CHAPTER 3

Canonical forms for similarity, and triangular factorizations

3.0 Introduction

†How can we tell if two given matrices are similar? The two matrices

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.0.0}$$

have the same eigenvalues, and hence they have the same characteristic polynomial, trace, and determinant. They also have the same rank, but $A^2 = 0$ and $B^2 \neq 0$, so A and B are not similar.

One approach to determining whether given square complex matrices A and B are similar would be to have in hand a set of special matrices of prescribed form, and see if both given matrices can be reduced by similarity to the same special matrix. If so, then A and B must be similar because the similarity relation is transitive and reflexive. If not, then we would like to be able to conclude that A and B are not similar. What sets of special matrices would be suitable for this purpose?

Every square complex matrix is similar to an upper triangular matrix. However, two upper triangular matrices with the same main diagonals but some different off-diagonal entries can still be similar (2.3.2b). Thus, we have a uniqueness problem: if we reduce A and B to two unequal upper triangular matrices with the same main diagonal, we cannot conclude from this fact alone that A and B are not similar.

The class of upper triangular matrices is too large for our purposes, but what about the smaller class of diagonal matrices? Uniqueness is no longer an issue,

† *Matrix Analysis*, second edition by Roger A. Horn and Charles R. Johnson, copyright Cambridge University Press 2009. June 8, 2009 version.

but now we have an existence problem: Some similarity equivalence classes contain no diagonal matrices.

The key to finding a suitable set of special matrices turns out to be a deft compromise between diagonal matrices and upper triangular matrices: A *Jordan matrix* is a special block upper triangular form that can be achieved by similarity for every complex matrix. Two Jordan matrices are similar if and only if they have the same diagonal blocks, without regard to their ordering. Moreover, no other matrix in the similarity equivalence class of a Jordan matrix J has strictly fewer nonzero off-diagonal entries than J.

Similarity is only one of many equivalence relations of interest in matrix theory; several others are listed in (0.11). Whenever we have an equivalence relation on a set of matrices, we want to be able to decide whether given matrices A and B are in the same equivalence class. A classical and broadly successful approach to this decision problem is to identify a set of representative matrices for the given equivalence relation such that (a) there is a representative in each equivalence class, and (b) distinct representatives are not equivalent. The test for equivalence of A and B is to reduce each via the given equivalence to a representative matrix and see if the two representative matrices are the same. Such a set of representatives is a *canonical form* for the equivalence relation.

For example, the spectral theorem (2.5.3) provides a canonical form for the set of normal matrices under unitary similarity: the diagonal matrices are a set of representative matrices (we identify two diagonal matrices if one is a permutation similarity of the other). Another example is the singular value decomposition (2.6.3), which provides a canonical form for M_n under unitary equivalence: the diagonal matrices $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ are the representative matrices.

3.1 The Jordan canonical form theorem

3.1.1 Definition. A *Jordan block* $J_k(\lambda)$ is a k-by-k upper triangular matrix of the form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & & \mathbf{0} \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ \mathbf{0} & & & & \lambda \end{bmatrix}; J_1(\lambda) = [\lambda], J_2(\lambda) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \tag{3.1.2}$$

The scalar λ appears k times on the main diagonal; if $k > 1$, there are $k - 1$ entries "$+1$" in the superdiagonal; all other entries are zero. A *Jordan matrix*

$J \in M_n$ is a direct sum of Jordan blocks

$$J = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_q}(\lambda_q), \qquad n_1 + n_2 + \cdots + n_q = n \quad (3.1.3)$$

Neither the block sizes n_i nor the scalars λ_i need be distinct.

The main result of this section is that every complex matrix is similar to an essentially unique Jordan matrix. We proceed to this conclusion in three steps, two of which have already been taken:

Step 1. Theorem 2.3.1 ensures that every complex matrix is similar to an upper triangular matrix whose eigenvalues appear on the main diagonal, and equal eigenvalues are grouped together.

Step 2. Theorem 2.4.6.1 ensures that a matrix of the form described in Step 1 is similar to a block diagonal upper triangular matrix (2.4.6.2) in which each diagonal block has equal diagonal entries.

Step 3. In this section, we show that an upper triangular matrix with equal diagonal entries is similar to a Jordan matrix.

We are also interested in concluding that if a matrix is real and has only real eigenvalues, then it can be reduced to a Jordan matrix via a *real* similarity. If a real matrix A has only real eigenvalues, then (2.3.1) and (2.4.6.1) ensure that there is a real similarity matrix S such that $S^{-1}AS$ is a (real) block diagonal upper triangular matrix of the form (2.4.6.2). Thus, it suffices to show that a real upper triangular matrix with equal main diagonal entries can be reduced to a direct sum of Jordan blocks via a real similarity.

The following lemma is helpful in taking Step 3; its proof is an entirely straightforward computation. The k-by-k Jordan block with eigenvalue zero is called a *nilpotent Jordan block*.

3.1.4 Lemma. Let $k \geq 2$ be given. Let $I_{k-1} \in M_{k-1}$ be an identity matrix, let e_i denote the ith standard unit basis vector, and let $x \in \mathbf{C}^k$ be given. Then

$$J_k^T(0)J_k(0) = \begin{bmatrix} 0 & 0 \\ 0 & I_{k-1} \end{bmatrix} \quad \text{and} \quad J_k(0)^p = 0 \quad \text{if} \quad p \geq k$$

Moreover, $J_k(0)e_{i+1} = e_i$ for $i = 1, 2, \ldots, k-1$ and $[I - J_k^T(0)J_k(0)]x = (x^T e_1)e_1$.

We now address the issue in Step 3.

3.1.5 Theorem. Let $A \in M_n$ be strictly upper triangular. There is a nonsingular $S \in M_n$ and there are integers $n_1, n_2, \ldots, n_m$ with $n_1 \geq n_2 \geq \cdots \geq n_m \geq 1$ and $n_1 + n_2 + \cdots + n_m = n$ such that

$$A = S\left(J_{n_1}(0) \oplus J_{n_2}(0) \oplus \cdots \oplus J_{n_m}(0)\right)S^{-1} \quad (3.1.6)$$

If A is real, the similarity matrix S may be chosen to be real.

Proof: If $n = 1$, $A = [0]$ and the result is trivial. We proceed by induction on n. Assume that $n > 1$ and that the result has been proved for all strictly upper triangular matrices of size less than n. Partition $A = \left[\begin{smallmatrix} 0 & a^T \\ 0 & A_1 \end{smallmatrix}\right]$, in which $a \in \mathbf{C}^{n-1}$ and $A_1 \in M_{n-1}$ is strictly upper triangular. By the induction hypothesis, there is a nonsingular $S_1 \in M_{n-1}$ such that $S_1^{-1}A_1S_1$ has the desired form (3.1.6); that is,

$$S_1^{-1}A_1S_1 = \begin{bmatrix} J_{k_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_{k_s} \end{bmatrix} = \begin{bmatrix} J_{k_1} & 0 \\ 0 & J \end{bmatrix} \tag{3.1.7}$$

in which $k_1 \geq k_2 \geq \cdots \geq k_s \geq 1, k_1 + k_2 + \cdots + k_s = n - 1, J_{k_i} = J_{k_i}(0)$, and $J = J_{k_2} \oplus \cdots \oplus J_{k_s} \in M_{n-k_1-1}$. No diagonal Jordan block in J has size greater than k_1, so $J^{k_1} = 0$. A computation reveals that

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & a^TS_1 \\ 0 & S_1^{-1}A_1S_1 \end{bmatrix} \tag{3.1.8}$$

Partition $a^TS_1 = [a_1^T a_2^T]$ with $a_1 \in \mathbf{C}^{k_1}$ and $a_2 \in \mathbf{C}^{n-k_1-1}$, and write (3.1.8) as

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & a_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}$$

Now consider the similarity

$$\begin{bmatrix} 1 & -a_1^TJ_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} 0 & a_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} \begin{bmatrix} 1 & a_1^TJ_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} 0 & a_1^T(I - J_{k_1}^TJ_{k_1}) & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} = \begin{bmatrix} 0 & (a_1^Te_1)e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} \tag{3.1.9}$$

in which we use the identity $(I - J_k^TJ_k)x = (x^Te_1)e_1$. There are now two possibilities, depending on whether $a_1^Te_1 \neq 0$ or $a_1^Te_1 = 0$.

If $a_1^Te_1 \neq 0$, then

$$\begin{bmatrix} 1/a_1^Te_1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & (1/a_1^Te_1)I \end{bmatrix} \begin{bmatrix} 0 & (a_1^Te_1)e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} \begin{bmatrix} a_1^Te_1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & a_1^Te_1I \end{bmatrix} \blacksquare$$

$$= \begin{bmatrix} 0 & e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} \equiv \begin{bmatrix} \tilde{J} & e_1 a_2^T \\ 0 & J \end{bmatrix}$$

Notice that $\tilde{J} = \left[\begin{smallmatrix} 0 & e_1^T \\ 0 & J_{k_1} \end{smallmatrix}\right] = J_{k_1+1}(0)$. Since $\tilde{J}e_{i+1} = e_i$ for $i = 1, 2, \ldots, k_1$, a computation reveals that

$$\begin{aligned} \begin{bmatrix} I & e_2 a_2^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{J} & e_1 a_2^T \\ 0 & J \end{bmatrix} \begin{bmatrix} I & -e_2 a_2^T \\ 0 & I \end{bmatrix} &= \begin{bmatrix} \tilde{J} & -\tilde{J}e_2 a_2^T + e_1 a_2^T + e_2 a_2^T J \\ 0 & J \end{bmatrix} \\ &= \begin{bmatrix} \tilde{J} & e_2 a_2^T J \\ 0 & J \end{bmatrix} \end{aligned}$$

We can proceed recursively to compute the sequence of similarities

$$\begin{bmatrix} I & e_{i+1} a_2^T J^{i-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{J} & e_i a_2^T J^{i-1} \\ 0 & J \end{bmatrix} \begin{bmatrix} I & -e_{i+1} a_2^T J^{i-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} \tilde{J} & e_{i+1} a_2^T J^i \\ 0 & J \end{bmatrix},$$

for $i = 2, 3, \ldots$. Since $J^{k_1} = 0$, after at most k_1 steps in this sequence of similarities, the off-diagonal term finally vanishes. We conclude that A is similar to $\left[\begin{smallmatrix} \tilde{J} & 0 \\ 0 & J \end{smallmatrix}\right]$, which is a strictly upper triangular Jordan matrix of the required form.

If $a_1^T e_1 = 0$, then (3.1.9) shows that A is similar to

$$\begin{bmatrix} 0 & 0 & a_2^T \\ 0 & \tilde{J}_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}$$

which is permutation similar to

$$\begin{bmatrix} J_{k_1} & 0 & 0 \\ 0 & 0 & a_2^T \\ 0 & 0 & J \end{bmatrix} \tag{3.1.10}$$

By the induction hypothesis, there is a nonsingular $S_2 \in M_{n-k_1}$ such that $S_2^{-1}\left[\begin{smallmatrix} 0 & a_2^T \\ 0 & J \end{smallmatrix}\right]S_2 = \hat{J} \in M_{n-k_1}$ is a Jordan matrix with zero main diagonal. Thus, the matrix (3.1.10), and therefore A itself, is similar to $\left[\begin{smallmatrix} J_{k_1} & 0 \\ 0 & \hat{J} \end{smallmatrix}\right]$, which is a Jordan matrix of the required form, except that the diagonal Jordan blocks might not be arranged in nonincreasing order of their size. A block permutation similarity, if necessary, produces the required form.

Finally, observe that if A is real then all the similarities in this proof are real, so A is similar via a real similarity to a Jordan matrix of the required form. $\square$

Theorem (3.1.5) essentially completes Step 3, as the general case is an easy consequence of the nilpotent case. If $A \in M_n$ is an upper triangular matrix with all diagonal entries equal to λ, then $A_0 = A - \lambda I$ is strictly upper triangular. If $S \in M_n$ is nonsingular and $S^{-1}A_0S$ is a direct sum of nilpotent Jordan blocks $J_{n_i}(0)$, as guaranteed by (3.1.5), then $S^{-1}AS = S^{-1}A_0S + \lambda I$ is a direct sum of Jordan blocks $J_{n_i}(\lambda)$ with eigenvalue λ. We have now established the existence assertion of the *Jordan canonical form theorem*:

3.1.11 Theorem. Let $A \in M_n$ be given. There is a nonsingular $S \in M_n$, positive integers q and $n_1, \ldots, n_q$ with $n_1 + n_2 + \cdots + n_q = n$, and scalars $\lambda_1, \ldots, \lambda_q \in \mathbf{C}$ such that

$$A = S \begin{bmatrix} J_{n_1}(\lambda_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_{n_q}(\lambda_q) \end{bmatrix} S^{-1} \qquad (3.1.12)$$

The Jordan matrix $J_A = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_q}(\lambda_q)$ is uniquely determined by A up to permutation of its direct summands. If A is real and has only real eigenvalues, then S can be chosen to be real.

The Jordan matrix J_A in the preceding theorem is the *Jordan canonical form* of A.

Two facts provide the key to understanding the uniqueness assertion in the Jordan canonical form theorem: (1) similarity of two matrices is preserved if they are both translated by the same scalar matrix, and (2) rank is a similarity invariant.

If $A, B, S \in M_n$, S is nonsingular, and $A = SBS^{-1}$, then for any $\lambda \in \mathbf{C}$, $A - \lambda I = SBS^{-1} - \lambda SS^{-1} = S(B - \lambda I)S^{-1}$. Moreover, for every $k = 1, 2, \ldots$, the matrices $(A - \lambda I)^k$ and $(B - \lambda I)^k$ are similar; in particular, their ranks are equal. We focus on this assertion when $B = J = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_q}(\lambda_q)$ is a Jordan matrix that is similar to A (the existence assertion of (3.1.11)) and λ is an eigenvalue of A. After a permutation of the diagonal blocks of J (a permutation similarity), we may assume that $J = J_{m_1}(\lambda) \oplus \cdots \oplus J_{m_p}(\lambda) \oplus \hat{J}$, in which the Jordan matrix $\hat{J}$ is a direct sum of Jordan blocks with eigenvalues different from λ. Then $A - \lambda I$ is similar to

$$\begin{aligned} J - \lambda I &= (J_{m_1}(\lambda) - \lambda I) \oplus \cdots \oplus (J_{m_p}(\lambda) - \lambda I) \oplus (\hat{J} - \lambda I) \\ &= J_{m_1}(0) \oplus \cdots \oplus J_{m_p}(0) \oplus (\hat{J} - \lambda I) \end{aligned}$$

which is a direct sum of p nilpotent Jordan blocks of various sizes and a nonsingular Jordan matrix $\hat{J} - \lambda I \in M_m$, in which $m = n - (m_1 + \cdots + m_p)$. Moreover, $(A - \lambda I)^k$ is similar to $(J - \lambda I)^k = J_{m_1}(0)^k \oplus \cdots \oplus J_{m_p}(0)^k \oplus (\hat{J} - \lambda I)^k$

for each $k = 1, 2, \ldots$. Since the rank of a direct sum is the sum of the ranks of the summands (0.9.2), we have

$$\begin{aligned}\operatorname{rank}(A-\lambda I)^k &= \operatorname{rank}(J-\lambda I)^k \\ &= \operatorname{rank} J_{m_1}(0)^k + \cdots + \operatorname{rank} J_{m_p}(0)^k + \operatorname{rank}(\hat{J}-\lambda I)^k \\ &= \operatorname{rank} J_{m_1}(0)^k + \cdots + \operatorname{rank} J_{m_p}(0)^k + m \qquad (3.1.13)\end{aligned}$$

for each $k = 1, 2, \ldots$.

What is the rank of a power of a nilpotent Jordan block? Inspection of (3.1.2) reveals that the first column of $J_\ell(0)$ is zero and its last $\ell - 1$ columns are independent (the only nonzero entries are ones in the first superdiagonal), so $\operatorname{rank} J_\ell(0) = \ell - 1$. The only nonzero entries in $J_\ell(0)^2$ are ones in the second superdiagonal, so its first two columns are zero, its last $\ell - 2$ columns are independent, and $\operatorname{rank} J_\ell(0)^2 = \ell - 2$. The ones move up one superdiagonal (so the number of zero columns increases by one and the rank drops by one) with each successive power until $J_\ell(0)^{\ell-1}$ has just one nonzero entry (in position $1, \ell$) and $\operatorname{rank} J_\ell(0)^{\ell-1} = 1 = \ell - (\ell - 1)$. Of course, $J_\ell(0)^k = 0$ for all $k = \ell, \ell + 1, \ldots$. In general, we have $\operatorname{rank} J_\ell(0)^k = \max\{\ell - k, 0\}$ for each $k = 1, 2, \ldots$, and so

$$\operatorname{rank} J_\ell(0)^{k-1} - \operatorname{rank} J_\ell(0)^k = \begin{cases} 1 \text{ if } \ell \geq k \\ 0 \text{ if } \ell < k \end{cases}, \; k = 1, 2, \ldots \qquad (3.1.14)$$

in which we observe the standard convention that $\operatorname{rank} J_\ell(0)^0 = \ell$.

Now let $A \in M_n$, let $\lambda \in \mathbf{C}$, let k be a positive integer, let

$$r_k(A, \lambda) = \operatorname{rank}(A - \lambda I)^k, \quad r_0(A, \lambda) := n \qquad (3.1.15)$$

and define

$$w_k(A, \lambda) = r_{k-1}(A, \lambda) - r_k(A, \lambda), \quad w_1(A, \lambda) := n - r_1(A, \lambda) \quad (3.1.16)$$

Exercise. If $A \in M_n$ and $\lambda \in \mathbf{C}$ is *not* an eigenvalue of A, explain why $w_k(A, \lambda) = 0$ for all $k = 1, 2, \ldots$.

Exercise. Consider the Jordan matrix

$$J = J_3(0) \oplus J_3(0) \oplus J_2(0) \oplus J_2(0) \oplus J_2(0) \oplus J_1(0) \qquad (3.1.16a)$$

Verify that $r_1(J, 0) = 7$, $r_2(J, 0) = 2$, and $r_3(J, 0) = r_4(J, 0) = 0$. Also verify that $w_1(J, 0) = 6$ is the number of blocks of size at least 1, $w_2(J, 0) = 5$ is the number of blocks of size at least 2, $w_3(J, 0) = 2$ is the number of blocks of size at least 3, and $w_4(J, 0) = 0$ is the number of blocks of size at least 4. Observe that $w_1(J, 0) - w_2(J, 0) = 1$ is the number of blocks of size 1, $w_2(J, 0) - w_3(J, 0) = 3$ is the number of blocks of size 2, and

$w_3(J,0) - w_4(J,0) = 2$ is the number of blocks of size 3. This is not an accident.

Use (3.1.13) and (3.1.14) to compute

$$\begin{aligned} w_k(A,\lambda) &= \left(\operatorname{rank} J_{m_1}(0)^{k-1} - \operatorname{rank} J_{m_1}(0)^k\right) + \\ &\quad \cdots + \left(\operatorname{rank} J_{m_p}(0)^{k-1} - \operatorname{rank} J_{m_p}(0)^k\right) \\ &= (1 \text{ if } m_1 \geq k) + \cdots + (1 \text{ if } m_p \geq k) \qquad (3.1.17) \\ &= \text{number of blocks with eigenvalue } \lambda \text{ that have size at least } k \end{aligned}$$

In particular, $w_1(A,\lambda)$ is the number of Jordan blocks of A of all sizes that have eigenvalue λ, which is the geometric multiplicity of λ as an eigenvalue of A.

Using the characterization (3.1.17), we see that $w_k(A,\lambda) - w_{k+1}(A,\lambda)$ is the number of blocks with eigenvalue λ that have size at least k but do not have size at least $k+1$; this is the number of blocks with eigenvalue λ that have size exactly k.

Exercise. Let $A, B \in M_n$ and $\lambda \in \mathbf{C}$ be given. If A and B are similar, explain why $w_k(A,\lambda) = w_k(B,\lambda)$ for all $k = 1, 2, \ldots$.

Exercise. Let $A \in M_n$ and $\lambda \in \mathbf{C}$ be given. Explain why $w_1(A,\lambda) \geq w_2(A,\lambda) \geq w_3(A,\lambda) \geq \cdots$, that is, the sequence $w_1(A,\lambda), w_2(A,\lambda), \ldots$ is nonincreasing. *Hint*: $w_k(A,\lambda) - w_{k+1}(A,\lambda)$ is always a nonnegative integer. Why?

