .

**Theorem 8.2.4.** *Let  $A \in M_n(\mathbb{C})$  with spectrum  $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$  and with invariant subspaces  $V_{\lambda_i}$ ,  $i = 1, 2, \dots, k$ .*

(i) *Let  $x \in V_{\lambda_i}$  be a generalized eigenvector of order  $p$ . Then the vectors*

$$x, (A - \lambda_i I)x, (A - \lambda_i I)^2 x, \dots, (A - \lambda_i I)^{p-1} x \quad (1)$$

*are linearly independent.*

(ii) *The subspace of  $\mathbb{C}_n$  generated by the vectors in (1) is an invariant subspace of  $A$ .*

*Proof.* (i) To prove linear independence of a set of vectors we suppose linear dependence. That is there is a smallest integer  $k$  and constants  $b_j$  such that

$$\sum_{j=0}^k x_j = \sum_{j=0}^k b_j (A - \lambda_i I)^j x = 0$$

where  $b_k \neq 0$ . Solving we obtain  $b_k (A - \lambda_i I)^k x = -\sum_{j=0}^{k-1} b_j (A - \lambda_i I)^j x$ . Now apply  $(A - \lambda_i I)^{p-k}$  to both sides and obtain a new linearly dependent set as the following calculation shows.

$$\begin{aligned} 0 &= b_k (A - \lambda_i I)^{p-k+k} x = -\sum_{j=0}^{k-1} b_j (A - \lambda_i I)^{j+p-k} x \\ &= -\sum_{j=p-k}^{p-1} b_{j+p-k} (A - \lambda_i I)^j x \end{aligned}$$

The key point to note here is that the lower limit of the sum is increased. This new linearly dependent set, which we denote with the notation  $\sum_{j=p-k}^{p-1} c_j (A - \lambda_i I)^j x$  can be split in the same way as before, where we assume with no loss in generality that  $c_{p-1} \neq 0$ . Then

$$c_{p-1} (A - \lambda_i I)^{p-1} x = -\sum_{j=p-k}^{p-2} c_j (A - \lambda_i I)^j x$$

Apply  $(A - \lambda_i I)$  to both sides to get

$$\begin{aligned} 0 &= c_{p-1} (A - \lambda_i I)^p x = -\sum_{j=p-k}^{p-2} c_j (A - \lambda_i I)^{j+1} x \\ &= -\sum_{j=p-k+1}^{p-1} c_{j-1} (A - \lambda_i I)^j x \end{aligned}$$

Thus we have obtained another linearly independent set with the lower limit of powers increased by one. Continue this process until the linear dependence of  $(A - \lambda_i I)^{p-1} x$  and  $(A - \lambda_i I)^{p-2} x$  is achieved. Thus we have

$$\begin{aligned} c (A - \lambda_i I)^{p-1} x &= d (A - \lambda_i I)^{p-2} x \\ (A - \lambda_i I) y &= \frac{d}{c} y \end{aligned}$$

where  $y = (A - \lambda_i I)^{p-2} x$ . This implies that  $\lambda_i + \frac{d}{c}$  is a new eigenvalue with eigenvector  $y \in V_{\lambda_i}$ , and of course this is a contradiction. (ii) The invariance under  $A$  is more straightforward. First note that while  $x_1 = x$ ,

$x_2 = (A - \lambda I)x = Ax - \lambda x$  so that  $Ax = x_2 + \lambda x_1$ . Consider any vector  $y$  defined by  $y = \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j x$ . It follows that

$$\begin{aligned}
 Ay &= A \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j x \\
 &= \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j Ax \\
 &= \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j (x_2 + \lambda x_1) \\
 &= \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j [(A - \lambda_i I)x_1 + \lambda x_1] \\
 &= \sum_{j=0}^{p-1} c_j (A - \lambda_i I)^j x
 \end{aligned}$$

where  $c_j = b_{j-1} - \lambda$  for  $j > 0$  and  $c_0 = -\lambda$ , which proves the result.  $\square$

### 8.3 The Jordan Normal Form

We need a lemma that points in the direction we are headed, that being the use of invariant subspaces as a basis for the (Jordan) block diagonalization of any matrix. These results were discussed in detail in the Section 8.2. The restatement here illustrates the “invariant subspace” nature of the result, irrespective of generalized eigenspaces. Its proof is elementary and is left to the reader.