The *Weyr characteristic* of $A \in M_n$ associated with $\lambda \in \mathbf{C}$ is the sequence of integers $w_1(A,\lambda), w_2(A,\lambda), \ldots$ defined by (3.1.16). We have just seen that the structure of a Jordan matrix J that is similar to A is completely determined by the Weyr characteristics of A associated with its distinct eigenvalues: If λ is an eigenvalue of A, and if J is a Jordan matrix that is similar to A, then the number of Jordan blocks $J_k(\lambda)$ in J is exactly $w_k(A,\lambda) - w_{k+1}(A,\lambda), k = 1, 2, \ldots$. This means that two essentially different Jordan matrices (that is, for some eigenvalue, their respective lists of nonincreasingly ordered block sizes associated with that eigenvalue are not identical) cannot both be similar to A because their Weyr characteristics must be different. We have now proved the uniqueness portion of the Jordan canonical form theorem (3.1.11) and a little more:

3.1.18 Lemma. Let λ be a given eigenvalue of $A \in M_n$ and let $w_1(A,\lambda)$, $w_2(A,\lambda), \ldots$ be the Weyr characteristic of A associated with λ. The number of blocks of the form $J_k(\lambda)$ in the Jordan canonical form of A is $w_k(A,\lambda) - w_{k+1}(A,\lambda)$, $k = 1, 2, \ldots$. Two square complex matrices of the same size are similar if and only if they have the same eigenvalues, and the same Weyr characteristics are associated with each eigenvalue.

Exercise. Let q denote the size of the largest Jordan block of A with eigenvalue λ, and consider the rank identity (3.1.13). Explain why $\operatorname{rank}(A-\lambda I)^k = \operatorname{rank}(A-\lambda I)^{k+1}$ for all $k \geq q$, $w_q(A,\lambda)$ is the number of Jordan blocks of A with eigenvalue λ and maximum size q, and $w_k(A,\lambda) - w_{k+1}(A,\lambda) = 0$ for all $k > q$. This integer q is called the *index* of λ as an eigenvalue of A.

Exercise. Let $A \in M_n$ and let λ have index q as an eigenvalue of A . Explain why (a) $w_1(A,\lambda)$ is the geometric multiplicity of λ (the number of Jordan blocks with eigenvalue λ in the Jordan canonical form of A); (b) $w_1(A,\lambda) + w_2(A,\lambda)+\cdots+w_q(A,\lambda)$ is the algebraic multiplicity of λ (the sum of the sizes of all the Jordan blocks of A with eigenvalue λ); (c) for each $p = 2, 3, \ldots q$, $w_p(A,\lambda) + w_{p+1}(A,\lambda) + \cdots + w_q(A,\lambda) = \operatorname{rank}(A-\lambda I)^{p-1}$.

The Jordan structure of a given $A \in M_n$ can be completely specified by giving, for each distinct eigenvalue λ of A, a list of the sizes of all the Jordan blocks of A that have eigenvalue λ. The *nonincreasingly ordered* list of sizes of Jordan blocks of A with eigenvalue λ

$$s_1(A,\lambda) \geq s_2(A,\lambda) \geq \cdots \geq s_{w_1(A,\lambda)}(A,\lambda) > 0 = s_{w_1(A,\lambda)+1}(A,\lambda) = \cdots \tag{3.1.19}$$

is called the *Segre characteristic* of A associated with the eigenvalue λ. It is convenient to define $s_k(A,\lambda) = 0$ for all $k > w_1(A,\lambda)$. Observe that $s_1(A,\lambda)$ is the index of λ as an eigenvalue of A (the size of the largest Jordan block of A with eigenvalue λ) and $s_{w_1(A,\lambda)}(A,\lambda)$ is the size of the smallest Jordan block of A with eigenvalue λ. For example, the Segre characteristic of the matrix (3.1.16a) associated with the zero eigenvalue is $3, 3, 2, 2, 2, 1$ ($s_1(J,0) = 3$ and $s_6(J,0) = 1$).

If $s_k = s_k(A,\lambda), k = 1, 2, \ldots$ is the Segre characteristic of $A \in M_n$ associated with the eigenvalue λ and $w_1 = w_1(A,\lambda)$, the part of the Jordan canonical form that contains all the Jordan blocks of A with eigenvalue λ is

$$\begin{bmatrix} J_{s_1}(\lambda) & & & \\ & J_{s_2}(\lambda) & & \\ & & \ddots & \\ & & & J_{s_{w_1}}(\lambda) \end{bmatrix} \tag{3.1.20}$$

It is easy to derive the Weyr characteristic if the Segre characteristic is known, and *vice versa*. For example, from the Segre characteristic $3, 3, 2, 2, 2, 1$ we see that there are 6 blocks of size 1 or greater, 5 blocks of size 2 or greater, and 2 blocks of size 3 or greater: the Weyr characteristic is 6,5,2. Conversely from the Weyr characteristic 6,5,2 we see that there are $6 - 5 = 1$ blocks of size 1, $5 - 2 = 3$ blocks of size 2, and $2 - 0 = 2$ blocks of size 3: the Segre characteristic is $3, 3, 2, 2, 2, 1$.

Our derivation of the Jordan canonical form is based on an explicit algorithm, but it cannot be recommended for implementation in a software package to compute Jordan canonical forms. A simple example illustrates the difficulty: If $A_\epsilon = \begin{bmatrix} \epsilon & 0 \\ 1 & 0 \end{bmatrix}$ and $\epsilon \neq 0$, then $A_\epsilon = S_\epsilon J_\epsilon S_\epsilon^{-1}$ with $S_\epsilon = \begin{bmatrix} 0 & \epsilon \\ 1 & 1 \end{bmatrix}$ and $J_\epsilon = \begin{bmatrix} 0 & 0 \\ 0 & \epsilon \end{bmatrix}$. If we let $\epsilon \to 0$, then $J_\epsilon \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = J_1(0) \oplus J_1(0)$, but $A_\epsilon \to A_0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, whose Jordan canonical form is $J_2(1)$. Small variations in the entries of a matrix can result in major changes in its Jordan canonical form. The root of the difficulty is that $\operatorname{rank} A$ is not a continuous function of the entries of A.

It is sometimes useful to know that every matrix is similar to a matrix of the form (3.1.12) in which all the "$+1$" entries in the Jordan blocks are replaced by any $\epsilon \neq 0$.

3.1.21 Corollary. Let $A \in M_n$ and a nonzero $\epsilon \in \mathbf{C}$ be given. Then there exists a nonsingular $S(\epsilon) \in M_n$ such that

$$A = S(\epsilon) \begin{bmatrix} J_{n_1}(\lambda_1, \epsilon) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_{n_k}(\lambda_k, \epsilon) \end{bmatrix} S(\epsilon)^{-1} \tag{3.1.22}$$

in which $n_1 + n_2 + \cdots + n_k = n$ and

$$J_m(\lambda, \epsilon) = \begin{bmatrix} \lambda & \epsilon & & \mathbf{0} \\ & \ddots & \ddots & \\ & & \ddots & \epsilon \\ \mathbf{0} & & & \lambda \end{bmatrix} \in M_m$$

If A is real and has real eigenvalues, and if $\epsilon \in \mathbf{R}$, then $S(\epsilon)$ may be taken to be real.

Proof: First find a nonsingular matrix $S_1 \in M_n$ such that $S_1^{-1} A S_1$ is a Jordan matrix of the form (3.1.3) (with a real S_1 if A is real and has real eigenvalues). Let $D_{\epsilon,i} = \operatorname{diag}(1, \epsilon, \epsilon^2, \ldots, \epsilon^{n_i - 1})$, define $D_\epsilon = D_{\epsilon,1} \oplus \cdots \oplus D_{\epsilon,q}$, and compute $D_\epsilon^{-1}(S_1^{-1} A S_1) D_\epsilon$. This matrix has the form (3.1.22), so $S(\epsilon) = S_1 D_\epsilon$ meets the stated requirements. $\square$

Problems

1. Supply the computational details to prove Lemma (3.1.4).

2. What are the Jordan canonical forms of the two matrices in (3.0.0)?

3. Suppose $A \in M_n$ has some non-real entries, but only real eigenvalues. Show that A is similar to a real matrix. Can the similarity matrix ever be chosen to be real?

4. Let $A \in M_n$ be given. If A is similar to cA for some complex scalar c with $|c| \neq 1$, show that $\sigma(A) = \{0\}$ and hence A is nilpotent. Conversely, if A is nilpotent, show that A is similar to cA for all nonzero $c \in \mathbf{C}$.

5. Explain why every Jordan block $J_k(\lambda)$ has a one-dimensional eigenspace associated with the eigenvalue λ. Conclude that λ has geometric multiplicity one and algebraic multiplicity k as an eigenvalue of $J_k(\lambda)$.

6. Carry out the three steps in the proof of (3.1.11) to find the Jordan canonical forms of

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 1 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Confirm your answers by using (3.1.18).

7. Let $A \in M_n$, let λ be an eigenvalue of A, and let $k \in \{1, \ldots, n\}$. Using (3.1.15-16), explain why $r_{k-1}(\lambda) - 2r_k(\lambda) + r_{k+1}(\lambda)$ is the number of Jordan blocks of A that have size k and eigenvalue λ.

8. Let $A \in M_n$ be given. Suppose that $\operatorname{rank} A = r \geq 1$ and $A^2 = 0$. Use the preceding problem or (3.1.18) to show that the Jordan canonical form of A is $J_2(0) \oplus \cdots \oplus J_2(0) \oplus 0_{n-2r}$ (there are r 2-by-2 blocks). Compare with Problem 23 in (2.6).

9. Let $n \geq 3$. Show that the Jordan canonical form of $J_n(0)^2$ is $J_m(0) \oplus J_m(0)$ if $n = 2m$ is even, and it is $J_{m+1}(0) \oplus J_m(0)$ if $n = 2m + 1$ is odd.

10. For any $\lambda \in \mathbf{C}$ and any positive integer k, show that the Jordan canonical form of $-J_k(\lambda)$ is $J_k(-\lambda)$. In particular, the Jordan canonical form of $-J_k(0)$ is $J_k(0)$.

11. The information contained in the Weyr characteristic of a matrix associated with a given eigenvalue can be presented as a *dot diagram*, sometimes called a *Ferrers diagram* or *Young diagram*. For example, consider the Jordan matrix J in (3.1.16a) and its Weyr characteristic $w_k = w_k(J, 0), k = 1, 2, 3$. Construct the dot diagram

$$\begin{array}{ccccccc} w_1 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ w_2 & \bullet & \bullet & \bullet & \bullet & \bullet & \\ w_3 & \bullet & \bullet & & & & \\ & s_1 & s_2 & s_3 & s_3 & s_5 & s_6 \end{array}$$

by putting w_1 dots in the first row, w_2 dots in the second row, and w_3 dots in the third row. We stop with the third row since $w_k = 0$ for all $k \geq 4$. Proceeding from the left, the respective column lengths are $3, 3, 2, 2, 2, 1$, which is the Segre characteristic $s_k = s_k(J, 0), k = 1, 2, \ldots, 6$. That is, J has 2 Jordan blocks of the form $J_3(0)$, 3 blocks of the form $J_2(0)$, and one block of the form $J_1(0)$. Conversely, if one first constructs a dot diagram by putting s_1 dots in the first column, s_2 dots in the second column, and so forth, then there are w_1 dots in the first row, w_2 dots in the second row, and w_3 dots in the third row. In this sense, the Segre and Weyr characteristics are *conjugate partitions* of their common sum n; either characteristic can be derived from the other via a dot diagram. In general, for $A \in M_n$ and a given eigenvalue λ of A, use the Weyr characteristic to construct a dot diagram with $w_k(A, \lambda)$ dots in row $k = 1, 2, \ldots$ so long as $w_k(A, \lambda) > 0$. (a) Explain why there are $s_j(A, \lambda)$ dots in column j for each $j = 1, 2, \ldots$. (b) Explain why one can also start with the Segre characteristic, construct the columns of a dot diagram from it, and then read off the Weyr characteristic from the rows.

12. Let $A \in M_n$. Write $w_k = w_k(A, \lambda)$ and $s_k = s_k(A, \lambda)$ for the Weyr and Segre characteristics of A associated with the eigenvalue λ. Show that: (a) $s_{w_k} \geq k$ if $w_k > 0$; (b) $k > s_{w_k+1}$ for all k; (c) $w_{s_k} \geq k$ if $s_k > 0$; (d) $k > w_{s_k+1}$ for all k.

13. Let k and m be given positive integers and consider the *block Jordan matrix*

$$J_k^+(\lambda I_m) := \begin{bmatrix} \lambda I_m & I_m & & \\ & \lambda I_m & \ddots & \\ & & \ddots & I_m \\ & & & \lambda I_m \end{bmatrix} \in M_{km}$$

(a block k-by-k matrix). Compute the Weyr characteristic of $J_k^+(\lambda I_m)$ and use it to show that the Jordan canonical form of $J_k^+(\lambda I_m)$ is $J_k(\lambda) \oplus \cdots \oplus J_k(\lambda)$ (m summands).

14. Let $A \in M_n$. Use (3.1.18) to show that A and A^T are similar. Are A and A^* similar?

15. Let $n \geq 2$, let $x, y \in \mathbf{C}^n$ be given nonzero vectors, and let $A = xy^*$. Explain why the Jordan canonical form of A is $B \oplus 0_{n-2}$, in which $B = \left[\begin{smallmatrix} y^*x & 0 \\ 0 & 0 \end{smallmatrix}\right]$ if $y^*x \neq 0$ and $B = J_2(0)$ if $y^*x = 0$.

16. Suppose that $\lambda \neq 0$ and $k \geq 2$. Then $J_k(\lambda)^{-1}$ is a polynomial in $J_k(\lambda)$ (2.4.3.4). (a) Explain why $J_k(\lambda)^{-1}$ is an upper triangular Toeplitz matrix, all of whose main diagonal entries are λ^{-1}. (b) Let $[\lambda^{-1}\ a_2\ \ldots\ a_n]$ be the first row of $J_k(\lambda)^{-1}$. Verify that the $1,2$ entry of $J_k(\lambda)J_k(\lambda)^{-1}$ is $\lambda a_2 + \lambda^{-1}$ and explain why all the entries in the first superdiagonal of $J_k(\lambda)^{-1}$ are $-\lambda^{-2}$; in particular, these entries are all nonzero. (c) Show that $\operatorname{rank}(J_k(\lambda)^{-1} - \lambda^{-1}I)^k = n - k$ for $k = 1, ..., n$ and explain why the Jordan canonical form of $J_k(\lambda)^{-1}$ is $J_k(\lambda^{-1})$. *Hint*: (3.1.18).

17. Suppose that $A \in M_n$ is nonsingular. Show that A is similar to A^{-1} if and only if for each eigenvalue λ of A with $\lambda \neq \pm 1$, the number of Jordan blocks of the form $J_k(\lambda)$ in the Jordan canonical form of A is equal to the number of blocks of the form $J_k(\lambda^{-1})$, that is, the blocks $J_k(\lambda)$ and $J_k(\lambda^{-1})$ occur in pairs if $\lambda \neq \pm 1$ (there is no restriction on the blocks with eigenvalues ± 1. *Hint*: Problem 16.

18. Suppose that $A \in M_n$ is nonsingular. (a) If each eigenvalue of A is either $+1$ or -1, explain why A is similar to A^{-1}. (b) Suppose that there are nonsingular $B, C, S \in M_n$ such that $A = BC$, $B^{-1} = SBS^{-1}$, and $C^{-1} = SCS^{-1}$. Show that A is similar to A^{-1}. *Hint*: Problem 17 and (1.3.22).

19. Let $x, y \in \mathbf{R}^n$ and $t \in \mathbf{R}$ be given. Define

$$A_{x,y,t} = \begin{bmatrix} 1 & x^T & t \\ 0 & I_n & y \\ 0 & 0 & 1 \end{bmatrix} \in M_{n+2}(\mathbf{R})$$

and let $\mathcal{H}_n(\mathbf{R}) = \{A_{x,y,t} : x, y \in \mathbf{R}^n \text{ and } t \in \mathbf{R}\}$. (a) Show that $A_{x,y,t}A_{\xi,\eta,\tau} = A_{x+\xi,y+\eta,t+\tau}$ and $(A_{x,y,t})^{-1} = A_{-x,-y,-t}$. (b) Explain why $\mathcal{H}_n(\mathbf{R})$ is a subgroup (called the n^{th} *Heisenberg group*) of the group of upper triangular matrices in $M_{n+2}(\mathbf{R})$ that have all main diagonal entries equal to $+1$. (c) Explain why: the Jordan canonical form of $A_{x,y,t}$ is $J_3(1) \oplus I_{n-1}$ if $x^Ty \neq 0$; if $x^Ty = 0$, it is either $J_2(1) \oplus J_2(1) \oplus I_{n-2}$ ($x \neq 0 \neq y$), or $J_2(1) \oplus I_n$ ($x = 0$ or $y = 0$ but not both), or I_{n+2} ($x = y = 0$). *Hint*: (3.1.18). (d) Explain why $A_{x,y,t}$ is always similar to its inverse.

20. Let $A \in M_n$ and suppose that $n > \operatorname{rank} A = r \geq 1$. If 0 is a semisimple eigenvalue of A, show that A has a nonsingular r-by-r principal submatrix (that is, A is rank principal (0.7.6)). *Hint*: Problem 16 in (1.3).

21. Let $A \in M_n$ be an unreduced upper Hessenberg matrix (0.9.9). (a) For each eigenvalue λ of A, explain why $w_1(A, \lambda) = 1$ and A is nonderogatory. (b) Suppose that A is diagonalizable (for example, A might be Hermitian and tridiagonal). Explain why A has n distinct eigenvalues.

22. Let $A \in M_n(\mathbf{R})$ be tridiagonal. (a) If $a_{i,i+1}a_{i+1,i} > 0$ for all $i = 1, \ldots, n-1$, show that A has n distinct real eigenvalues. *Hint*: Show that there is a positive diagonal D such that DAD^{-1} is symmetric and apply Problem 21. (b) If $a_{i,i+1}a_{i+1,i} \geq 0$ for all $i = 1, \ldots, n-1$, show that all the eigenvalues of A are real. *Hint*: Perturb A and use continuity.

23. Let $A = [a_{ij}] \in M_n$ be tridiagonal with a_{ii} real for all $i = 1, \ldots, n$. (a) If $a_{i,i+1}a_{i+1,i}$ is real and positive for $i = 1, \ldots, n-1$, show that A has n distinct real eigenvalues. *Hint*: Proceed as in Problem 22; choose a positive diagonal D such that DAD^{-1} is Hermitian. (b) If $a_{i,i+1}a_{i+1,i}$ is real and nonnegative for all $i = 1, \ldots, n-1$, show that all the eigenvalues of A are real.

24. Consider the 4-by-4 matrices $A = [A_{ij}]_{i,j=1}^2$ and $B = [B_{ij}]_{i,j=1}^2$, in which $A_{11} = A_{22} = B_{11} = B_{22} = J_2(0)$, $A_{21} = B_{21} = 0_2$, $A_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, and $B_{12} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. (a) For all $k = 1, 2, \ldots$, show that A^k and B^k are $0-1$ matrices (that is, every entry is 0 or 1) that have the same number of entries equal to 1. (b) Explain why A and B are nilpotent and similar. What is their Jordan canonical form? (c) Explain why two permutation similar $0-1$ matrices have the same number of entries equal to 1. (d) Show that A and B are not permutation similar. *Hint*: Consider the directed graphs of A and B (6.2).

Notes and Further Readings. Camille Jordan published his eponymous canonical form in C. Jordan, *Traité des Substitutions et des Équations Algébriques*, Gauthier-Villars, Paris, 1870; see §157. Our proof of (3.1.11) is in the spirit of R. Fletcher and D. Sorensen, An Algorithmic Derivation of the Jordan Canonical Form, *Amer. Math. Monthly* 90 (1983) 12–16. For a combinatorial approach, see R. Brualdi, The Jordan Canonical Form: An Old Proof, *Amer. Math. Monthly* 94 (1987) 257-267.

3.2 The Jordan canonical form: some observations and applications

3.2.1 The structure of a Jordan matrix. The Jordan matrix

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_k}(\lambda_k) \end{bmatrix}, \quad n_1 + n_2 + \cdots + n_k = n \quad (3.2.1.1)$$

has a definite structure that makes apparent certain basic properties of any matrix that is similar to it.

1. The number k of Jordan blocks (counting multiple occurrences of the same block) is the maximum number of linearly independent eigenvectors of J.
2. The matrix J is diagonalizable if and only if $k = n$, that is, if and only if all the Jordan blocks are 1-by-1.
3. The number of Jordan blocks corresponding to a given eigenvalue is the geometric multiplicity of the eigenvalue, which is the dimension of the associated eigenspace. The sum of the sizes of all the Jordan blocks corresponding to a given eigenvalue is its algebraic multiplicity.
4. Let $A \in M_n$ be a given nonzero matrix, and suppose that λ is an eigenvalue of A. Using (3.1.14) and the notation of (3.1.15), we know that there is some positive integer q such that

   $$r_1(A, \lambda) > r_2(A, \lambda) > \cdots > r_{q-1}(A, \lambda) > r_q(A, \lambda) = 0$$

   This integer q is the index of λ as an eigenvalue of A; it is also the size of the largest Jordan block of A with eigenvalue λ.

3.2.2 Linear systems of ordinary differential equations. One application of the Jordan canonical form that is of considerable theoretical importance is to the analysis of solutions of a system of first order linear ordinary differential equations with constant coefficients. Let $A \in M_n$ be given, and consider the first-order initial value problem

$$\begin{aligned} x'(t) &= Ax(t) \\ x(0) &= x_0 \text{ is given} \end{aligned} \tag{3.2.2.1}$$

in which $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$, and the prime ($'$) denotes differentiation with respect to t. If A is not a diagonal matrix, this system of equations is *coupled*; that is, $x_i'(t)$ is related not only to $x_i(t)$ but to the other entries of the vector $x(t)$ as well. This coupling makes the problem hard to solve, but if A can be transformed to diagonal (or almost diagonal) form, the amount of coupling can be reduced or even eliminated and the problem may be easier to solve. If $A = SJS^{-1}$ and J is the Jordan canonical form of A, then (3.2.2.1) becomes

$$\begin{aligned} y'(t) &= \quad Jy(t) \\ y(0) &= y_0 \text{ is given} \end{aligned} \tag{3.2.2.2}$$

in which $x(t) = Sy(t)$ and $y_0 = S^{-1}x_0$. If the problem (3.2.2.2) can be solved, then each entry of the solution $x(t)$ to (3.2.2.1) is just a linear combination of the entries of the solution to (3.2.2.2), and the linear combinations are given by S.

If A is diagonalizable, then J is a diagonal matrix, and (3.2.2.2) is just an uncoupled set of equations of the form $y_k'(t) = \lambda_k y_k(t)$, which have the solutions $y_k(t) = y_k(0)e^{\lambda_k t}$. If the eigenvalue λ_k is real, this is a simple exponential, and if $\lambda_k = a_k + ib_k$ is not real, $y_k(t) = y_k(0)e^{a_k t}[\cos(b_k t) + i\sin(b_k t)]$ is an oscillatory term with a real exponential factor if $a_k \neq 0$.

If J is not diagonal, the solution is more complicated but it can be described explicitly. The entries of $y(t)$ that correspond to distinct Jordan blocks in J are not coupled, so it suffices to consider the case in which $J = J_m(\lambda)$ is a single Jordan block. The system (3.2.2.2) is

$$\begin{array}{rcl} y_1'(t) & = & \lambda y_1(t) + y_2(t) \\ \vdots & \vdots & \vdots \\ y_{m-1}'(t) & = & \lambda y_{m-1}(t) + y_m(t) \\ y_m'(t) & = & \lambda y_m(t) \end{array}$$

which can be solved in a straightforward way from the bottom up. Starting with the last equation, we obtain

$$y_m(t) = y_m(0)e^{\lambda t}$$

so that

$$y_{m-1}'(t) = \lambda y_{m-1}(t) + y_m(0)e^{\lambda t}$$

This has the solution

$$y_{m-1}(t) = e^{\lambda t}[y_m(0)t + y_{m-1}(0)]$$

which can now be used in the next equation. It becomes

$$y_{m-2}'(t) = \lambda y_{m-2}(t) + y_m(0)te^{\lambda t} + y_{m-1}(0)e^{\lambda t}$$

which has the solution

$$y_{m-2}(t) = e^{\lambda t}[y_m(0)\frac{t^2}{2} + y_{m-1}(0)t + y_{m-2}(0)]$$

and so forth. Each entry of the solution has the form

$$y_k(t) = e^{\lambda t}q_k(t) = e^{\lambda t}\sum_{i=k}^{m} y_i(0)\frac{t^{i-k}}{(i-k)!}$$

so $q_k(t)$ is an explicitly determined polynomial of degree at most $m-k$, $k = 1, \dots, m$.

From this analysis, we conclude that the entries of the solution $x(t)$ of the problem (3.2.2.1) have the form

$$x_j(t) = e^{\lambda_1 t} p_1(t) + e^{\lambda_2 t} p_2(t) + \cdots + e^{\lambda_k t} p_k(t)$$

in which $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of A and each $p_j(t)$ is a polynomial whose degree is strictly less than the size of the largest Jordan block corresponding to the eigenvalue λ_j (that is, strictly less than the index of λ_j). Real eigenvalues are associated with terms that contain a real exponential factor, while non-real eigenvalues are associated with terms that contain an oscillatory factor and possibly also a real exponential factor.

3.2.3 Similarity of a matrix and its transpose. Let K_m be the m-by-m reversal matrix (0.9.5.1), which is symmetric and involutory: $K_m = K_m^T = K_m^{-1}$.

Exercise. Verify that $K_m J_m(\lambda) = J_m(\lambda)^T K_m$. Deduce that $K_m J_m(\lambda)$ is symmetric and $J_m(\lambda) = K_m^{-1} J_m(\lambda)^T K_m = K_m J_m(\lambda)^T K_m$.

The preceding exercise shows that each Jordan block is similar to its transpose via a reversal matrix. Therefore, if J is a given Jordan matrix (3.2.1.1), then J^T is similar to J via the symmetric involutory matrix $K = K_{n_1} \oplus \cdots \oplus K_{n_k}$: $J^T = KJK$. If $A = SJS^{-1}$, then $J = S^{-1}AS$,

$$\begin{aligned} A^T &= S^{-T} J^T S^T = S^{-T} KJKS^T = S^{-T} K(S^{-1}AS)KS^T \\ &= (S^{-T}KS^{-1})A(SKS^T) \end{aligned}$$

and SKS^T is symmetric. The conclusion is that every square complex matrix is similar to its transpose, and this similarity can be accomplished with a symmetric matrix. If A is nonderogatory, we can say more: *every* similarity between A and A^T must be via a symmetric matrix; see (3.2.4.4).

Moreover, we can write

$$A = SJS^{-1} = (SKS^T)(S^{-T}KJS^{-1})$$

in which KJ is symmetric. The conclusion is that every complex matrix is a product of two symmetric matrices.

For *any* field $\mathbf{F}$, it is also the case that every matrix in $M_n(\mathbf{F})$ is similar, via some symmetric matrix in $M_n(\mathbf{F})$, to its transpose.

3.2.4 Commutativity and nonderogatory matrices. For any polynomial $p(t)$ and any $A \in M_n$, $p(A)$ always commutes with A. What about the converse? If $A, B \in M_n$ are given and if A commutes with B, is there some polynomial $p(t)$ such that $B = p(A)$? Not always, for if we take $A = I$, then A commutes with every matrix and $p(I) = p(1)I$ is a scalar matrix; no non-scalar matrix can be a polynomial in I. The problem is that the form of A permits it to commute with many matrices, but permits it to generate only a limited set of matrices of the form $p(A)$.

What can we say if $A = J_m(\lambda)$ is a single Jordan block of size 2 or greater?

Exercise. Let $\lambda \in \mathbf{C}$ and an integer $m \geq 2$ be given. Show that $B \in M_m$ commutes with $J_m(\lambda)$ if and only if it commutes with $J_m(0)$. *Hint*: $J_m(\lambda) = \lambda I_m + J_m(0)$.

Exercise. Show that $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \in M_2$ commutes with $J_2(0)$ if and only if $b_{21} = 0$ and $b_{11} = b_{22}$; this is the case if and only if $B = b_{11}I_2 + b_{12}J_2(0)$, which is a polynomial in $J_2(0)$.

Exercise. Show that $B = [b_{ij}] \in M_3$ commutes with $J_3(0)$ if and only if B is upper triangular, $b_{11} = b_{22} = b_{33}$, and $b_{12} = b_{23}$; that is, if and only if B is an upper triangular Toeplitz matrix (0.9.7). This is the case if and only if $B = b_{11}I_3 + b_{12}J_3(0) + b_{13}J_3(0)^2$, which is a polynomial in $J_3(0)$.

Exercise. What can you say about $B = [b_{ij}] \in M_4$ if it commutes with $J_4(0)$?

3.2.4.1 Definition. A square complex matrix is *nonderogatory* if each of its eigenvalues has geometric multiplicity one.

Since the geometric multiplicity of a given eigenvalue of a Jordan matrix is equal to the number of Jordan blocks corresponding to that eigenvalue, a matrix is nonderogatory if and only if each of its distinct eigenvalues corresponds to exactly one block in its Jordan canonical form. Examples of nonderogatory matrices $A \in M_n$ are: any matrix with n distinct eigenvalues or any matrix with only one eigenvalue, which has geometric multiplicity one (that is, A is similar to a single Jordan block). A scalar matrix is the antithesis of a nonderogatory matrix.

Exercise. If $A \in M_n$ is nonderogatory, why is $\operatorname{rank} A \geq n - 1$?

3.2.4.2 Theorem. Suppose that $A \in M_n$ is nonderogatory. If $B \in M_n$ commutes with A, then there is a polynomial $p(t)$ of degree at most $n-1$ such that $B = p(A)$.

Proof: Let $A = SJ_AS^{-1}$ be the Jordan canonical form of A. If $BA = AB$, then $BSJ_AS^{-1} = SJ_AS^{-1}B$ and hence $(S^{-1}BS)J_A = J(S^{-1}BS)$. If we can show that $S^{-1}BS = p(J_A)$, then $B = Sp(J_A)S^{-1} = p(SJ_AS^{-1}) = p(A)$ is a polynomial in A. Thus, it suffices to assume that A is itself a Jordan matrix.

Assume that (a) $A = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k)$, in which $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct, and (b) A commutes with B. If we partition $B = [B_{ij}]_{i,j=1}^{k}$ conformally with J, then (2.4.4.2) ensures that $B = B_{11} \oplus \cdots \oplus B_{kk}$ is block diagonal. Moreover, $B_{ii}J_{n_i}(0) = J_{n_i}(0)B_{ii}$ for each $i = 1, 2, \ldots, k$. A computation reveals that each B_{ii} must be an upper triangular Toeplitz matrix (0.9.7), that is,

$$B_{ii} = \begin{bmatrix} b_1^{(i)} & b_2^{(i)} & \cdots & b_{n_i}^{(i)} \\ & \ddots & \ddots & \vdots \\ & & \ddots & b_2^{(i)} \\ & & & b_1^{(i)} \end{bmatrix} \tag{3.2.4.3}$$

which is a polynomial in $J_{n_i}(0)$, and hence also a polynomial in $J_{n_i}(\lambda)$:

$$\begin{aligned} B_{ii} &= b_1^{(i)} I_{n_i} + b_2^{(i)} J_{n_i}(0) + \cdots + b_{n_i}^{(i)} J_{n_i}(0)^{n_i-1} \\ &= b_1^{(i)} (J_{n_i}(\lambda) - \lambda_i I_{n_i})^0 + b_2^{(i)} (J_{n_i}(\lambda) - \lambda_i I_{n_i})^1 + \cdots + b_{n_i}^{(i)} (J_{n_i}(\lambda) - \lambda_i I_{n_i})^{n_i-1} \end{aligned}$$

If we can construct polynomials $p_i(t)$ of degree at most $n-1$ with the property that $p_i(J_{n_j}(\lambda_j)) = 0$ for all $i \neq j$, and $p_i(J_{n_i}(\lambda_i)) = B_{ii}$, then

$$p(t) = p_1(t) + \cdots + p_k(t)$$

fulfills the assertions of the theorem. Define

$$q_i(t) = \prod_{\substack{j=1 \\ j \neq i}}^{k} (t - \lambda_j)^{n_j}, \qquad \text{degree } q_i(t) = n - n_i$$

and observe that $q_i(J_{n_j}(\lambda_j)) = 0$ whenever $i \neq j$ because $(J_{n_j}(\lambda_j) - \lambda_j I)^{n_j} = 0$. The upper triangular Toeplitz matrix $q_i(J_{n_i}(\lambda_i))$ is nonsingular because its main diagonal entries $q_i(\lambda_i)$ are nonzero.

The key to our construction of the polynomials $p_i(t)$ is observing that the product of two upper triangular Toeplitz matrices is upper triangular Toeplitz, and the inverse of a nonsingular upper triangular Toeplitz matrix has the same form (0.9.7). Thus, $[q_i(J_{n_i}(\lambda_i))]^{-1}B_{ii}$ is an upper triangular Toeplitz matrix, which is therefore a polynomial in $J_{n_i}(\lambda_i)$:

$$[q_i(J_{n_i}(\lambda_i))]^{-1}B_{ii} = r_i(J_{n_i}(\lambda_i))$$

in which $r_i(t)$ is a polynomial of degree at most $n_i - 1$. The polynomial $p_i(t) = q_i(t)r_i(t)$ has degree at most $n - 1$,

$$p_i(J_{n_j}(\lambda_j)) = q_i(J_{n_j}(\lambda_j))r_i(J_{n_j}(\lambda_j)) = 0 \text{ whenever } i \neq j$$

and

$$\begin{aligned} p_i(J_{n_i}(\lambda_i)) &= q_i(J_{n_i}(\lambda_i))r_i(J_{n_i}(\lambda_i)) \\ &= q_i(J_{n_i}(\lambda_i))(q_i(J_{n_i}(\lambda_i))]^{-1}B_{ii}) = B_{ii} \end{aligned}$$

□

There is a converse to the preceding theorem; see Problem 2.

An illustrative application of (3.2.4.2) is the following strengthening of (3.2.3) in a special case.

3.2.4.4 Corollary. Let $A, B, S \in M_n$ be given and suppose that A is nonderogatory.
(a) If $AB = BA^T$ then B is symmetric.
(b) If S is nonsingular and $A^T = S^{-1}AS$, then S is symmetric.

Proof: (a) There is a symmetric nonsingular $R \in M_n$ such that $A^T = RAR^{-1}$ (3.2.3), so $AB = BA^T = BRAR^{-1}$ and hence $A(BR) = (BR)A$. Then (3.2.4.2) ensures that there is a polynomial $p(t)$ such that $BR = p(A)$. Compute $RB^T = (BR)^T = p(A)^T = p(A^T) = p(RAR^{-1}) = Rp(A)R^{-1} = R(BR)R^{-1} = RB$. Since R is nonsingular, it follows that $B^T = B$. (b) If $A^T = S^{-1}AS$ then $SA^T = AS$, so (a) ensures that S is symmetric. □

3.2.5 Convergent matrices. A matrix $A \in M_n$ with the property that all entries of A^m tend to zero as $m \to \infty$ is said to be *convergent*. Convergent matrices play an important role in the analysis of algorithms in numerical linear algebra. If A is a diagonal matrix, then A is convergent if and only if all the eigenvalues of A have modulus strictly less than 1; the same is true of non-diagonalizable matrices, but a careful analysis is required to come to this conclusion.

If $A = SJ_AS^{-1}$ is the Jordan canonical form of A, then $A^m = SJ_A^mS^{-1}$, so $A^m \to 0$ as $m \to \infty$ if and only if $J_A^m \to 0$ as $m \to \infty$. Since J_A is a direct sum of Jordan blocks, it suffices to consider the behavior of powers of a single Jordan block $J_k(\lambda) = \lambda I_k + J_k(0)$, which we can compute using the binomial theorem. We have $J_k(0)^m = 0$ for all $m \geq k$, so

$$\begin{aligned} J_k(\lambda)^m &= (\lambda I + J_k(0))^m = \sum_{j=0}^{m} \binom{m}{m-j} \lambda^{m-j} J_k(0)^j \\ &= \sum_{j=0}^{k-1} \binom{m}{m-j} \lambda^{m-j} J_k(0)^j \end{aligned}$$

for all $m \geq k$. The diagonal entries of $J_k(\lambda)^m$ are all equal to λ^m, so $J_k(\lambda)^m \to 0$ implies that $\lambda^m \to 0$, which means that $|\lambda| < 1$. Conversely, if $|\lambda| < 1$, it suffices to prove that

$$\binom{m}{m-j} \lambda^{m-j} \to 0 \text{ as } m \to \infty \text{ for each } j = 0, 1, 2, \ldots, k-1$$

There is nothing to prove if $\lambda = 0$ or $j = 0$, so suppose that $0 < |\lambda| < 1$ and $j \geq 1$; compute

$$\left| \binom{m}{m-j} \lambda^{m-j} \right| = \left| \frac{m(m-1)(m-2)\cdots(m-j+1)\lambda^m}{j!\lambda^j} \right| \leq \left| \frac{m^j \lambda^m}{j!\lambda^j} \right|$$

It suffices to show that $m^j |\lambda|^m \to 0$ as $m \to \infty$. One way to see this is to take logarithms and observe that $j \log m + m \log |\lambda| \to -\infty$ as $m \to \infty$ because $\log |\lambda| < 0$ and l'Hopital's rule ensures that $(\log m)/m \to 0$ as $m \to \infty$.

The preceding argument makes essential use of the Jordan canonical form of A to show that $A^m \to 0$ as $m \to \infty$ if and only if all the eigenvalues of A have modulus strictly less than 1. Another proof, which is completely independent of the Jordan canonical form, is given in (5.6.12).

3.2.6 The geometric multiplicity–algebraic multiplicity inequality. The geometric multiplicity of an eigenvalue λ of a given $A \in M_n$ is the number of Jordan blocks of A corresponding to λ. This number is less than or equal to the sum of the sizes of all the Jordan blocks corresponding to λ; this sum is the algebraic multiplicity of λ. Thus, the geometric multiplicity of an eigenvalue is not greater than its algebraic multiplicity. We have already discussed this fundamental inequality from very different points of view: see (1.2.18), (1.3.7), and (1.4.10).

3.2.7 Diagonalizable + nilpotent: the Jordan decomposition. For any Jordan block, we have the identity $J_k(\lambda) = \lambda I_k + J_k(0)$, and $J_k(0)^k = 0$. Thus, any Jordan block is the sum of a diagonal matrix and a nilpotent matrix.

More generally, a Jordan matrix (3.2.1.1) can be written as $J = D + N$, in which D is a diagonal matrix whose main diagonal is the same as that of J, and $N = J - D$. The matrix N is nilpotent, and $N^k = 0$ if k is the size of the largest Jordan block in J, which is the index of 0 as an eigenvalue of N.

Finally, if $A \in M_n$ and $A = SJ_AS^{-1}$ is its Jordan canonical form, then $A = S(D+N)S^{-1} = SDS^{-1} + SNS^{-1} \equiv A_D + A_N$, in which A_D is diagonalizable and A_N is nilpotent. Moreover, $A_DA_N = A_NA_D$ because both D and N are conformal block diagonal matrices, and the diagonal blocks in D are scalar matrices. Of course, A_D and A_N also commute with $A = A_D + A_N$.

The preceding discussion establishes the existence of a *Jordan decomposition*: any square complex matrix is a sum of two commuting matrices, one of which is diagonalizable and the other is nilpotent. For the uniqueness of the Jordan decomposition, see Problem 18.

3.2.8 The Jordan canonical form of a direct sum. Let $A_i \in M_{n_i}$ be given for $i = 1, \ldots, m$ and suppose that each $A_i = S_iJ_iS_i^{-1}$, in which each J_i is a Jordan matrix. Then the direct sum $A = A_1 \oplus \cdots \oplus A_m$ is similar to the direct sum $J = J_1 \oplus \cdots \oplus J_m$ via $S = S_1 \oplus \cdots \oplus S_m$. Moreover, J is a direct sum of direct sums of Jordan blocks, so it is a Jordan matrix and hence uniqueness of the Jordan canonical form ensures that it is the Jordan canonical form of A.

3.2.9 An optimality property of the Jordan canonical form. The Jordan canonical form of a matrix is a direct sum of upper triangular matrices that have nonzero off-diagonal entries only in the first superdiagonal, so it has many zero entries. However, among all the matrices that are similar to a given matrix, the Jordan canonical form need not have the smallest number of nonzero entries. For example,

$$A = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.2.9.1}$$

has 5 nonzero entries, but its Jordan canonical form $J = J_2(1) \oplus J_2(-1)$ has 6 nonzero entries. However, A has 5 nonzero *off-diagonal* entries, while J has only 2 nonzero off-diagonal entries. We now explain why no matrix similar to A can have fewer than 2 nonzero off-diagonal entries.

3.2.9.2 Observation. Suppose that $B = [b_{ij}] \in M_m$ has fewer than $m - 1$

nonzero off-diagonal entries. Then there exists a permutation matrix P such that $P^TBP = B_1 \oplus B_2$ in which each $B_i \in M_{n_i}$ and each $n_i \geq 1$.

Why is this? Here is an informal argument that can be made precise: Consider m islands $C_1, \ldots, C_m$ located near each other in the sea. There is a footbridge between two different islands C_i and C_j if and only if $i \neq j$ and either $b_{ij} \neq 0$ or $b_{ji} \neq 0$. Suppose that $C_1, C_{j_2}, \ldots, C_{j_\nu}$ are all the different islands that one can walk to starting from C_1. The minimum number of bridges required to link up all the islands is $m-1$. We are assuming that there are fewer than $m-1$ bridges, so $\nu < m$. Relabel all the islands (1 through m again) in any way that gives the new labels $1, 2, \ldots, \nu$ to $C_1, C_{j_2}, \ldots, C_{j_\nu}$. Let $P \in M_m$ be the permutation matrix corresponding to the relabeling. Then $P^TBP = B_1 \oplus B_2$, in which $B_1 \in M_\nu$. The direct sum structure reflects the fact that no bridge joins any of the first (relabeled) ν islands to any of the remaining $n - \nu$ islands.