**Lemma 8.3.1.** *Let  $A \in M_n(\mathbb{C})$  with invariant subspace  $V \subset \mathbb{C}^n$ .*

(i) *Suppose  $v_1 \dots v_k$  is a basis for  $V$  and  $S$  is an invertible matrix with the first  $k$  columns given by  $v_1 \dots v_k$ . Then*

$$S^{-1}AS = \left[ \begin{array}{c|c} & \\ \hline * & * \\ \hline 0 & * \end{array} \right]^{1..k}$$

(ii) Suppose that  $V_1, V_2 \subset \mathbb{C}^n$  are two invariant subspaces of  $A$  and  $\mathbb{C}^n = V_1 \oplus V_2$ . Let the (invertible) matrix  $S$  consist respectively of bases from  $V_1$  and  $V_2$  as its columns. Then

$$S^{-1}AS = \left[ \begin{array}{c|c} * & 0 \\ \hline 0 & * \end{array} \right].$$

**Definition 8.3.1.** Let  $\lambda \in \mathbb{C}$ . A Jordan block  $J_k(\lambda)$  is a  $k \times k$  upper triangular matrix of the form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ 0 & & \ddots & 1 \\ & & & \lambda \end{bmatrix}.$$

A Jordan matrix is any matrix of the form

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_k}(\lambda_k) \end{bmatrix}.$$

where the matrices  $J_{n_i}$  are Jordan blocks. If  $J \in M_n(\mathbb{C})$ , then  $n_1 + n_2 + \cdots + n_k = n$ .

**Theorem 8.3.1 (Jordan normal form).** Let  $A \in M_n(\mathbb{C})$ . Then there is a nonsingular matrix  $S \in M_n$  such that

$$A = S \begin{bmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_k}(\lambda_k) \end{bmatrix} S^{-1} = SJS^{-1}$$

where  $J_{n_i}(\lambda_i)$  is a Jordan block, where  $n_1 + n_2 + \cdots + n_k = n$ .  $J$  is unique up to permutations of the blocks. The eigenvalues  $\lambda_1, \dots, \lambda_k$  are not necessarily distinct. If  $A$  is real with real eigenvalues, then  $S$  can be taken as real.

*Proof.* This result is proved in four steps.

- (1) Block diagonalize (by similarity) into invariant subspaces pertaining to  $\sigma(A)$ . This is accomplished as follows. First block diagonalize the matrix according to the generalized eigenspaces  $V_{\lambda_i} = \{x \in \mathbb{C}^n \mid (A - \lambda_i I)^n x =$

0} as discussed in the previous section. Beginning with the highest order eigenvector in  $x \in V_{\lambda_i}$ , construct the invariant subspace as in (1) of Section 8.2. Repeat this process until all generalized eigenvectors have been included in an invariant subspace. This includes of course first order eigenvectors that are not associated with higher order eigenvectors. These invariant subspaces have dimension one. Each of these invariant subspaces is linearly independent from the others. Continue this process for all the generalized eigenspaces. This exhausts  $\mathbb{C}^n$ . Each of the blocks contains exactly one eigenvector. The dimensions of these invariant subspaces can range from one to  $m_{\lambda_i}$ , there being at least one subspace of dimension  $m_{\lambda_i}$ .

- (2) Triangularize each block by Schur's theorem, so that each block has the form

$$K(\lambda) = \begin{bmatrix} \lambda & & * \\ & \ddots & \\ 0 & & \lambda \end{bmatrix}$$

You will note that

$$K(\lambda) = \lambda I + N$$

where  $N$  is nilpotent, or  $K(\lambda)$  is  $1 \times 1$ .

- (3) "Jordanize" each triangular block. Assume that  $K_1(\lambda)$  is  $m \times m$ , where  $m > 1$ . By construction  $K_1(\lambda)$  pertains to an invariant subspace for which there is a unique vector  $x$  for which

$$N^{m-1}x \neq 0 \quad \text{and} \quad N^m x = 0.$$

Thus  $N^{m-1}x$  is an eigenvector of  $K_1(\lambda)$ , the unique eigenvector. Define

$$y_i = N^{i-1}x \quad i = 1, 2, \dots, m.$$

Expand the set  $\{y_i\}_{i=1}^m$  as a basis of  $\mathbb{C}_m$ . Define

$$S_1 = \begin{bmatrix} y_m & y_{m-1} & \cdots & y_1 \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}.$$

Then

$$NS_1 = \begin{bmatrix} 0 & y_m & y_{m-1} & \cdots & y_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}.$$

So

$$S_1^{-1}NS_1 = \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & 0 & \ddots \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{bmatrix}.$$