We say that a given $B \in M_m$ is *indecomposable under permutation similarity* if there is no permutation matrix P such that $P^TBP = B_1 \oplus B_2$, in which each $B_i \in M_{n_i}$ and each $n_i \geq 1$. Then (3.2.9.2) says that if $B \in M_m$ is indecomposable under permutation similarity, it has at least $m-1$ nonzero off-diagonal entries.

3.2.9.3 Observation. Any given $B \in M_n$ is permutation similar to a direct sum of matrices that are indecomposable under permutation similarity.

Proof: Consider the finite set $\mathcal{S} = \{P^TBP : P \in M_n \text{ is a permutation matrix}\}$. Some of the elements of $\mathcal{S}$ are block diagonal (take $P = I_n$, for example). Let q be the largest positive integer such that B is permutation similar to $B_1 \oplus \cdots \oplus B_q$, each $B_i \in M_{n_i}$, and each $n_i \geq 1$; maximality of q ensures that no direct summand B_i is decomposable under permutation similarity. □

The number of nonzero off-diagonal entries in a square matrix is not changed by a permutation similarity, so we can combine the two preceding observations to obtain a lower bound on the number of Jordan blocks in the Jordan canonical form of a matrix.

3.2.9.4 Observation. Suppose that a given $B \in M_n$ has p nonzero off-diagonal entries, and that its Jordan canonical form J_B contains r Jordan blocks. Then $r \geq n - p$.

Proof: Suppose that B is permutation similar to $B_1 \oplus \cdots \oplus B_q$, in which each $B_i \in M_{n_i}$ is indecomposable under permutation similarity, and each $n_i \geq 1$. The number of nonzero off-diagonal entries in B_i is at least $n_i - 1$, so the number of nonzero off-diagonal entries in B is at least $(n_1 - 1) + \cdots + (n_q - 1) = n - q$. That is, $p \geq n - q$, so $q \geq n - p$. But (3.2.8) ensures that J_B contains at least q Jordan blocks, so $r \geq q \geq n - p$. □

Our final observation is that the number of nonzero off-diagonal entries in an n-by-n Jordan matrix $J = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_r}(\lambda_r)$ is exactly $(n_1 - 1) + \cdots + (n_r - 1) = n - r$.

3.2.9.5 Theorem. Let $A, B \in M_n$ be given. Suppose that B has exactly p nonzero off-diagonal entries and is similar to A. Let J_A be the Jordan canonical form of A and suppose that J_A consists of r Jordan blocks. Then $p \geq n - r$, which is the number of nonzero off-diagonal entries of J_A.

Proof: Since B is similar to A, J_A is also the Jordan canonical form of B, and (3.2.9.4) ensures that $r \geq n - p$, so $p \geq n - r$. □

3.2.10 The index of an eigenvalue of a block upper triangular matrix. The *index* of an eigenvalue λ of $A \in M_n$ (the index of λ in A) is, equivalently, (a) the size of the largest Jordan block of A with eigenvalue λ or (b) the smallest value of $m = 1, 2, \ldots, n$ such that $\operatorname{rank}(A - \lambda I)^m = \operatorname{rank}(A - \lambda I)^{m+1}$ (and hence $\operatorname{rank}(A - \lambda I)^m = \operatorname{rank}(A - \lambda I)^{m+k}$ for all $k = 1, 2, \ldots$). If the index of λ in $A_{11} \in M_{n_1}$ is ν_1 and the index of λ in $A_{22} \in M_{n_2}$ is ν_2, then the index of λ in the direct sum $A_{11} \oplus A_{22}$ is $\max\{\nu_1, \nu_2\}$.

Exercise. Consider $A = \begin{bmatrix} J_2(0) & I_2 \\ 0 & J_2(0)^T \end{bmatrix}$, so the index of the eigenvalue 0 in each diagonal block is 2. Show that the index of 0 as an eigenvalue of A is 4.

If λ is an eigenvalue of A_{11} or A_{22} in the block upper triangular matrix $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, and if $A_{12} \neq 0$, what can we say about the index of λ as an eigenvalue of A? For convenience, take $\lambda = 0$. Let the index of λ in $A_{11} \in M_{n_1}$ be ν_1 and let the index of λ in $A_{22} \in M_{n_2}$ be ν_2. Any power of A is block upper triangular,

$$A^m = \begin{bmatrix} A_{11}^m & \sum_{k=0}^{m} A_{11}^k A_{12} A_{22}^{m-k} \\ 0 & A_{22}^m \end{bmatrix}$$

the rank of A^m is at least the sum of the ranks of its diagonal blocks (0.9.4), and we have the lower bound

$$\begin{aligned} \operatorname{rank} A^{\nu_1+\nu_2+1} &\geq \operatorname{rank} A_{11}^{\nu_1+\nu_2+1} + \operatorname{rank} A_{22}^{\nu_1+\nu_2+1} \\ &= \operatorname{rank} A_{11}^{\nu_1} + \operatorname{rank} A_{22}^{\nu_2} \end{aligned}$$

Now compute

$$\begin{aligned}
A^{\nu_1+\nu_2} &= \begin{bmatrix} A_{11}^{\nu_1+\nu_2} & \sum_{k=0}^{\nu_1+\nu_2} A_{11}^k A_{12} A_{22}^{\nu_1+\nu_2-k} \\ 0 & A_{22}^{\nu_1+\nu_2} \end{bmatrix} \\
&= \begin{bmatrix} A_{11}^{\nu_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_{11}^{\nu_2} & \sum_{k=0}^{\nu_2} A_{11}^k A_{12} A_{22}^{\nu_2-k} \\ 0 & 0 \end{bmatrix} \\
&\quad + \begin{bmatrix} 0 & \sum_{k=0}^{\nu_1-1} A_{11}^k A_{12} A_{22}^{\nu_1-k} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & A_{22}^{\nu_2} \end{bmatrix}
\end{aligned}$$

The basic rank inequalities (0.4.5c,d) now imply the upper bound

$$\begin{aligned}
\operatorname{rank} A^{\nu_1+\nu_2} &\leq \operatorname{rank} \begin{bmatrix} A_{11}^{\nu_1} & 0 \\ 0 & 0 \end{bmatrix} + \operatorname{rank} \begin{bmatrix} 0 & 0 \\ 0 & A_{22}^{\nu_2} \end{bmatrix} \\
&= \operatorname{rank} A_{11}^{\nu_1} + \operatorname{rank} A_{22}^{\nu_2}
\end{aligned}$$

Combining our upper and lower bounds gives

$$\begin{aligned}
\operatorname{rank} A^{\nu_1+\nu_2} &\leq \operatorname{rank} A_{11}^{\nu_1} + \operatorname{rank} A_{22}^{\nu_2} \leq \operatorname{rank} A^{\nu_1+\nu_2+1} \\
&\leq \operatorname{rank} A^{\nu_1+\nu_2}
\end{aligned}$$

which tells us that $\operatorname{rank} A^{\nu_1+\nu_2} = \operatorname{rank} A^{\nu_1+\nu_2+1}$. The conclusion is that the index of 0 in A is at most $\nu_1 + \nu_2$. An induction permits us to extend this conclusion to any block upper triangular matrix.

Theorem 3.2.10.1. Let $A = [A_{ij}]_{i,j=1}^p \in M_n$ be block upper triangular, so each A_{ii} is square and $A_{ij} = 0$ for all $i > j$. Suppose that the index of λ as an eigenvalue of each diagonal block A_{ii} is ν_i, $i = 1, \ldots, p$. Then the index of λ as an eigenvalue of A is at most $\nu_1 + \cdots + \nu_p$.

Exercise. Provide details for the induction required to prove the preceding theorem.

Corollary 3.2.10.2. Let $\lambda \in \mathbf{C}$, let $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & \lambda I_{n_2} \end{bmatrix}$, and suppose that $A_{11} \in M_{n_1}$ is diagonalizable. Then every Jordan block of A with eigenvalue different from λ is 1-by-1, and every Jordan block of A with eigenvalue λ is either 1-by-1 or 2-by-2.

3.2.11 AB **vs.** BA**.** If $A \in M_{m,n}$ and $B \in M_{n,m}$, (1.3.22) ensures that the nonzero eigenvalues of AB and BA are the same, including their multiplicities. In fact, we can make a much stronger statement: the nonsingular parts of the Jordan canonical forms of AB and BA are identical.

Theorem 3.2.11.1. Suppose that $A \in M_{m,n}$ and $B \in M_{n,m}$. For each nonzero eigenvalue λ of AB and for each $k = 1, 2, \ldots$, the respective Jordan canonical forms of AB and BA contain the same number of Jordan blocks $J_k(\lambda)$.

Proof: In the proof of Theorem (1.3.22), we found that $C_1 = \left[\begin{smallmatrix} AB & 0 \\ B & 0_n \end{smallmatrix}\right]$ and $C_2 = \left[\begin{smallmatrix} 0_m & 0 \\ B & BA \end{smallmatrix}\right]$ are similar. Let $\lambda \neq 0$ be given and let k be any given positive integer. First observe that the row rank of

$$(C_1 - \lambda I_{m+n})^k = \begin{bmatrix} (AB - \lambda I_m)^k & 0 \\ \bigstar & (-\lambda I_n)^k \end{bmatrix}$$

is $n + \operatorname{rank}((AB - \lambda I_m)^k)$, then observe that the column rank of

$$(C_2 - \lambda I_{m+n})^k = \begin{bmatrix} (-\lambda I_m)^k & 0 \\ \bigstar & (BA - \lambda I_n)^k \end{bmatrix}$$

is $m + \operatorname{rank}((BA - \lambda I_n)^k)$. But $(C_1 - \lambda I_{m+n})^k$ is similar to $(C_2 - \lambda I_{m+n})^k$, so their ranks are equal, that is,

$$\operatorname{rank}((AB - \lambda I_m)^k) = \operatorname{rank}((BA - \lambda I_n)^k) + m - n$$

for each $k = 1, 2, \ldots$, which implies that

$$\begin{aligned} & \operatorname{rank}((AB - \lambda I_m)^{k-1}) - \operatorname{rank}((AB - \lambda I_m)^k) \\ = \; & \operatorname{rank}((BA - \lambda I_n)^{k-1}) - \operatorname{rank}((BA - \lambda I_n)^k) \end{aligned}$$

for each $k = 1, 2, \ldots$. Thus, the respective Weyr characteristics of AB and BA associated with any given nonzero eigenvalue λ of AB are identical, so (3.1.18) ensures that their respective Jordan canonical forms contain exactly the same number of blocks $J_k(\lambda)$ for each $k = 1, 2, \ldots$. □

3.2.12 The Drazin inverse A singular matrix does not have an inverse, but several types of *generalized inverse* are available, each of which has some (but of course not all) features of the ordinary inverse. The generalized inverse that we consider in this section is the *Drazin inverse*.

3.2.12.1 Definition. Let $A \in M_n$ and suppose that

$$A = S \begin{bmatrix} B & 0 \\ 0 & N \end{bmatrix} S^{-1} \tag{3.2.12.2}$$

in which S and B are square and nonsingular, and N is nilpotent. The direct summand B is absent if A is nilpotent; N is absent if A is nonsingular. The *Drazin inverse* of A is

$$A^D = S \begin{bmatrix} B^{-1} & 0 \\ 0 & 0 \end{bmatrix} S^{-1} \tag{3.2.12.3}$$

Every $A \in M_n$ has a representation of the form (3.2.12.2): use the Jordan canonical form (3.1.12) in which B is a direct sum of all the nonsingular Jordan blocks of A and N is a direct sum of all the nilpotent blocks.

In addition to (3.2.12.2), suppose that A is represented as

$$A = T \begin{bmatrix} C & 0 \\ 0 & N' \end{bmatrix} T^{-1} \tag{3.2.12.4}$$

in which T and C are square and nonsingular, and N' is nilpotent. Then $A^n = S\left[\begin{smallmatrix} B^n & 0 \\ 0 & 0 \end{smallmatrix}\right]S^{-1} = T\left[\begin{smallmatrix} C^n & 0 \\ 0 & 0 \end{smallmatrix}\right]T^{-1}$, so $\operatorname{rank} A^n = \operatorname{rank} B^n = \operatorname{rank} B$ is the size of B since it is nonsingular; for the same reason, it is also the size of C. We conclude that B and C have the same size, and hence N and N' have the same size. Since $A = S\left[\begin{smallmatrix} B & 0 \\ 0 & N \end{smallmatrix}\right]S^{-1} = T\left[\begin{smallmatrix} C & 0 \\ 0 & N' \end{smallmatrix}\right]T^{-1}$, it follows that $R\left[\begin{smallmatrix} B & 0 \\ 0 & N \end{smallmatrix}\right] = \left[\begin{smallmatrix} C & 0 \\ 0 & N' \end{smallmatrix}\right]R$, in which $R = T^{-1}S$. Partition $R = [R_{ij}]_{i,j=1}^2$ conformally with $\left[\begin{smallmatrix} B & 0 \\ 0 & N \end{smallmatrix}\right]$. Then (2.4.4.2) ensures that $R_{12} = 0$ and $R_{21} = 0$, so $R = R_{11} \oplus R_{22}$, R_{11} and R_{22} are nonsingular, $C = R_{11}BR_{11}^{-1}$, $N' = R_{22}NR_{22}^{-1}$, and $T = SR^{-1}$. Finally, compute the Drazin inverse using (3.2.12.4):

$$\begin{aligned} T \begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} T^{-1} &= SR^{-1} \begin{bmatrix} \left(R_{11}BR_{11}^{-1}\right)^{-1} & 0 \\ 0 & 0 \end{bmatrix} RS^{-1} \\ &= S \begin{bmatrix} R_{11}^{-1} & 0 \\ 0 & R_{22}^{-1} \end{bmatrix} \begin{bmatrix} R_{11}B^{-1}R_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix} S^{-1} \\ &= S \begin{bmatrix} B^{-1} & 0 \\ 0 & 0 \end{bmatrix} S^{-1} = A^D \end{aligned}$$

We conclude that the Drazin inverse is well defined by (3.2.12.3).

Exercise. Explain why $A^D = A^{-1}$ if A is nonsingular.

Let q be the index of the eigenvalue 0 of A and consider the three identities

$$AX = XA \tag{3.2.12.5}$$

$$A^{q+1}X = A^q \tag{3.2.12.6}$$

$$XAX = X \tag{3.2.12.7}$$

Exercise. Use (3.2.12.2) and (3.2.12.3) to explain why A and $X = A^D$ satisfy the preceding three identities if and only if $A = \left[\begin{smallmatrix} B & 0 \\ 0 & N \end{smallmatrix}\right]$ and $X = \left[\begin{smallmatrix} B^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right]$ satisfy them. Verify that they do.

There is a converse to the result in the preceding exercise: *If X satisfies*

(3.2.12.5-7), then $X = A^D$. To verify this assertion, proceed as in the Exercise to replace A by $\begin{bmatrix} B & 0 \\ 0 & N \end{bmatrix}$ and partition the unknown matrix $X = [X_{ij}]_{i,j=1}^2$ conformally. We must show that $X_{11} = B^{-1}$ and that X_{12}, X_{21}, and X_{22} are zero blocks. Combining the first identity (3.2.12.5) with (2.4.4.2) ensures that $X_{12} = 0$ and $X_{21} = 0$; in addition, $NX_{22} = X_{22}N$. The second identity (3.2.12.6) says that $\begin{bmatrix} B^{q+1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{11} & 0 \\ 0 & X_{22} \end{bmatrix} = \begin{bmatrix} B^q & 0 \\ 0 & 0 \end{bmatrix}$, so $B^{q+1}X_{11} = B^q$, $BX_{11} = I$, and $X_{11} = B^{-1}$. The third identity (3.2.12.7) ensures that

$$X_{22} = X_{22}NX_{22} = NX_{22}^2 \qquad (3.2.12.8)$$

which implies that $N^{q-1}X_{22} = N^{q-1}NX_{22}^2 = N^qX_{22}^2 = 0$, so $N^{q-1}X_{22} = 0$. Using (3.2.12.8) again, we see that $N^{q-2}X_{22} = N^{q-2}NX_{22}^2 = (N^{q-1}X_{22})X_{22} = 0$, so $N^{q-2}X_{22} = 0$. Continuing this argument reveals that $N^{q-3}X_{22} = 0, \ldots, NX_{22} = 0$, and finally $X_{22} = 0$.

Our last observation is that *the Drazin inverse A^D is a polynomial in A.*

Exercise. Represent A as in (3.2.12.2). According to (2.4.3.4) there is a polynomial $p(t)$ such that $p(B^{q+1}) = (B^{q+1})^{-1}$. Let $g(t) = t^q p(t^{q+1})$. Verify that $g(A) = A^D$.

Exercise. Let $A \in M_n$ and suppose λ is a nonzero eigenvalue of A. If $x \neq 0$ and $Ax = \lambda x$, explain why $A^D x = \lambda^{-1} x$.

Problems

1. Let $\mathcal{F} = \{A_\alpha : \alpha \in \mathcal{I}\} \subset M_n$ be a given family of matrices, indexed by the index set $\mathcal{I}$, and suppose there is a nonderogatory matrix $A_0 \in \mathcal{F}$ such that $A_\alpha A_0 = A_\alpha A_0$ for all $\alpha \in \mathcal{I}$. Show that for every $\alpha \in \mathcal{I}$ there is a polynomial $p_\alpha(t)$ of degree at most $n-1$ such that $A_\alpha = p_\alpha(A_0)$, and hence $\mathcal{F}$ is a commuting family.

2. Let $A \in M_n$. If every matrix that commutes with A is a polynomial in A, show that A is nonderogatory. *Hint*: Why does it suffice to consider the case in which A is a Jordan matrix? Suppose that $A = J_k(\lambda) \oplus J_\ell(\lambda) \oplus J$, in which J is either empty or is a Jordan matrix and $k, \ell \geq 1$. For any polynomial $p(t)$ the leading $k + \ell$ diagonal entries of $p(A)$ are all equal to $p(\lambda)$. But $-I_k \oplus I_\ell \oplus I_{n-k-\ell}$ commutes with A.

3. Let $A \in M_n$. Show that there is a bounded set containing all of the entries of the family $\{A^m : m = 1, 2, \ldots\}$ (that is, A is *power-bounded*) if and only if every eigenvalue λ of A satisfies (a) $|\lambda| \leq 1$ and (b) if $|\lambda| = 1$ then no Jordan block of A with eigenvalue λ has size greater than 1 (that is, every eigenvalue with modulus 1 is semisimple). *Hint*: $J_2(\lambda)^m = \begin{bmatrix} \lambda^m & m \\ 1 & \lambda^m \end{bmatrix}$.

4. Suppose $A \in M_n$ is singular and let $r = \operatorname{rank} A$. In Problem 28 of (2.4) we learned that there is a polynomial of degree $r+1$ that annihilates A. Provide details for the following argument to show that $h(t) = p_A(t)/t^{n-r-1}$ is such a polynomial. Let the Jordan canonical form of A be $J \oplus J_{n_1}(0) \oplus \cdots \oplus J_{n_k}(0)$, in which the Jordan matrix J is nonsingular. Let $\nu = n_1 + \cdots + n_k$ and let $n_{\max} = \max_i n_i$ be the index of the eigenvalue zero. (a) Explain why $p_A(t) = p_1(t)t^{\nu}$, in which $p_1(t)$ is a polynomial and $p_1(0) \neq 0$. (b) Show that $p(t) = p_1(t)t^{n_{\max}}$ annihilates A, so $p_A(t) = (p_1(t)t^{n_{\max}})t^{\nu - n_{\max}}$ (c) Explain why $k = n - r$, $\nu - n_{\max} \geq k - 1 = n - r - 1$, and $h(A) = 0$.

5. What is the Jordan canonical form of $A = \begin{bmatrix} i & 1 \\ 1 & -i \end{bmatrix}$?

6. The linear transformation $d/dt : p(t) \to p'(t)$ acting on the vector space of all polynomials with degree at most 3 has the basis representation

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

in the basis $B = \{1, t, t^2, t^3\}$. What is the Jordan canonical form of this matrix?

7. What are the possible Jordan forms of a matrix $A \in M_n$ such that $A^3 = I$?

8. What are the possible Jordan canonical forms for a matrix $A \in M_6$ with characteristic polynomial $p_A(t) = (t+3)^4(t-4)^2$?

9. Suppose that $k \geq 2$. Explain why the Jordan canonical form of $\operatorname{adj} J_k(\lambda)$ is $J_k(\lambda^{k-1})$ if $\lambda \neq 0$, and it is $J_2(0) \oplus 0_{k-2}$ if $\lambda = 0$.

10. Suppose that the Jordan canonical form of a given nonsingular $A \in M_n$ is $J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k)$. Explain why the Jordan canonical form of $\operatorname{adj} A$ is $J_{n_1}(\mu_1) \oplus \cdots \oplus J_{n_k}(\mu_k)$, in which each $\mu_i = \lambda_i^{n_i - 1} \prod_{j \neq i} \lambda_j^{n_j}$, $i = 1, \ldots, k$.

11. Suppose that the Jordan canonical form of a given singular $A \in M_n$ is $J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_{k-1}}(\lambda_{k-1}) \oplus J_{n_k}(0)$. Explain why the Jordan canonical form of $\operatorname{adj} A$ is $J_2(0) \oplus 0_{n-2}$ if $n_k \geq 2$, and it is $\prod_{i=1}^{k-1} \lambda_i^{n_i} \oplus 0_{n-1}$ if $n_k = 1$; the former case is characterized by $\operatorname{rank} A < n-1$ and the latter case is characterized by $\operatorname{rank} A = n - 1$.

12. Explain why $\operatorname{adj} A = 0$ if the Jordan canonical form of A contains two or more singular Jordan blocks.

13. Let $A \in M_n$ and $B, C \in M_m$ be given. Show that $\left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right] \in M_{n+m}$ is similar to $\left[\begin{smallmatrix} A & 0 \\ 0 & C \end{smallmatrix}\right]$ if and only if B is similar to C.

14. Let $B, C \in M_m$ and a positive integer k be given. Show that

$$\underbrace{B \oplus \cdots \oplus B}_{k \text{ summands}} \quad \text{and} \quad \underbrace{C \oplus \cdots \oplus C}_{k \text{ summands}}$$

are similar if and only if B and C are similar.

15. Let $A \in M_n$ and $B, C \in M_m$ be given. Show that

$$A \oplus \underbrace{B \oplus \cdots \oplus B}_{k \text{ summands}} \quad \text{and} \quad A \oplus \underbrace{C \oplus \cdots \oplus C}_{k \text{ summands}}$$

are similar if and only if B and C are similar. *Hint:* Use the two preceding problems.

16. Let $A \in M_n$ have Jordan canonical form $J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k)$. If A is nonsingular, show that the Jordan canonical form of A^2 is $J_{n_1}(\lambda_1^2) \oplus \cdots \oplus J_{n_k}(\lambda_k^2)$; that is, the Jordan canonical form of A^2 is composed of precisely the same collection of Jordan blocks as A, but the respective eigenvalues are squared. However, the Jordan canonical form of $J_m(0)^2$ is not $J_m(0^2)$ if $m \geq 2$; explain.

17. Let $A \in M_n$ be given. Show that $\operatorname{rank} A = \operatorname{rank} A^2$ if and only if the geometric and algebraic multiplicities of the eigenvalue $\lambda = 0$ are equal; that is, if and only if all the Jordan blocks corresponding to $\lambda = 0$ (if any) in the Jordan canonical form of A are 1-by-1. Explain why A is diagonalizable if and only if $\operatorname{rank}(A - \lambda I) = \operatorname{rank}(A - \lambda I)^2$ for all $\lambda \in \sigma(A)$.

18. Let $A \in M_n$ be given. In (3.2.7) we used the Jordan canonical form to write A as a sum of two commuting matrices, one of which is diagonalizable and the other is nilpotent: the Jordan decomposition $A = A_D + A_N$. The goal of this problem is to show that *the Jordan decomposition is unique*. That is, suppose that (a) $A = B + C$, (b) B commutes with C, (c) B is diagonalizable, and (d) C is nilpotent; we claim that $B = A_D$ and $C = A_N$. It is helpful to use the fact that there are polynomials $p(t)$ and $q(t)$ such that $A_D = p(A)$ and $A_N = q(A)$; see Problem 14(d) in Section 6.1 of [HJ]. Provide details for the following: (a) B and C commute with A. (b) B and C commute with A_D and A_N. (c) B and A_D are simultaneously diagonalizable, so $A_D - B$ is diagonalizable. (d) C and A_N are simultaneously upper triangularizable, so $C - A_N$ is nilpotent. (e) $A_D - B = C - A_N$ is both diagonalizable and nilpotent, so it is a zero matrix. The (uniquely determined) matrix A_D is called the *diagonalizable part* of A; A_N is the *nilpotent part* of A.

19. Let $A \in M_n$ be given and let λ be an eigenvalue of A. (a) Prove that the following two assertions are equivalent: (i) Every Jordan block of A with eigenvalue λ has size two or greater; (ii) Every eigenvector of A corresponding to λ is in the range of $A - \lambda I$. (b) Prove that the following five assertions are equivalent: (i) Some Jordan block of A is 1-by-1; (ii) There is a nonzero vector x such that $Ax = \lambda x$ but x is not in the range of $A - \lambda I$; (iii) There is a nonzero vector x such that $Ax = \lambda x$ but x is not orthogonal to the null space of $A^* - \bar{\lambda} I$; (iv) There are nonzero vectors x and y such that $Ax = \lambda x$, $y^* A = \lambda y^*$, and $x^* y \neq 0$; (v) A is similar to $[\lambda] \oplus B$ for some $B \in M_{n-1}$. *Hint*: (1.4.7).

20. Let $A, B \in M_n$ be given. Show that AB is similar to BA if and only $\operatorname{rank}(AB)^k = \operatorname{rank}(BA)^k$ for each $k = 1, 2, \ldots, n$.

21. Let $A = \left[\begin{smallmatrix} J_2(0) & 0 \\ x^T & 0 \end{smallmatrix}\right] \in M_3$ with $x^T = [1\ 0]$, and let $B = I_2 \oplus [0] \in M_3$. Show that the Jordan canonical form of AB is $J_3(0)$, while that of BA is $J_2(0) \oplus J_1(0)$.

22. Let $A \in M_n$. Show that both AA^D and $I - AA^D$ are projections (idempotents), and that $AA^D(I - AA^D) = 0$.

23. Let $A \in M_n$ and let q be the index of 0 as an eigenvalue of A. Show that $A^D = \lim_{t \to 0}(A^{k+1} + tI)^{-1} A^k$ for any $k \geq q$.

24. Let $A, B \in M_n$, let $D = AB - BA^T$, and suppose that $AD = DA^T$. Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of A. (a) (an analog of Problem 12(c) in (2.4)) If A is diagonalizable, show that $D = 0$, that is, $AB = BA^T$. *Hint*: Let $A = S\Lambda S^{-1}$ with $\Lambda = \lambda_1 I_{n_1} \oplus \cdots \oplus \lambda_d I_{n_d}$; let $\mathcal{D} = S^{-1} D S$ and $\mathcal{B} = S^{-1} B S^{-T}$. Then $\Lambda \mathcal{D} = \mathcal{D} \Lambda$ and $\mathcal{D} = \Lambda \mathcal{B} - \mathcal{B} \Lambda$. Conclude that $\mathcal{D}$ and $\mathcal{B}$ are block diagonal conformal to Λ and $\mathcal{D} = 0$. (b) Suppose that A is nonderogatory. Then (3.2.4.4) ensures that D is symmetric. In addition, show that $\operatorname{rank} D \leq n - d$, so the geometric multiplicity of 0 as an eigenvalue of D is at least d. *Hint*: Let $A = SJS^{-1}$ with $J = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_d}(\lambda_d)$; let $\mathcal{D} = S^{-1} D S$ and $\mathcal{B} = S^{-1} B S^{-T}$. Then $J\mathcal{D} = \mathcal{D} J^T$ and $\mathcal{D} = J\mathcal{B} - \mathcal{B} J^T$. Conclude that $\mathcal{D} = \mathcal{D}_1 \oplus \cdots \oplus \mathcal{D}_d$ and $\mathcal{B} = \mathcal{B}_1 \oplus \cdots \oplus \mathcal{B}_d$ are block diagonal conformal to J, $J_i \mathcal{D}_i = \mathcal{D}_i J_i^T$, and $\mathcal{D}_i = J_i \mathcal{B}_i - \mathcal{B}_i J_i^T$ for each $i = 1, \ldots, d$ ($J_i := J_{n_i}(\lambda_i)$). Let $J_i^T = S_i J_i S_i^{-1}$. Then $J_i(\mathcal{D}_i S_i) = (\mathcal{D}_i S_i) J_i$, and $(\mathcal{D}_i S_i) = J_i(\mathcal{B}_i S_i) - (\mathcal{B}_i S_i) J_i$. Invoke Jacobson's lemma.

25. Let $A \in M_n$ be given and suppose that A^2 is nonderogatory. Explain why: (a) A is nonderogatory; (b) If λ is a nonzero eigenvalue of A, then $-\lambda$ is not an eigenvalue of A; (c) If A is singular, then 0 has algebraic multiplicity one as an eigenvalue of A; (d) $\operatorname{rank} A \geq n-1$; (e) There is a polynomial $p(t)$ such that $A = p(A^2)$.

26. Let $A, B \in M_n$ be given and suppose that A^2 is nonderogatory. If $AB = B^T A$ and $BA = AB^T$, show that B is symmetric. *Hint*: Show that $(B - B^T)A = 0$ and explain why $\operatorname{rank}(B - B^T) \leq 1$. See Problem 27 in (2.6).

Notes and Further Readings. For a detailed discussion of the optimality property (3.2.9.4) and a characterization of the case of equality, see R. Brualdi, P. Pei, and X. Zhan, An extremal sparsity property of the Jordan canonical form, *Linear Algebra Appl.* 429 (2008) 2367-2372. Problem 21 illustrates that the nilpotent Jordan structures of AB and BA need not be the same, but in the following sense they cannot differ by much: If $m_1 \geq m_2 \geq \cdots$ are the sizes of the nilpotent Jordan blocks of AB while $n_1 \geq n_2 \geq \cdots$ are the sizes of the nilpotent Jordan blocks of BA (append zero sizes to one list or the other, if necessary to achieve lists of equal length) then $|m_i - n_i| \leq 1$ for all i. For a discussion and proof, see C. Johnson and E. Schreiner, The relationship between AB and BA, *Amer. Math. Monthly* 103 (1996) 578-582.

3.3 The minimal polynomial and the companion matrix

A polynomial $p(t)$ is said to *annihilate* $A \in M_n$ if $p(A) = 0$. The Cayley–Hamilton theorem (2.4.2) guarantees that for each $A \in M_n$ there is a monic polynomial $p_A(t)$ of degree n (the characteristic polynomial) such that $p_A(A) = 0$. Of course, there may be a monic polynomial of degree $n-1$ that annihilates A, or one of degree $n-2$ or less. Of special interest is a monic polynomial of minimum degree that annihilates A. It is clear that such a polynomial exists; the following theorem says that it is unique.

3.3.1 Theorem. Let $A \in M_n$ be given. There exists a unique monic polynomial $q_A(t)$ of minimum degree that annihilates A. The degree of $q_A(t)$ is at most n. If $p(t)$ is any monic polynomial such that $p(A) = 0$, then $q_A(t)$ divides $p(t)$, that is, $p(t) = h(t)q_A(t)$ for some monic polynomial $h(t)$.

Proof: The set of monic polynomials that annihilate A contains $p_A(t)$, which has degree n. Let $m = \min\{k : p(t)$ is a monic polynomial of degree k and $p(A) = 0\}$; necessarily $m \leq n$. If $p(t)$ is any monic polynomial that annihilates A, and if $q(t)$ is a monic polynomial of degree m that annihilates A, then the degree of $p(t)$ is m or greater. The Euclidean algorithm ensures that there is a monic polynomial $h(t)$ and a polynomial $r(t)$ of degree strictly less than m such that $p(t) = q(t)h(t) + r(t)$. But $0 = p(A) = q(A)h(A) + r(A) =$

$0h(A) + r(A)$, so $r(A) = 0$. If $r(t)$ is not the zero polynomial, we could normalize it and obtain a monic annihilating polynomial of degree less than m, which would be a contradiction. We conclude that $r(t)$ is the zero polynomial, so $q(t)$ divides $p(t)$ with quotient $h(t)$. If there are two monic polynomials of minimum degree that annihilate A, this argument shows that each divides the other; since the degrees are the same, one must be a scalar multiple of the other. But since both are monic, the scalar factor must be $+1$ and they are identical. □

3.3.2 Definition. Let $A \in M_n$ be given. The unique monic polynomial $q_A(t)$ of minimum degree that annihilates A is called the *minimal polynomial* of A.

3.3.3 Corollary. Similar matrices have the same minimal polynomial.

Proof: If $A, B, S \in M_n$ and if $A = SBS^{-1}$, then $q_B(A) = q_B(SBS^{-1}) = Sq_B(B)S^{-1} = 0$, so $q_B(t)$ is a monic polynomial that annihilates A and hence the degree of $q_A(t)$ is less than or equal to the degree of $q_B(t)$. But $B = S^{-1}AS$, so the same argument shows that the degree of $q_B(t)$ is less than or equal to the degree of $q_A(t)$. Thus, $q_A(t)$ and $q_B(t)$ are monic polynomials of minimum degree that annihilate A, so (3.3.1) ensures that they are identical. □

Exercise. Consider $A = J_2(0) \oplus J_2(0) \in M_4$ and $B = J_2(0) \oplus 0_2 \in M_4$. Explain why A and B have the same minimal polynomial but are not similar.

3.3.4 Corollary. For each $A \in M_n$, the minimal polynomial $q_A(t)$ divides the characteristic polynomial $p_A(t)$. Moreover, $q_A(\lambda) = 0$ if and only if λ is an eigenvalue of A, so every root of $p_A(t) = 0$ is a root of $q_A(t) = 0$.

Proof: Since $p_A(A) = 0$, the fact that there is a polynomial $h(t)$ such that $p_A(t) = h(t)q_A(t)$ follows from (3.2.1). This factorization makes it clear that every root of $q_A(t) = 0$ is a root of $p_A(t) = 0$, and hence every root of $q_A(t) = 0$ is an eigenvalue of A. If λ is an eigenvalue of A, and if x is an associated eigenvector, then $Ax = \lambda x$ and $0 = q_A(A)x = q_A(\lambda)x$, so $q_A(\lambda) = 0$ since $x \neq 0$. □

The preceding corollary shows that if the characteristic polynomial $p_A(t)$ has been completely factored as

$$p_A(t) = \prod_{i=1}^{d}(t - \lambda_i)^{s_i}, \quad 1 \le s_i \le n, \quad s_1 + s_2 + \cdots + s_d = n$$

with $\lambda_1, \lambda_2, \ldots, \lambda_d$ distinct, then the minimal polynomial $q_A(t)$ must have the form

$$q_A(t) = \prod_{i=1}^{d}(t - \lambda_i)^{r_i}, \quad 1 \le r_i \le s_i$$

In principle, this gives an algorithm for finding the minimal polynomial of a given matrix A:

1. First compute the eigenvalues of A, together with their algebraic multiplicities, perhaps by finding the characteristic polynomial and factoring it completely. By some means, determine the factorization (3.3.5a).
2. There are finitely many polynomials of the form (3.3.5b). Starting with the product in which all $r_i = 1$, determine by explicit calculation the product of minimal degree that annihilates A; it is the minimal polynomial.

Numerically, this is not a good algorithm if it involves factoring the characteristic polynomial of a large matrix, but it can be very effective for hand calculations involving small matrices of simple form. Another approach to computing the minimal polynomial that does not involve knowing either the characteristic polynomial or the eigenvalues is outlined in Problem 5.

There is an intimate connection between the Jordan canonical form of $A \in M_n$ and the minimal polynomial of A. Suppose that $A = SJS^{-1}$ is the Jordan canonical form of A, and suppose first that $J = J_n(\lambda)$ is a single Jordan block. The characteristic polynomial of A is $(t - \lambda)^n$, and since $(J - \lambda I)^k \ne 0$ if $k < n$, the minimal polynomial of J is also $(t - \lambda)^n$. However, if

$$J = \begin{bmatrix} J_{n_1}(\lambda) & 0 \\ 0 & J_{n_2}(\lambda) \end{bmatrix} \in M_n$$

with $n_1 \ge n_2$, then the characteristic polynomial of J is still $(t - \lambda)^n$, but now $(J - \lambda I)^{n_1} = 0$ and no lower power vanishes. The minimal polynomial of J is therefore $(t - \lambda)^{n_1}$. If there are more Jordan blocks with eigenvalue λ, the conclusion is the same: The minimal polynomial of J is $(t - \lambda)^r$, in which r is the size of the largest Jordan block corresponding to λ. If J is a general Jordan matrix, the minimal polynomial must contain a factor $(t-\lambda_i)^{r_i}$ for each distinct eigenvalue λ_i, and r_i must be the size of the largest Jordan block corresponding to λ_i; no smaller power annihilates all the Jordan blocks corresponding to λ_i, and no greater power is needed. Since similar matrices have the same minimal polynomial, we have proved the following theorem.

3.3.6 Theorem. Let $A \in M_n$ be a given matrix whose distinct eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_d$. The minimal polynomial of A is

$$q_A(t) = \prod_{i=1}^{d} (t - \lambda_i)^{r_i} \tag{3.3.7}$$

in which r_i is the size of the largest Jordan block of A corresponding to the eigenvalue λ_i.

In practice, this result is not very helpful in computing the minimal polynomial since it is usually harder to determine the Jordan canonical form of a matrix than it is to determine its minimal polynomial. Indeed, if only the eigenvalues of a matrix are known, its minimal polynomial can be determined by simple trial and error. There are important theoretical consequences, however. Since a matrix is diagonalizable if and only if all its Jordan blocks have size 1, a necessary and sufficient condition for diagonalizability is that all $r_i = 1$ in (3.3.7).

3.3.8 Corollary. Let $A \in M_n$ have distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ and let

$$q(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_d) \tag{3.3.9}$$

Then A is diagonalizable if and only if $q(A) = 0$.

This criterion is actually useful for determining if a given matrix is diagonalizable, provided that we know its distinct eigenvalues: form the polynomial (3.3.9) and see if it annihilates A. If it does, it must be the minimal polynomial of A, since no lower-order polynomial could have as zeros all the distinct eigenvalues of A. If it does not annihilate A, then A is not diagonalizable. It is sometimes useful to have this result formulated in several equivalent ways:

3.3.10 Corollary. Let $A \in M_n$ and let $q_A(t)$ be its minimal polynomial. The following are equivalent:

(a) $q_A(t)$ has distinct linear factors.
(b) Every eigenvalue of A has multiplicity 1 as a root of $q_A(t) = 0$.
(c) For every eigenvalue λ of A, $q_A'(t) \neq 0$.
(d) A is diagonalizable.

We have been considering the problem of finding, for a given $A \in M_n$, a monic polynomial of minimum degree that annihilates A. But what about the converse? Given a monic polynomial

$$p(t) = t^n + a_{n-1}t^{n-1} + a_{n-2}t^{n-2} + \cdots + a_1 t + a_0 \tag{3.3.11}$$

is there a matrix A for which $p(t)$ is the minimal polynomial? If so, the size of A must be at least *n*-by-*n*. Consider

$$A = \begin{bmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ 0 & & & 1 & -a_{n-1} \end{bmatrix} \in M_n \tag{3.3.12}$$

and observe that

$$\begin{array}{rcccc} Ie_1 & = & e_1 & = & A^0 e_1 \\ Ae_1 & = & e_2 & = & Ae_1 \\ Ae_2 & = & e_3 & = & A^2 e_1 \\ Ae_3 & = & e_4 & = & A^3 e_1 \\ \vdots & & \vdots & & \vdots \\ Ae_{n-1} & = & e_n & = & A^{n-1} e_1 \end{array}$$

In addition,

$$\begin{aligned} Ae_n &= -a_{n-1}e_n - a_{n-2}e_{n-1} - \cdots - a_1 e_2 - a_0 e_1 \\ &= -a_{n-1}A^{n-1}e_1 - a_{n-2}A^{n-2}e_1 - \cdots - a_1 A e_1 - a_0 e_1 = A^n e_1 \\ &= [A^n - p(A)]e_1 \end{aligned}$$

Thus,

$$\begin{aligned} p(A)e_1 &= (a_0 e_1 + a_1 A e_1 + a_2 A^2 e_1 + \cdots + a_{n-1}A^{n-1}e_1) + A^n e_1 \\ &= [p(A) - A^n]e_1 + [A^n - p(A)]e_1 = 0 \end{aligned}$$

Furthermore, $p(A)e_k = p(A)A^{k-1}e_1 = A^{k-1}p(A)e_1 = A^{k-1}0 = 0$ for each $k = 1, 2, \ldots, n$. Since $p(A)e_k = 0$ for every basis vector e_k, we conclude that $p(A) = 0$. Thus $p(t)$ is a monic polynomial of degree n that annihilates A. If there were a polynomial $q(t) = t^m + b_{m-1}t^{m-1} + \cdots + b_1 t + b_0$ of lower degree $m < n$ that annihilates A, then

$$\begin{aligned} 0 &= q(A)e_1 = A^m e_1 + b_{m-1}A^{m-1}e_1 + \cdots + b_1 A e_1 + b_0 e_1 \\ &= e_{m+1} + b_{m-1}e_m + \cdots + b_1 e_2 + b_0 e_1 = 0 \end{aligned}$$

which is impossible since $\{e_1, e_2, \ldots, e_{m+1}\}$ is linearly independent. We conclude that n^{th} degree polynomial $p(t)$ is a monic polynomial of *minimum* degree that annihilates A, so it is the minimal polynomial of A. The characteristic polynomial $p_A(t)$ is also a monic polynomial of degree n that annihilates A, so (3.3.1) ensures that $p(t)$ is also the characteristic polynomial of the matrix (3.3.12).