We conclude that

$$S_1^{-1}K_1(\lambda)S_1 = \begin{bmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \lambda & 1 & \\ & & & \ddots & \ddots \\ & & & & \ddots & 1 \\ 0 & & & & & \lambda \end{bmatrix}$$

- (4) Assemble all of the blocks to form the Jordan form. For example, the block  $K_1(\lambda)$  and the corresponding similarity transformation  $S_1$  studied above can be treated in the assembly process as follows: Define the  $n \times n$  matrix

$$\hat{S}_1 = \begin{bmatrix} I & 0 & 0 \\ 0 & S_1 & 0 \\ 0 & 0 & I \end{bmatrix}$$

where  $S_1$  is the block constructed above and placed in the  $n \times n$  matrix in the position that  $K_1(\lambda)$  was extracted from the block triangular form of  $A$ . Repeat this for each of the blocks pertaining to minimally invariant subspaces. This gives a sequence of block diagonal matrices  $\hat{S}_1, \hat{S}_2, \dots, \hat{S}_k$ . Define  $T = \hat{S}_1\hat{S}_2\dots\hat{S}_k$ . It has the form

$$T = \begin{bmatrix} \hat{S}_1 & & & 0 \\ & \hat{S}_2 & & \\ & & \ddots & \\ 0 & & & \hat{S}_k \end{bmatrix}$$

Together with the original matrix  $P$  that transformed the matrix to the minimal invariant subspace blocked form and the unitary matrix



*Proof.* We assume  $A$  is a Jordan matrix. Each Jordan block  $J_k(\lambda)$  can be written as

$$J_k(\lambda) = \lambda I_k + N_k$$

where

$$N_k = \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{bmatrix} \text{ is nilpotent.}$$

Now

$$A = \begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{m_k}(\lambda_k) \end{pmatrix}.$$

We compute, for  $m > n_k$

$$\begin{aligned} (J_{n_k}(\lambda_k))^m &= (\lambda I + N)^m \\ &= \lambda^m I + \sum_{j=0}^{m_k} \lambda^{m-j} N^j \binom{m}{j} \end{aligned}$$

because  $N^j = 0$  for  $j > n_k$ . We have

$$\lambda^{m-j} \binom{m}{j} \rightarrow 0 \text{ as } m \rightarrow \infty$$

since  $|\lambda| < 1$ . The results follows.  $\square$

## 8.5 Exercises

1. Prove Theorem 8.2.1.
2. Find a  $3 \times 3$  matrix that has the same eigenvalues are the squares of the roots of the equation  $\lambda^3 - 3\lambda^2 + 4\lambda - 5 = 0$ .
3. Suppose that  $A$  is a square matrix with  $\sigma(A) = \{3\}$ ,  $m_a(3) = 6$ , and  $m_g(3) = 3$ . Up to permutations of the blocks show all possible Jordan normal forms for  $A$ .

4. Let  $A \in M_n(\mathbb{C})$  and let  $x_1 \in \mathbb{C}^n$ . Define  $x_{i+1} = Ax_i$  for  $i = 1, \dots, n-1$ . Show that  $V = \mathfrak{S}(\{x_1, \dots, x_n\})$  is an invariant subspace of  $A$ . Show that  $V$  contains an eigenvector of  $A$ .
5. Referring to the previous problem, let  $A \in M_n(\mathbb{R})$  be a permutation matrix. (i) Find starting vectors so that  $\dim V = n$ . (ii) Find starting vectors so that  $\dim V = 1$ . (iii) Show that if  $\lambda = 1$  is a simple eigenvalue of  $A$  then  $\dim V = 1$  or  $\dim V = n$ .

## Chapter 9

# Hermitian and Symmetric Matrices

**Example 9.0.1.** Let  $f: D \rightarrow R$ ,  $D \subset R^n$ . The *Hessian* is defined by

$$H(x) = h_{ij}(x) \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \in M_n.$$

Since for functions  $f \in C^2$  it is known that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

it follows that  $H(x)$  is symmetric.

**Definition 9.0.1.** A function  $f: R \rightarrow R$  is *convex* if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for  $x, y \in D$  (domain) and  $0 \leq \lambda \leq 1$ .

**Proposition 9.0.1.** If  $f \in C^2(D)$  and  $f''(x) \geq 0$  on  $D$  then  $f(x)$  is convex.

*Proof.* Because  $f'' \geq 0$ , this implies that  $f'(x)$  is increasing. Therefore if  $x < x_m < y$  we must have

$$f(x_m) \leq f(x) + f'(x_m)(x_m - x)$$

and

$$f(x_m) \leq f(y) + f'(x_m)(x_m - y)$$