3.3.13 Definition. The matrix (3.3.12) is the *companion matrix* of the polynomial (3.3.11).

We have proved the following:

3.3.14 Theorem. Every monic polynomial is both the minimal polynomial and the characteristic polynomial of its companion matrix.

If the minimal polynomial of $A \in M_n$ has degree n, then the exponents in (3.3.7) satisfy $r_1 + \cdots + r_d = n$; that is, the *largest* Jordan block corresponding to each eigenvalue is the *only* Jordan block corresponding to each eigenvalue. Such a matrix is nonderogatory. In particular, every companion matrix is nonderogatory. A nonderogatory matrix $A \in M_n$ need not be a companion matrix, of course, but A and the companion matrix C of the characteristic polynomial of A have the same Jordan canonical form (one block $J_{r_i}(\lambda_i)$ corresponding to each distinct eigenvalue λ_i), so A is similar to C.

Exercise. Provide details for a proof of the following theorem.

3.3.15 Theorem. Let $A \in M_n$ have minimal polynomial $q_A(t)$ and characteristic polynomial $p_A(t)$. The following are equivalent:

(a) $q_A(t)$ has degree n.
(b) $p_A(t) = q_A(t)$.
(c) A is nonderogatory.
(d) A is similar to the companion matrix of $p_A(t)$.

Problems

1. Let $A, B \in M_3$ be nilpotent. Show that A and B are similar if and only if A and B have the same minimal polynomial. Is this true in M_4?

2. Suppose $A \in M_n$ has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$. Explain why the minimal polynomial of A (3.3.7) is determined by the following algorithm: For each $i = 1, 2, \ldots, d$ compute $(A - \lambda_i I)^k$ for $k = 1, 2, \ldots, n$. Let r_i be the smallest value of k for which $\operatorname{rank}(A - \lambda_i I)^k = \operatorname{rank}(A - \lambda_i I)^{k+1}$.

3. Use (3.3.10) to show that every idempotent matrix is diagonalizable. *Hint:* Show that $t^2 - t = t(t-1)$ annihilates A. What is the minimal polynomial of A? What can you say if A is *tripotent* ($A^3 = A$)? What if $A^k = A$?

4. If $A \in M_n$ and $A^k = 0$ for some $k > n$, use properties of the minimal polynomial to explain why $A^r = 0$ for some $r \leq n$.

5. Show that the following application of the Gram–Schmidt process permits the minimal polynomial of a given $A \in M_n$ to be computed without knowing either the characteristic polynomial of A or any of its eigenvalues.

(a) Let the mapping $T : M_n \to \mathbf{C}^{n^2}$ be defined as follows: For any $A \in M_n$ partitioned according to columns as $A = [a_1 a_2 \dots a_n]$, let $T(A)$ denote the unique vector in $\mathbf{C}^{n^2}$ whose first n entries are the entries of the first column a_1, whose entries from $n+1$ to $2n$ are the entries of the second column a_2, and so forth. Show that this mapping T is an isomorphism (linear, one-to-one, and onto) of the vector spaces M_n and $\mathbf{C}^{n^2}$.

(b) Consider the vectors

$$v_0 = T(I), v_1 = T(A), v_2 = T(A^2), \dots, v_k = T(A^k), \dots$$

in $\mathbf{C}^{n^2}$ for $k = 0, 1, 2, \dots, n$. Use the Cayley–Hamilton theorem to show that $\{v_0, v_1, \dots, v_n\}$ is a dependent set.

(c) Apply the Gram–Schmidt process to the set $\{v_0, v_1, \dots, v_n\}$ in the given order until it stops by producing a first zero vector. Why must a zero vector be produced?

(d) If the Gram–Schmidt process produces a first zero vector at the kth step, argue that $k-1$ is the degree of the minimal polynomial of A.

(e) If the kth step of the Gram–Schmidt process produces the vector $\alpha_0 v_0 + \alpha_1 v_1 + \cdots + \alpha_{k-1} v_{k-1} = 0$, show that

$$\begin{aligned} &T^{-1}(\alpha_0 v_0 + \alpha_1 v_1 + \cdots + \alpha_{k-1} v_{k-1}) \\ &\quad = \alpha_0 I + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_{k-1} A^{k-1} = 0 \end{aligned}$$

and conclude that $q_A(t) = (\alpha_{k-1} t^{k-1} + \cdots + \alpha_2 t^2 + \alpha_1 t + \alpha_0)/\alpha_{k-1}$ is the minimal polynomial of A. Why is $\alpha_{k-1} \neq 0$?

6. Carry out the computations required by the algorithm in Problem 5 to determine the minimal polynomials of $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$.

7. Consider $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ to show that the minimal polynomials of AB and BA need not be the same. However, if $C, D \in M_n$, why must the characteristic polynomials of CD and DC be the same?

8. Let $A_i \in M_{n_i}, i = 1, 2, \dots, k$ and let $q_{A_i}(t)$ be the minimal polynomial of each A_i. Show that the minimal polynomial of $A = A_1 \oplus \cdots \oplus A_k$ is the least common multiple of $q_{A_1}(t), \dots, q_{A_k}(t)$. This is the unique monic polynomial of minimum degree that is divisible by each $q_i(t)$. Use this result to give a different proof for (1.3.10).

9. If $A \in M_5$ has characteristic polynomial $p_A(t) = (t-4)^3(t+6)^2$ and minimal polynomial $q_A(t) = (t-4)^2(t+6)$, what is the Jordan canonical form of A?

10. Show by direct computation that the polynomial (3.3.11) is the characteristic polynomial of the companion matrix (3.3.12). *Hint:* Use cofactors to compute the determinant.

11. Let $A \in M_n$ be the companion matrix (3.3.12) of the polynomial $p(t)$ in (3.3.11). Let K_n be the n-by-n reversal matrix. Let $A_2 = K_n A K_n$, $A_3 = A^T$, and $A_4 = K_n A^T K_n$. (a) Write A_2, A_3, and A_4 as explicit arrays like the one in (3.3.12). (b) Explain why $p(t)$ is both the minimal and characteristic polynomial of A_2, A_3, and A_4, each of which is encountered in the literature as an alternative definition of *companion matrix*.

12. Show that there is no real 3-by-3 matrix whose minimal polynomial is $x^2 + 1$, but that there is a real 2-by-2 matrix as well as a complex 3-by-3 matrix with this property. *Hint:* Use (3.3.4).

13. Explain why any n complex numbers can be the eigenvalues of an n-by-n companion matrix. However, the singular values of a companion matrix are subject to some very strong restrictions. Write the companion matrix (3.3.12) as a block matrix $A = \begin{bmatrix} 0 & \alpha \\ I_{n-1} & \xi \end{bmatrix}$, in which $\alpha = -a_0$ and $\xi = [-a_1 \ \dots \ -a_{n-1}]^T \in \mathbf{C}^{n-1}$. Verify that $A^*A = \begin{bmatrix} I_{n-1} & \xi \\ \xi^* & s \end{bmatrix}$, in which $s = |a_0|^2 + |a_1|^2 + \cdots + |a_{n-1}|^2$. Let $\sigma_1 \geq \cdots \geq \sigma_n$ denote the ordered singular values of A. Use (1.2.20) to show that $\sigma_2 = \cdots = \sigma_{n-1} = 1$ and

$$\sigma_1^2, \sigma_n^2 = \frac{1}{2}\left(s + 1 \pm \sqrt{(s+1)^2 - 4|a_0|^2}\right)$$

14. Use the example in the exercise preceding (3.3.4) to show that there are nonsimilar $A, B \in M_n$ such that for every polynomial $p(t)$, $p(A) = 0$ if and only if $p(B) = 0$.

15. Let $A \in M_n$ be given, and let $P(A) = \{p(A) : p(t) \text{ is a polynomial}\}$. Show that $P(A)$ is a subspace of M_n and that it is even a subalgebra of M_n. Explain why the dimension of $P(A)$ is the degree of the minimal polynomial of A.

16. Let $A, B \in M_n$. Suppose that $p_A(t) = p_B(t) = q_A(t) = q_B(t)$. Explain why A and B are similar. Use this fact to show that the alternative forms for the companion matrix noted in Problem 11 are all similar to (3.3.12).

17. Explain why any matrix that commutes with a companion matrix C must be a polynomial in C.

18. Newton's identities (2.4.18-19) can be proved by applying standard matrix analytic identities to the companion matrix. Adopt the notation of Problems

3 and 9 in (2.4) and let $A \in M_n$ be the companion matrix of $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$. Provide details for the following: (a) Since $p(t) = p_A(t)$, we have $p(A) = 0$ and $0 = \operatorname{tr}(A^k p(A)) = \mu_{n+k} + a_{n-1}\mu_{n+k-1} + \cdots + a_1\mu_{k+1} + a_0\mu_k$ for $k = 0, 1, 2, \ldots$, which is (2.4.19). (b) Use (2.4.13) to show that

$$\operatorname{tr}(\operatorname{adj}(tI - A)) = nt^{n-1} + \operatorname{tr} A_{n-2}t^{n-2} + \cdots + \operatorname{tr} A_1 t + \operatorname{tr} A_0 \quad (3.3.13)$$

and use (2.4.17) to show that $\operatorname{tr} A_{n-k-1} = \mu_k + a_{n-1}\mu_{k-1} + \cdots + a_{n-k+1}\mu_1 + na_{n-k}$, which is the coefficient of t^{n-k-1} in the right-hand side of (3.3.13) for $k = 1, \ldots, n-1$. Use (0.8.10.2) to show that $\operatorname{tr}(\operatorname{adj}(tI - A)) = nt^{n-1} + (n-1)a_{n-1}t^{n-2} + \cdots + 2a_2 t + a_1$, so $(n-k)a_{n-k}$ is the coefficient of t^{n-k-1} in the left-hand side of (3.3.13) for $k = 1, \ldots, n-1$. Conclude that $(n-k)a_{n-k} = \mu_k + a_{n-1}\mu_{k-1} + \cdots + a_{n-k+1}\mu_1 + na_{n-k}$ for $k = 1, \ldots, n-1$, which is equivalent to (2.4.17).

19. Let $A, B \in M_n$ and let $C = AB - BA$ be their commutator. In Problem 12 of (2.4) we learned that if C commutes with *either* A or B, then $C^n = 0$. If C commutes with *both* A and B, show that $C^{n-1} = 0$. What does this say if $n = 2$? *Hint*: Suppose $C^{n-1} \neq 0$ and use (3.2.4.2).

20. Let $A, B \in M_n$ be companion matrices (3.3.12) and let $\lambda \in \mathbf{C}$. (a) Show that λ is an eigenvalue of A if and only if $x_\lambda = [1\ \lambda\ \lambda^2\ \ldots\ \lambda^{n-1}]^T$ is an eigenvector of A^T. (b) If λ is an eigenvalue of A, show that every eigenvector of A^T associated with λ is a scalar multiple of x_λ. Deduce that every eigenvalue of A has geometric multiplicity one. (c) Explain why A^T and B^T have a common eigenvector if and only if they have a common eigenvalue. (d) If A commutes with B, why must A and B have a common eigenvalue?

21. Let $n \geq 2$, let C_n be the companion matrix (3.3.12) of $p(t) = t^n + 1$, let $L_n \in M_n$ be the strictly lower triangular matrix whose entries below the main diagonal are all equal to $+1$, let $E_n = L_n - L_n^T$, and let $\theta_k = \frac{\pi}{n}(2k+1)$, $k = 0, 1, \ldots, n-1$. Provide details for the following proof that the spectral radius of E_n is $\cot \frac{\pi}{2n}$. (a) The eigenvalues of C_n are $\lambda_k = e^{i\theta_k}$, $k = 0, 1, \ldots, n-1$ with respective associated eigenvectors $x_k = [1\ \lambda_k\ \ldots\ \lambda_k^{n-1}]^T$. (b) $E_n = C_n + C_n^2 + \cdots + C_n^{n-1}$ has eigenvectors x_k, $k = 0, 1, \ldots, n-1$ with respective associated eigenvalues

$$\begin{aligned}
\lambda_k + \lambda_k^2 + \cdots + \lambda_k^{n-1} &= \frac{\lambda_k - \lambda_k^n}{1 - \lambda_k} = \frac{1 + \lambda_k}{1 - \lambda_k} \\
&= \frac{e^{-i\theta_k/2} + e^{i\theta_k/2}}{e^{-i\theta_k/2} - e^{i\theta_k/2}} = i \cot \frac{\theta_k}{2}
\end{aligned}$$

for $k = 0, 1, \ldots, n-1$. (c) $\rho(E_n) = \cot \frac{\pi}{2n}$.

22. Let $A \in M_n$. Explain why the degree of the minimal polynomial of A is at least $\operatorname{rank} A + 1$, and show by example that this lower bound on the degree is best possible: for each $r = 1, \ldots, n-1$ there is some $A \in M_n$ such that the degree of $q_A(t)$ is $\operatorname{rank} A + 1$. *Hint*: Problem 4 in (3.2).

23. Show that a companion matrix is diagonalizable if and only if it has distinct eigenvalues.

24. Let $A \in M_n$ be a companion matrix (3.3.12). Show that: (a) If $n = 2$, then A is normal if and only if $|a_0| = 1$ and $a_1 = -a_0\overline{a_1}$; it is unitary if and only if $|a_0| = 1$ and $a_1 = 0$. (b) If $n \geq 3$, then A is normal if and only if $|a_0| = 1$ and $a_1 = \cdots = a_{n-1} = 0$, that is, if and only if $p_A(t) = t^n - c$ and $|c| = 1$; (c) If $n \geq 3$ and A is normal, then A is unitary and there is a $\varphi \in [0, 2\pi/n)$ such that the eigenvalues of A are $e^{i\varphi}e^{2\pi i k/n}$, $k = 0, 1, \ldots, n-1$.

25. If $a_0 \neq 0$, show that the inverse of the companion matrix A in (3.3.12) is

$$A^{-1} = \begin{bmatrix} \frac{-a_1}{a_0} & 1 & 0 & \cdots & 0 \\ \frac{-a_2}{a_0} & 0 & 1 & & 0 \\ \vdots & \vdots & \ddots & \ddots & \\ \frac{-a_{n-1}}{a_0} & 0 & & \ddots & 1 \\ \frac{-1}{a_0} & 0 & \cdots & \cdots & 0 \end{bmatrix} \tag{3.3.14}$$

and that its characteristic polynomial is

$$t^n + \frac{a_1}{a_0}t^{n-1} + \cdots + \frac{a_{n-1}}{a_0}t + \frac{1}{a_0} = \frac{t^n}{a_0}p_A(\frac{1}{t}) \tag{3.3.15}$$

26. This problem is a generalization of Problem 16 in (2.4). Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of $A \in M_n$, let $q_A(t) = (t - \lambda_1)^{\mu_1} \cdots (t - \lambda_d)^{\mu_d}$ be the minimal polynomial of A. For $i = 1, \ldots, d$, let $q_i(t) = q_A(t)/(t - \lambda_i)$ and let ν_i denote the number of blocks $J_{\mu_i}(\lambda_i)$ in the Jordan canonical form of A. Show that: (a) For each $i = 1, \ldots, d$, $q_i(A) \neq 0$, each of its nonzero columns is an eigenvector of A associated with λ_i, and each of its nonzero rows is the complex conjugate of a left eigenvector of A associated with λ_i; (b) for each $i = 1, \ldots, d$, $q_i(A) = X_i Y_i^*$, in which $X_i, Y_i \in M_{n,\nu_i}$ each have rank ν_i, $AX_i = \lambda_i X_i$, and $Y_i^* A = \lambda_i Y_i^*$; (c) $\operatorname{rank} q_i(A) = \nu_i, i = 1, \ldots, d$; (d) If $\nu_i = 1$ for some $i = 1, \ldots, d$, then there exists a polynomial $p(t)$ such that $\operatorname{rank} p(A) = 1$; (e) If A is nonderogatory, then there is a polynomial $p(t)$ such that $\operatorname{rank} p(A) = 1$; (f) The converse of the assertion in (d) is correct as well—can you prove it? *Hint*: $(\lambda_i I - A)q_i(A) = q_A(A)$.

27. The n^{th} order linear homogeneous ordinary differential equation

$$y^{(n)} + a_{n-1}y^{(n-1)} + a_{n-2}y^{(n-2)} + \cdots + a_1y' + a_0y = 0$$

for a complex-valued function $y(t)$ of a real parameter t can be transformed into a first order homogeneous system of ordinary differential equations $x' = Ax, A \in M_n, x = [x_1 \ \ldots \ x_n]^T$ by introducing auxiliary variables $x_1 = y, x_2 = y', \ldots, x_n = y^{(n-1)}$. Perform this transformation and show that A^T is the companion matrix (3.3.12).

28. Suppose that $K \in M_n$ is an involution. Explain why K is diagonalizable, and why K is similar to $I_m \oplus (-I_{n-m})$ for some $m \in \{0, 1, \ldots, n\}$. *Hint*: $K^2 = I$, so there are three possibilities for the minimal polynomial of K.

29. Suppose that $A, K \in M_n$, K is an involution, and $A = KAK$. Show that: (a) there is some $m \in \{0, 1, \ldots, n\}$ and matrices $A_{11} \in M_m$, $A_{22} \in M_{n-m}$ such that A is similar to $A_{11} \oplus A_{22}$ and KA is similar to $A_{11} \oplus (-A_{22})$; (b) λ is an eigenvalue of A if and only if either $+\lambda$ or $-\lambda$ is an eigenvalue of KA; (c) if $A \in M_n$ is centrosymmetric (0.9.10) and $K = K_n$ is the reversal matrix (0.9.5.1), then λ is an eigenvalue of A if and only if either $+\lambda$ or $-\lambda$ is an eigenvalue of K_nA, which presents the rows of A in reverse order. *Hint*: Let $K = SDS^{-1}$ with $D = I_m \oplus (-I_{n-m})$ and let $\mathcal{A} = S^{-1}AS = [A_{ij}]_{i,j=1}^2$. Then KA is similar to $D\mathcal{A}$ and $\mathcal{A} = D\mathcal{A}D \Rightarrow A_{12} = 0$ and $A_{22} = 0$.

30. Suppose that $A, K \in M_n$, K is an involution, and $A = -KAK$. Show that: (a) there is some $m \in \{0, 1, \ldots, n\}$ and matrices $A_{12} \in M_{m,n-m}$, $A_{21} \in M_{n-m,m}$ such that A is similar to $\mathcal{B} = \begin{bmatrix} 0_m & A_{12} \\ A_{21} & 0_{n-m} \end{bmatrix}$ and KA is similar to $\begin{bmatrix} 0_m & A_{12} \\ -A_{21} & 0_{n-m} \end{bmatrix}$; (b) A is similar to iKA, so λ is an eigenvalue of A if and only if $i\lambda$ is an eigenvalue of KA; (c) if $A \in M_n$ is skew-centrosymmetric (0.9.10) and K_n is the reversal matrix (0.9.5.1), then A is similar to iK_nA (thus, λ is an eigenvalue of A if and only if $i\lambda$ is an eigenvalue of K_nA, which presents the rows of A in reverse order). *Hint*: Let $T = iI_m \oplus I_{n-m}$ and compute $T\mathcal{B}T^{-1}$.

3.4 The real Jordan and Weyr canonical forms

In this section we discuss a real version of the Jordan canonical form for real matrices, as well as an alternative to the Jordan canonical form for complex matrices that is especially useful in problems involving commutativity.

3.4.1 The real Jordan canonical form Suppose that $A \in M_n(\mathbf{R})$, so any nonreal eigenvalues must occur in complex conjugate pairs. We have $\operatorname{rank}(A-$

$\lambda I)^k = \operatorname{rank}\overline{(A-\lambda I)^k} = \operatorname{rank}(\overline{A-\lambda I})^k = \operatorname{rank}(A-\bar{\lambda}I)^k$ for any $\lambda \in \mathbf{C}$ and all $k = 1, 2, \ldots$, so the Weyr characteristics of A associated with any complex conjugate pair of eigenvalues are the same (that is, $w_k(A,\lambda) = w_k(A,\bar{\lambda})$ for all $k = 1, 2, \ldots$). Lemma 3.1.18 ensures that the Jordan structure of A corresponding to any eigenvalue λ is the same as the Jordan structure of A corresponding to the eigenvalue $\bar{\lambda}$ (that is, $s_k(A,\lambda) = s_k(A,\bar{\lambda})$ for all $k = 1, 2, \ldots$). Thus, all the Jordan blocks of A of all sizes with nonreal eigenvalues occur in conjugate pairs of equal size.

For example, if λ is a nonreal eigenvalue of $A \in M_n(\mathbf{R})$, and if k blocks $J_2(\lambda)$ are in the Jordan canonical form of A, then there are k blocks $J_2(\bar{\lambda})$ as well. The block matrix

$$\begin{bmatrix} J_2(\lambda) & 0 \\ 0 & J_2(\bar{\lambda}) \end{bmatrix} = \left[\begin{array}{cc|cc} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ \hline 0 & 0 & \bar{\lambda} & 1 \\ 0 & 0 & 0 & \bar{\lambda} \end{array}\right]$$

is permutation-similar (interchange rows and columns 2 and 3) to the block matrix

$$\left[\begin{array}{cc|cc} \lambda & 0 & 1 & 0 \\ 0 & \bar{\lambda} & 0 & 1 \\ \hline 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \bar{\lambda} \end{array}\right] = \begin{bmatrix} D(\lambda) & I_2 \\ 0 & D(\lambda) \end{bmatrix}$$

in which $D(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix} \in M_2$.

In general, any Jordan matrix of the form

$$\begin{bmatrix} J_k(\lambda) & 0 \\ 0 & J_k(\bar{\lambda}) \end{bmatrix} \in M_{2k} \tag{3.4.1.1}$$

is permutation similar to the block upper triangular matrix

$$\begin{bmatrix} D(\lambda) & I_2 & & & \\ & D(\lambda) & I_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & I_2 \\ & & & & D(\lambda) \end{bmatrix} \in M_{2k} \tag{3.4.1.2}$$

which has k 2-by-2 blocks $D(\lambda)$ on the main block diagonal and $k-1$ blocks I_2 on the block superdiagonal.

Let $\lambda = a + ib, a, b \in \mathbf{R}$. A computation reveals that $D(\lambda)$ is similar to a real matrix

$$C(a,b) := \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = SD(\lambda)S^{-1} \qquad (3.4.1.3)$$

in which $S = \left[\begin{smallmatrix} -i & -i \\ 1 & -1 \end{smallmatrix}\right]$ and $S^{-1} = \frac{1}{2i}\left[\begin{smallmatrix} -1 & i \\ -1 & -i \end{smallmatrix}\right]$. Moreover, every block matrix of the form (3.4.1.2) with a nonreal λ is similar to a real block matrix of the form

$$C_k(a,b) := \begin{bmatrix} C(a,b) & I_2 & & & \\ & C(a,b) & I_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & I_2 \\ & & & & C(a,b) \end{bmatrix} \in M_{2k} \qquad (3.4.1.4)$$

via the similarity matrix $S \oplus \cdots \oplus S$ (k direct summands). Thus, every block matrix of the form (3.4.1.1) is similar to the matrix $C_k(a,b)$ in (3.4.1.4). These observations lead us to the *real Jordan canonical form theorem*.

3.4.1.5 Theorem. Each $A \in M_n(\mathbf{R})$ is similar via a real similarity to a real block diagonal matrix of the form

$$\begin{bmatrix} C_{n_1}(a_1,b_1) & & & & & \\ & \ddots & & & & \\ & & C_{n_p}(a_p,b_p) & & & \\ & & & J_{m_1}(\mu_1) & & \\ & & & & \ddots & \\ & & & & & J_{m_r}(\mu_r) \end{bmatrix} \qquad (3.4.1.6)$$

in which $\lambda_k = a_k + ib_k$, $k = 1, 2, \ldots, p$, are nonreal eigenvalues of A, each a_k and b_k is real and $b_k > 0$, and $\mu_1, \ldots, \mu_r$ are real eigenvalues of A. Each real block triangular matrix $C_{n_k}(a_k, b_k) \in M_{2n_k}$ is of the form (3.4.1.4) and corresponds to a pair of conjugate Jordan blocks $J_{n_k}(\lambda_k), J_{n_k}(\bar{\lambda}_k) \in M_{n_k}$ with nonreal λ_k in the Jordan canonical form (3.1.12) of A. The real Jordan blocks $J_{m_k}(\mu_k)$ in (3.4.6) are the Jordan blocks in (3.1.12) that have real eigenvalues.

Proof: We have shown that A is similar to (3.4.1.6) over $\mathbf{C}$. Theorem 1.3.28 ensures that A is similar to (3.4.6) over $\mathbf{R}$. □

The block matrix (3.4.1.6) is the *real Jordan canonical form* of A.

3.4.2 The Weyr canonical form The Weyr characteristic (3.1.16) played a key role in our discussion of uniqueness of the Jordan canonical form. It can also be used to define a canonical form for similarity that has certain advantages over the Jordan form.

Suppose that q is the index of an eigenvalue λ of $A \in M_n$, and let $w_k = w_k(A, \lambda), k = 1, 2, \ldots$ be the Weyr characteristic of A associated with λ. The *Weyr block* $W_A(\lambda)$ of A associated with λ is the upper triangular q-by-q block bidiagonal matrix

$$W_A(\lambda) = \begin{bmatrix} \lambda I_{w_1} & G_{1,2} & & & \\ & \lambda I_{w_2} & G_{2,3} & & \\ & & \ddots & \ddots & \\ & & & \ddots & G_{q-1,q} \\ & & & & \lambda I_{w_q} \end{bmatrix} \tag{3.4.2.1}$$

in which

$$G_{i,i+k} := \begin{bmatrix} I_{w_{i+k}} \\ 0 \end{bmatrix} \in M_{w_i, w_{i+k}}, \quad k = 1, 2, \ldots$$

Notice that $\operatorname{rank} G_{i,i+k} = w_{i+k}$.

For example, the Weyr characteristic of the Jordan matrix J in (3.1.16a) associated with the eigenvalue 0 is $w_1 = 6, w_2 = 5, w_3 = 2$, so

$$W_J(0) = \begin{bmatrix} 0_6 & G_{6,5} & \\ & 0_5 & G_{5,2} \\ & & 0_2 \end{bmatrix} \tag{3.4.2.2}$$

Exercise. Let λ be an eigenvalue of $A \in M_n$. Explain why the size of the Weyr block $W_A(\lambda)$ is the algebraic multiplicity of λ, which is the sum of the sizes of all the Jordan blocks of A with eigenvalue λ.

The Weyr block $W_A(\lambda)$ in (3.4.2.1) may be thought of as a q-by-q block matrix analog of a Jordan block. The number of diagonal blocks (the parameter q) is the *index* of λ (rather than its algebraic multiplicity), the diagonal blocks are *scalar matrices* λI (rather than scalars) with nonincreasingly ordered sizes, and the superdiagonal blocks are *full-column-rank blocks* $\begin{bmatrix} I \\ 0 \end{bmatrix}$ (rather than 1s) whose sizes are dictated by the sizes of the diagonal blocks.

Exercise. For the Weyr block (3.4.2.2), show by explicit calculation that

$$W_J(0)^2 = \begin{bmatrix} 0_6 & 0_{6,5} & G_{6,2} \\ & 0_5 & 0_{5,2} \\ & & 0_2 \end{bmatrix}$$

and $W_J(0)^3 = 0$. Explain why $\operatorname{rank} W_J(0) = 7 = w_2 + w_3$ and $\operatorname{rank} W_J(0)^2 =$

$2 = w_3$, and why the Weyr characteristic of $W_J(0)$ is $6, 5, 2$. Deduce that J is similar to $W_J(0)$. In fact, J is similar to $W_J(0)$ via a permutation similarity. Describe it.

Exercise. The size of the Weyr block $W_A(\lambda)$ in (3.4.2.1) is $w_1 + w_2 + \cdots + w_q$. Explain why $\operatorname{rank}(W_A(\lambda) - \lambda I) = w_2 + \cdots + w_q$.

Exercise. Verify that $G_{k-1,k}G_{k,k+1} = G_{k-1,k+1}$, that is,

$$\begin{bmatrix} I_{w_k} \\ 0_{w_{k-1}-w_k,w_k} \end{bmatrix} \begin{bmatrix} I_{w_{k+1}} \\ 0_{w_k-w_{k+1},w_{k+1}} \end{bmatrix} = \begin{bmatrix} I_{w_{k+1}} \\ 0_{w_{k-1}-w_{k+1},w_{k+1}} \end{bmatrix}$$

Using the preceding exercise, we find that $(W_A(\lambda) - \lambda I)^2 =$

$$\begin{bmatrix} 0_{w_1} & 0 & G_{1,3} & & \\ & 0_{w_2} & 0 & \ddots & \\ & & 0_{w_3} & \ddots & G_{q-2,q} \\ & & & \ddots & 0 \\ & & & & 0_{w_q} \end{bmatrix}$$

so $\operatorname{rank}(W_A(\lambda) - \lambda I)^2 = w_3 + \cdots + w_q$. Moving from one power to the next, each nonzero superdiagonal block $G_{k-p,k} \in M_{w_{k-p},w_k}$, $k = p+1, \ldots, q$ in $(W_A(\lambda) - \lambda I)^p$ moves up one block row into a higher superdiagonal of $(W_A(\lambda) - \lambda I)^{p+1}$ whose blocks are $G_{k-p-1,k} \in M_{w_{k-p-1},w_k}, k = p+2, \ldots, q$. In particular, $\operatorname{rank}(W_A(\lambda) - \lambda I)^p = w_{p+1} + \cdots + w_q, k = 1, 2, \ldots$. Observe that $\operatorname{rank}(W_A(\lambda) - \lambda I)^{p-1} - \operatorname{rank}(W_A(\lambda) - \lambda I)^p = w_p$, so the Weyr characteristic of $W_A(\lambda)$ associated with λ is $w_1, w_2, \ldots, w_q$; this is also the Weyr characteristic of A associated with λ.

A *Weyr matrix* is a direct sum of Weyr blocks with *distinct* eigenvalues.

We can now state the *Weyr canonical form theorem*:

3.4.2.3 Theorem. Let $A \in M_n$ be given, let $\lambda_1, \ldots, \lambda_d$ be its distinct eigenvalues in any prescribed order, let $w_k(A, \lambda_j), k = 1, 2, \ldots$, be the Weyr characteristic of A associated with the eigenvalue $\lambda_j, j = 1, \ldots, d$, and let $W_A(\lambda_j)$ be the Weyr block (3.4.2.1) for $j = 1, 2, \ldots, d$. Then there is a nonsingular $S \in M_n$ such that

$$A = S \begin{bmatrix} W_A(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & W_A(\lambda_d) \end{bmatrix} S^{-1}$$

The Weyr matrix $W_A = W_A(\lambda_1) \oplus \cdots \oplus W_A(\lambda_d)$ is uniquely determined by

A up to permutation of its direct summands. If A is real and has only real eigenvalues, then S can be chosen to be real.

Proof: The preceding observations show that W_A and A have identical Weyr characteristics associated with each of their distinct eigenvalues. Lemma 3.1.18 ensures that W_A and A are similar (and that W_A is unique up to permutation of its direct summands) since they are both similar to the same Jordan canonical form. If A and all its eigenvalues are real, then W_A is real and (1.3.28) ensures that A is similar to W_A via a real similarity. $\square$

The Weyr matrix $W_A = W_A(\lambda_1) \oplus \cdots \oplus W_A(\lambda_d)$ in the preceding theorem is the *Weyr canonical form* of A. The Weyr and Jordan canonical forms W_A and J_A contain the same information about A, but presented differently. Given one form, one can use the information incorporated into it to write down the other. Moreover, W_A is permutation similar to J_A and vice versa.

Exercise. Let $A \in M_n$ be given. Verify that the Jordan and Weyr canonical forms of A have the same number of nonzero entries.

Exercise. Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of $A \in M_n$. (a) If A is nonderogatory, explain why: (i) $w_1(A, \lambda_i) = 1$ for each $i = 1, \ldots, d$; (ii) in each Weyr block $W_A(\lambda_i), i = 1, \ldots, d$, every diagonal and superdiagonal sub-block of (3.4.2.1) is 1-by-1; (iii) The Weyr canonical form of A is the same matrix as its Jordan canonical form. (b) If $w_1(A, \lambda_i) = 1$ for each $i = 1, \ldots, d$, why must A be nonderogatory? *Hint*: $w_1(A, \lambda_i)$ is the geometric multiplicity of λ_i.

Exercise. Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of $A \in M_n$. (a) If A is diagonalizable, explain why: (i) $w_2(A, \lambda_i) = 0$ for all $i = 1, \ldots, d$; (ii) each Weyr block $W_A(\lambda_i), i = 1, \ldots, d$ (3.4.2.1) consists of a single block, which is a scalar matrix; (iii) the Weyr canonical form of A is the same matrix as its Jordan canonical form. (b) If $w_2(A, \lambda_i) = 0$ for some i, why is $w_1(A, \lambda_i)$ equal to the algebraic multiplicity of λ_i (it is *always* equal to the geometric multiplicity)? (c) If $w_2(A, \lambda_i) = 0$ for all $i = 1, \ldots, d$, why must A be diagonalizable?

Exercise. Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of $A \in M_n$. Explain why: For each $i = 1, \ldots, d$ there are at most p Jordan blocks of A with eigenvalue λ_i if and only if $w_1(A, \lambda_i) \leq p$ for each $i = 1, \ldots, d$, which is equivalent to requiring that *every* diagonal block of *every* Weyr block $W_A(\lambda_i)$ (3.4.2.1) is at most p-by-p.

In (3.2.4) we investigated the set of matrices that commute with a single given nonderogatory matrix. The key to understanding the structure of this set is the following observation: a matrix A commutes with a single Jordan block if and only if A is an upper triangular Toeplitz matrix (3.2.4.3). Thus, a matrix commutes with a *nonderogatory* Jordan matrix J if and only if it is a direct sum (conformal to J) of upper triangular Toeplitz matrices; in particular, it is upper triangular. The Jordan and Weyr canonical forms of a nonderogatory matrix A are identical; they are not the same if A is derogatory, and there is an important difference in how the matrices that commute with them can be described.

Exercise. Let $J = J_2(\lambda) \oplus J_2(\lambda)$. Show that: (a) $W_J = \begin{bmatrix} \lambda I_2 & I_2 \\ 0_2 & \lambda I_2 \end{bmatrix}$; (b) a matrix commutes with J if and only if it has the form $\begin{bmatrix} B & C \\ D & E \end{bmatrix}$ in which each of $B, C, D, E \in M_2$ is upper triangular Toeplitz; (c) a matrix commutes with W_J if and only if it has the block upper triangular form $\begin{bmatrix} B & C \\ 0 & B \end{bmatrix}$, in which $B, C \in M_2$.

A matrix that commutes with a derogatory Jordan matrix need not be block upper triangular. However, the preceding exercise suggests that the situation might be different for a derogatory Weyr block. The following lemma identifies the feature of a Weyr block that forces any matrix that commutes with it to have a block upper triangular structure.

3.4.2.4 Lemma. Let $\lambda \in \mathbf{C}$ and positive integers $n_1 \geq n_2 \geq \cdots \geq n_k \geq 1$ be given. Consider the upper triangular and identically partitioned matrices

$$F = [F_{ij}]_{i,j=1}^k = \begin{bmatrix} \lambda I_{n_1} & F_{12} & & \bigstar \\ & \lambda I_{n_2} & \ddots & \\ & & \ddots & F_{k-1,k} \\ & & & \lambda I_{n_k} \end{bmatrix} \in M_n$$

and

$$F' = [F'_{ij}]_{i,j=1}^k = \begin{bmatrix} \lambda I_{n_1} & F'_{12} & & \bigstar \\ & \lambda I_{n_2} & \ddots & \\ & & \ddots & F'_{k-1,k} \\ & & & \lambda I_{n_k} \end{bmatrix} \in M_n$$

Assume that all of the superdiagonal blocks $F'_{i,i+1}$ have full column rank. If $A \in M_n$ and $AF = F'A$, then A is block upper triangular conformal to F and F'. If, in addition, A is normal, then A is block diagonal conformal to F and F'.

Proof: Partition $A = [A_{ij}]_{i,j=1}^{k}$ conformally to F and F'. Our strategy is to inspect corresponding blocks of the identity $AF = F'A$ in a particular order. In block position $k-1, 1$ we have $\lambda A_{k-1,1} = \lambda A_{k-1,1} + F'_{k-1,k} A_{k1}$, so $F'_{k-1,k} A_{k1} = 0$ and hence $A_{k1} = 0$ since $F'_{k-1,k}$ has full column rank. In block position $k-2, 1$ we have $\lambda A_{k-2,1} = \lambda A_{k-2,1} + F'_{k-2,k-1} A_{k-1,1}$ (since $A_{k1} = 0$), so $F'_{k-2,k-1} A_{k-1,1} = 0$ and $A_{k-1,1} = 0$. Proceeding upward in the first block column of A and using at each step the fact that the lower blocks in that block column have been shown to be zero blocks, we find that $A_{i1} = 0$ for each $i = k, k-1, \dots, 2$. Now inspect block position $k-1, 2$ and proceed upward in the same fashion to show that $A_{i2} = 0$ for each $i = k, k-1, \dots, 3$. Continuing this process left to right and bottom to top, we find that A is block upper triangular conformal to F and F'. If A is normal and block triangular, (2.5.2) ensures that it is block diagonal. □

3.4.2.5 Corollary. Let $A \in M_n$ be given, let $\lambda_1, \dots, \lambda_d$ be its distinct eigenvalues in any prescribed order, let $w_k(A, \lambda_j)$, $k = 1, 2, \dots$, be the Weyr characteristic of A associated with the eigenvalue λ_j, $j = 1, \dots, d$, and let $W_A(\lambda_j)$ be the Weyr block (3.4.2.1) for $j = 1, 2, \dots, d$. Let $S \in M_n$ be nonsingular and such that $A = S(W_A(\lambda_1) \oplus \cdots \oplus W_A(\lambda_d))S^{-1}$. Suppose that $B \in M_n$ and $AB = BA$. Then (1) $S^{-1}BS = B^{(1)} \oplus \cdots \oplus B^{(k)}$ is block diagonal conformal to $W_A(\lambda_1) \oplus \cdots \oplus W_A(\lambda_d)$, and (2) each matrix $B^{(\ell)}$ is block upper triangular conformal to the partition (3.4.2.1) of the Weyr block $W_A(\lambda_\ell)$.

Proof: The assertion (1) follows from the basic result (2.4.4.2); the assertion (2) follows from the preceding lemma. □

Any matrix that commutes with a Weyr matrix is block upper triangular, but we can say a little more. Consider once again the Jordan matrix J in (3.1.16a), whose Weyr canonical form $W_J = W_J(0)$ is (3.4.2.2). In order to expose certain identities among the blocks of a (necessarily block upper triangular) matrix that commutes with W_J, we impose a finer partition on W_J. Let $m_k = w_{k-1} - w_k$, $k = 1, 2, 3$, so each m_k is the number of Jordan blocks of size k in J: $m_3 = 2$, $m_2 = 3$, and $m_1 = 1$. We have $w_1 = m_3 + m_2 + m_1 = 6$, $w_2 = m_3 + m_2 = 5$, and $w_3 = m_3 = 2$. Now re-partition W_J (3.4.2.2) with diagonal block sizes $2, 3, 1; 2, 3; 2$—this is known as the *standard partition*: the coarsest partition of a Weyr block such that every diagonal block is a scalar matrix (square) and every off-diagonal block is either an identity matrix (square) or a zero matrix (not necessarily square). In the standard partition, W_J has the form

$$W_J = \begin{bmatrix} 0_2 & 0 & 0 & I_2 & 0 & 0 \\ & 0_3 & 0 & 0 & I_3 & 0 \\ & & 0_1 & 0 & 0 & 0 \\ & & & 0_2 & 0 & I_2 \\ & & & & 0_3 & 0 \\ & & & & & 0_2 \end{bmatrix} \qquad (3.4.2.6)$$

Although the diagonal blocks in a Weyr block (3.4.2.1) are arranged in nonincreasing order of size, after imposing the standard partition the new, smaller, diagonal blocks need not occur in nonincreasing order of size. A computation reveals that N commutes with W_J if and only if it has the following block structure, conformal to that of (3.4.2.6):

$$N = \begin{bmatrix} B & C & \bigstar & D & \bigstar & \bigstar \\ & F & \bigstar & E & \bigstar & \bigstar \\ & & G & 0 & \bigstar & \bigstar \\ & & & B & C & D \\ & & & & F & E \\ & & & & & B \end{bmatrix} \qquad (3.4.2.7)$$

in which there are no constraints on the entries of the $\bigstar$ blocks. It is easier to see how the equalities among the blocks of (3.4.2.7) are structured if we collapse its standard partition to the coarser partition of (3.4.2.2): $N = [N_{ij}]_{i,j=1}^{3}$ with $N_{11} \in M_{w_1} = M_6$, $N_{22} \in M_{w_2} = M_5$, and $N_{33} \in M_{w_3} = M_2$. Then

$$\begin{aligned} N_{33} &= [B], \quad N_{23} = \begin{bmatrix} D \\ E \end{bmatrix}, \quad N_{22} = \begin{bmatrix} B & C \\ 0 & F \end{bmatrix}, \\ N_{12} &= \begin{bmatrix} D & \bigstar \\ E & \bigstar \\ 0 & \bigstar \end{bmatrix}, \quad N_{11} = \begin{bmatrix} B & C & \bigstar \\ 0 & F & \bigstar \\ 0 & 0 & \bigstar \end{bmatrix} \end{aligned}$$

that is,

$$N_{22} = \begin{bmatrix} N_{33} & \bigstar \\ 0 & \bigstar \end{bmatrix}, \quad N_{11} = \begin{bmatrix} N_{22} & \bigstar \\ 0 & \bigstar \end{bmatrix}, \quad N_{12} = \begin{bmatrix} N_{23} & \bigstar \\ 0 & \bigstar \end{bmatrix}$$

The pattern

$$N_{i-1,j-1} = \begin{bmatrix} N_{ij} & \bigstar \\ 0 & \bigstar \end{bmatrix} \qquad (3.4.2.8)$$

permits us to determine all the equalities among the blocks in the standard partition (including the positions of the off-diagonal zero block(s)) starting with the blocks in the last block column and working backwards up their block diagonals.

One final structural simplification of (3.4.2.7) is available to us. Let $U_3, \Delta_3 \in M_{m_3}$, $U_2, \Delta_2 \in M_{m_2}$, and $U_1, \Delta_1 \in M_{m_1}$ be unitary and upper triangular matrices (2.3.1) such that $B = U_3 \Delta_3 U_3^*$, $F = U_2 \Delta_2 U_2^*$, and $G = U_1 \Delta_1 U_1^*$ (the last factorization is trivial in this case). Let

$$U = U_3 \oplus U_2 \oplus U_1 \oplus U_3 \oplus U_2 \oplus U_3$$

Then

$$N' := U^* N U = \begin{bmatrix} \Delta_3 & C' & \bigstar & D' & \bigstar & \bigstar \\ & \Delta_2 & \bigstar & E' & \bigstar & \bigstar \\ & & \Delta_1 & 0 & \bigstar & \bigstar \\ & & & \Delta_3 & C' & D' \\ & & & & \Delta_2 & E' \\ & & & & & \Delta_3 \end{bmatrix} \qquad (3.4.2.9)$$

is upper triangular, in which $C' = U_3^* C U_2$, $D' = U_3^* D U_3$, and $E' = U_2^* E U_3$. The equalities among the blocks of N' on and above the block diagonal are the same as those of N. Moreover, W_J is unchanged after a similarity via U: $U^* W_J U = W_J$.

We can draw a remarkable conclusion from the preceding example. Suppose that: $A \in M_{13}$ has the Jordan canonical form (3.1.16a); $\mathcal{F} = \{A, B_1, B_2, \ldots\}$ is a commuting family; and $S \in M_{13}$ is nonsingular and $S^{-1}AS = W_A$ is the Weyr canonical form (3.4.2.11). Then $S^{-1}\mathcal{F}S = \{W_A, S^{-1}B_1S, S^{-1}B_2S, \ldots\}$ is a commuting family. Since each matrix $S^{-1}B_iS$ commutes with W_A, it has the block upper triangular form (3.4.2.7) in the standard partition. Thus, for each $j = 1, \ldots, 6$ the diagonal blocks in position j, j of all the matrices $S^{-1}B_iS$ constitute a commuting family, which can be upper triangularized by a single unitary matrix U_j (2.3.3). For each $i = 1, 2, \ldots$ the diagonal blocks of $S^{-1}B_iS$ in positions $(1,1)$, $(4,4)$, and $(6,6)$ are constrained to be the same, so we may (and do) insist that $U_1 = U_4 = U_6$. For the same reason, we insist that $U_2 = U_5$. Let $U = U_1 \oplus \cdots \oplus U_6$. Then each $U^*(S^{-1}B_iS)U$ is upper triangular and has the form (3.4.2.9), and $U^*S^{-1}ASU = U^*W_AU = W_A$. The conclusion is that there is a simultaneous similarity of the commuting family $\{A, B_1, B_2, \ldots\}$ (via $T = SU$, that is, $\mathcal{F} \to T^{-1}\mathcal{F}T$) that reduces A to Weyr canonical form and reduces every B_i to the upper triangular form (3.4.2.9).

All the essential features of the general case are captured in the preceding example, and by following its development one can prove the following theorem.

3.4.2.10 Theorem. Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of a given $A \in$

M_n in any prescribed order, let their respective indices be $q_1, \ldots, q_d$, and let their respective algebraic multiplicities be $p_1, \ldots, p_d$. For each $i = 1, \ldots, d$, let $w_1(A, \lambda_i), \ldots, w_{q_i}(A, \lambda_i)$ be the Weyr characteristic of A associated with λ_i and let $W_A(\lambda_i)$ be the Weyr block (3.4.2.1) of A associated with λ_i. Let

$$W_A = W_A(\lambda_1) \oplus \cdots \oplus W_A(\lambda_d) \tag{3.4.2.11}$$

be the Weyr canonical form of A, and let $A = SW_AS^{-1}$. Then
(1) Suppose that $B \in M_n$ commutes with A. Then $S^{-1}BS = B^{(1)} \oplus \cdots \oplus B^{(d)}$ is block diagonal conformal to W_A. For each $\ell = 1, \ldots, d$, partition $B^{(\ell)} = [B_{ij}^{(\ell)}]_{i,j=1}^{q_\ell} \in M_{p_\ell}$, in which each $B_{jj}^{(\ell)} \in M_{w_j(A,\lambda_\ell)}$, $j = 1, \ldots, q_\ell$. In this partition, $B^{(\ell)}$ is block upper triangular conformal to $W_A(\lambda_\ell)$ and its blocks along the k^{th} block superdiagonal are related by the identities

$$B_{j-k-1,j-1}^{(\ell)} = \begin{bmatrix} B_{j-k,j}^{(\ell)} & \bigstar \\ 0 & \bigstar \end{bmatrix}, \qquad \begin{array}{l} k = 0, 1, \ldots, q_\ell - 1; \\ j = q_\ell, q_\ell - 1, \ldots, k+1 \end{array} \tag{3.4.2.12}$$

(2) Let $\mathcal{F} = \{A, A_1, A_2, \ldots\} \subset M_n$ be a commuting family. There is a nonsingular $T \in M_n$ such that $T^{-1}\mathcal{F}T = \{W_A, T^{-1}A_1T, T^{-1}A_2T, \ldots\}$ is an upper triangular family. Each matrix $T^{-1}A_iT$ is block diagonal conformal to (3.4.2.11). If the diagonal block of $T^{-1}A_iT$ corresponding to $W_A(\lambda_\ell)$ is partitioned with diagonal block sizes $w_1(A, \lambda_\ell), w_2(A, \lambda_\ell), \ldots, w_{q_\ell}(A, \lambda_\ell)$, then its blocks along its k^{th} block superdiagonal are related by identities of the form (3.4.2.12).

3.4.3 The unitary Weyr form Theorem 3.4.2.3 and the QR factorization imply a refinement of Schur's unitary triangularization theorem (2.3.1) that incorporates the block structure of the Weyr canonical form.

3.4.3.1 Theorem. Let $\lambda_1, \ldots, \lambda_d$ be the distinct eigenvalues of a given $A \in M_n$ in any prescribed order, let $q_1, \ldots, q_d$ be their respective indices, and let $q = q_1 + \cdots + q_d$. Then A is unitarily similar to an upper triangular matrix of the form

$$F = \begin{bmatrix} \mu_1 I_{n_1} & F_{12} & F_{13} & \cdots & F_{1p} \\ & \mu_2 I_{n_2} & F_{23} & \cdots & F_{2p} \\ & & \mu_3 I_{n_3} & \ddots & \vdots \\ & & & \ddots & F_{p-1,p} \\ & & & & \mu_p I_{n_p} \end{bmatrix} \tag{3.4.3.2}$$

in which (a) $\mu_1 = \cdots = \mu_{q_1} = \lambda_1; \mu_{q_1+1} = \cdots = \mu_{q_1+q_2} = \lambda_2; \ldots;$

$\mu_{p-q_d+1} = \cdots = \mu_p = \lambda_d$; (b) For each $j = 1, \ldots, d$ the q_j integers $n_i, \ldots, n_{i+q_j-1}$ for which $\mu_i = \cdots = \mu_{i+q_j-1} = \lambda_j$ are the Weyr characteristic of λ_j as an eigenvalue of A, that is, $n_i = w_1(A, \lambda_j) \geq \cdots \geq n_{i+q_j-1} = w_{q_j}(A, \lambda_j)$; (c) if $\mu_i = \mu_{i+1}$ then $n_i \geq n_{i+1}$, $F_{i,i+1} \in M_{n_i, n_{i+1}}$ is upper triangular, and its diagonal entries are real and positive.

If $A \in M_n(\mathbf{R})$ and if $\lambda_1, \ldots, \lambda_d \in \mathbf{R}$, then A is real orthogonally similar to a real matrix F of the form (3.4.3.2) that satisfies conditions (a), (b), and (c).

The matrix F in (3.4.3.2) is determined by A up to the following equivalence: If A is unitarily similar to a matrix F' of the form (3.4.3.2) that satisfies the conditions (a), (b), and (c), then there is a block diagonal unitary matrix $U = U_1 \oplus \cdots \oplus U_p$ conformal to F such that $F' = UFU^*$, that is, $F'_{ij} = U_i^* F_{ij} U_j$, $i \leq j$, $i, j = 1, \ldots, p$.

Proof: Let $S \in M_n$ be nonsingular and such that

$$A = SW_A S^{-1} = S(W_A(\lambda_1) \oplus \cdots \oplus W_A(\lambda_d))S^{-1}$$

Let $S = QR$ be a QR factorization (2.1.14), so Q is unitary, R is upper triangular with positive diagonal entries, and $A = Q(RW_A R^{-1})Q^*$ is unitarily similar to the upper triangular matrix $RW_A R^{-1}$. Partition $R = [R_{ij}]_{i,j=1}^d$ conformally to W_A and compute

$$RW_A R^{-1} = \begin{bmatrix} R_{11} W(A, \lambda_1) R_{11}^{-1} & & \bigstar \\ & \ddots & \\ & & R_{dd} W(A, \lambda_d) R_{dd}^{-1} \end{bmatrix}$$

It suffices to consider only the diagonal blocks, that is, matrices of the form $TW(A, \lambda)T^{-1}$. The matrix T is upper triangular with positive diagonal entries; we partition $T = [T_{ij}]_{i,j=1}^q$ and $T^{-1} = [T^{ij}]_{i,j=1}^q$ conformally to $W(A, \lambda)$, whose diagonal block sizes are $w_1 \geq \cdots \geq w_q \geq 1$. The diagonal blocks of $TW(A, \lambda)T^{-1}$ are $T_{ii} \lambda I_{w_i} T^{ii} = \lambda I_{w_i}$ since $T^{ii} = T_{ii}^{-1}$ (0.9.10); the superdiagonal blocks are $T_{ii} G_{i,i+1} T^{i+1,i+1} + \lambda(T_{ii} T^{i,i+1} + T_{i,i+1} T^{i+1,i+1}) = T_{ii} G_{i,i+1} T^{i+1,i+1}$ (the term in parentheses is the $(i, i+1)$ block entry of $TT^{-1} = I$). If we partition $T_{ii} = \begin{bmatrix} C & \bigstar \\ 0 & D \end{bmatrix}$ with $C \in M_{w_i}$ (C is upper triangular with positive diagonal entries), then

$$T_{ii} G_{i,i+1} T^{i+1,i+1} = \begin{bmatrix} C & \bigstar \\ 0 & D \end{bmatrix} \begin{bmatrix} I_{w_{i+1}} \\ 0 \end{bmatrix} T_{i+1,i+1}^{-1} = \begin{bmatrix} CT_{i+1,i+1}^{-1} \\ 0 \end{bmatrix}$$

is upper triangular and has positive diagonal entries, as asserted.

If A is real and has real eigenvalues, (2.3.1), (3.4.2.3), and (2.1.14) ensure that the reductions in the preceding argument (as well as the QR factorization) can be achieved with real matrices.

Finally, suppose that $V_1, V_2 \in M_n$ are unitary, $A = V_1 F V_1^* = V_2 F' V_2^*$, and both F and F' satisfy the conditions (a), (b), and (c). Then $(V_2^* V_1)F = F'(V_2^* V_1)$, so (3.4.2.4) ensures that $V_2^* V_1 = U_1 \oplus \cdots \oplus U_p$ is block diagonal conformal with F and F', that is, $V_1 = V_2(U_1 \oplus \cdots \oplus U_p)$ and $F' = UFU^*$. $\square$

The following corollary illustrates how (3.4.3.1) can be used.

3.4.3.3 Corollary. Let $A \in M_n$ be given and suppose that $A^2 = A$. Let

$$\sigma_1 \geq \cdots \geq \sigma_g > 1 \geq \sigma_{g+1} \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots$$

be the singular values of A, so $r = \operatorname{rank} A$ and g is the number of singular values of A that are greater than 1. Then A is unitarily similar to

$$\begin{bmatrix} 1 & (\sigma_1^2 - 1)^{1/2} \\ 0 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 1 & (\sigma_g^2 - 1)^{1/2} \\ 0 & 0 \end{bmatrix} \oplus I_{r-g} \oplus 0_{n-r-g}$$

Proof: The minimal polynomial of A is $q_A(t) = t(t-1)$, so A is diagonalizable; its distinct eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 0$; their respective indices are $q_1 = q_2 = 1$; and their respective Weyr characteristics are $w_1(A,1) = r = \operatorname{tr} A$ and $w_1(A,0) = n - r$. Theorem 3.4.3.1 ensures that A is unitarily similar to $F = \left[\begin{smallmatrix} I_r & F_{12} \\ 0 & 0_{n-r} \end{smallmatrix}\right]$ and that F_{12} is determined up to unitary equivalence. Let $h = \operatorname{rank} F_{12}$ and let $F_{12} = V\Sigma W^*$ be a singular value decomposition: $V \in M_r$ and $W \in M_{n-r}$ are unitary, and $\Sigma \in M_{r,n-r}$ is diagonal with diagonal entries $s_1 \geq \cdots \geq s_h > 0 = s_{h+1} = \cdots$. Then F is unitarily similar (via $V \oplus W$) to $\left[\begin{smallmatrix} I_r & \Sigma \\ 0 & 0_{n-r} \end{smallmatrix}\right]$, which is permutation similar to

$$C = \begin{bmatrix} 1 & s_1 \\ 0 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 1 & s_h \\ 0 & 0 \end{bmatrix} \oplus I_{r-h} \oplus 0_{n-r-h}$$

The singular values of C (and hence also of A) are $(s_1^2+1)^{1/2}, \ldots, (s_h^2+1)^{1/2}$ together with $r-h$ ones and $n-r-h$ zeroes. It follows that $h = g$ and $s_i = (\sigma_i^2 - 1)^{1/2}, i = 1, \ldots, g$. $\square$

Exercise. Provide details for the preceding proof. Explain why two involutory matrices of the same size are unitarily similar if and only if they are unitarily equivalent, that is, if and only if they have the same singular values.

Problems

1. Suppose that $A \in M_n(\mathbf{R})$ and $A^2 = -I_n$. Show that n must be even, and that there is a nonsingular $S \in M_n(\mathbf{R})$ such that

$$S^{-1}AS = \begin{bmatrix} 0 & -I_{n/2} \\ I_{n/2} & 0 \end{bmatrix}$$

Hint: What is the real Jordan canonical form of A?

In the following three problems, for a given $A \in M_n$, $\mathcal{C}(A) = \{B \in M_n : AB = BA\}$ denotes the *centralizer* of A: the set of matrices that commute with A.

2. Explain why $\mathcal{C}(A)$ is an algebra.

3. Let $J \in M_{13}$ be the matrix in (3.1.16a). (a) Use (3.4.2.7) to show that $\dim \mathcal{C}(J) = 65$. (b) Show that $w_1(J,0)^2 + w_2(J,0)^2 + w_3(J,0)^2 = 65$.

4. Let the distinct eigenvalues of $A \in M_n$ be $\lambda_1, \ldots, \lambda_d$ with respective indices $q_1, \ldots, q_d$. (a) Show that $\dim \mathcal{C}(A) = \sum_{j=1}^{d} \sum_{i=1}^{q_j} w_i(A, \lambda_j)^2$. *Hint*: Use (3.4.2.10) and the identities (3.4.2.12). (b) Show that $\dim \mathcal{C}(A) \geq n$, with equality if and only if A is nonderogatory. *Hint*: $w_i(A,\lambda_j)^2 \geq w_i(A,\lambda_j)$. (c) Let the Segre characteristic of each eigenvalue λ_j of A be $s_i(A,\lambda_j), i = 1, \ldots, w_1(A,\lambda_j)$. It is known that $\dim \mathcal{C}(A) = \sum_{j=1}^{d} \sum_{i=1}^{w_1(A,\lambda)} (2i-1) s_i(A,\lambda_j)$; see Problem 9 in Section 4.4 of [HJ]. Explain why

$$\sum_{j=1}^{d} \sum_{i=1}^{q} w_i(A,\lambda)^2 = \sum_{j=1}^{d} \sum_{i=1}^{w_1(A,\lambda_j)} (2i-1) s_i(A,\lambda_j)$$

Verify this identity for the matrix in (3.1.16a).

5. Let $A \in M_n$ be given and suppose that $A^2 = 0$, that is, A is *self-annihilating*. Let $r = \operatorname{rank} A$ and let $\sigma_1 \geq \cdots \geq \sigma_r$ be the positive singular values of A. Show that A is unitarily similar to

$$\begin{bmatrix} 0 & \sigma_1 \\ 0 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 0 & \sigma_r \\ 0 & 0 \end{bmatrix} \oplus 0_{n-2r}$$

Explain why two self-annihilating matrices of the same size are unitarily similar if and only if they have the same singular values, that is, if and only if they are unitarily equivalent. *Hint*: Use (3.4.3.1) as in (3.4.3.3); the unitary Weyr form of A is again block 2-by-2, but now F_{12} has full column rank.

6. Show that $A \in M_2(\mathbf{R})$ is similar to $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ if and only if $A = \begin{bmatrix} 1+\alpha & (1+\alpha^2)/\beta \\ -\beta & 1-\alpha \end{bmatrix}$ for some $\alpha, \beta \in \mathbf{R}$ with $\beta \neq 0$.

Further Readings. Eduard Weyr announced his eponymous characteristic and canonical form in E. Weyr, Répartition des matrices en espèces et formation de toutes les espèces, *C. R. Acad. Sci. Paris* 100 (1885) 966-969. For an exposition of the Weyr characteristic and canonical form see H. Shapiro, The Weyr characteristic, *Amer. Math. Monthly* 196 (1999) 919-929; for technical details (in particular, for a proof of (3.4.2.10)) see V. Sergeichuk, Canonical matrices for linear matrix problems, *Linear Algebra Appl.* 317 (2000) 53-102.

3.5 Triangular factorizations

If a linear system $Ax = b$ has a nonsingular triangular (0.9.3) coefficient matrix $A \in M_n$, computation of the unique solution x is remarkably easy. If, for example, $A = [a_{ij}]$ is upper triangular and nonsingular, then all $a_{ii} \neq 0$ and one can employ *back substitution*: $a_{nn}x_n = b_n$ determines x_n; $a_{n-1,n-1}x_{n-1} + a_{n-1,n}x_n = b_{n-1}$ then determines x_{n-1} since x_n is known and $a_{n-1,n-1} \neq 0$; proceeding in the same fashion upward through successive rows of A one determines $x_{n-2}, x_{n-3}, \ldots, x_2, x_1$.

Exercise. Describe *forward substitution* as a solution technique for $Ax = b$ if $A \in M_n$ is nonsingular and lower triangular.

If $A \in M_n$ is not triangular, one can still use forward and back substitution to solve $Ax = b$ provided that A is nonsingular and can be factored as $A = LU$, in which L is lower triangular and U is upper triangular: First use forward substitution to solve $Ly = b$, and then use back substitution to solve $Ux = y$.

3.5.1 Definition. Let $A \in M_n$. A presentation $A = LU$, in which $L \in M_n$ is lower triangular and $U \in M_n$ is upper triangular, is called an *LU factorization* of A.

3.5.2 Lemma. Suppose that $A \in M_n$ and that $A = LU$ is an LU factorization. For any block 2-by-2 partition

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}, \quad U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

with $A_{11}, L_{11}, U_{11} \in M_k, k \leq n$, we have

$$\begin{aligned} &A_{11} = L_{11}U_{11} \\ &A_{12} = L_{11}U_{12}, \quad A_{21} = L_{21}U_{11} \\ &A_{22} = L_{21}U_{12} + L_{22}U_{22} \end{aligned}$$

Consequently, each leading principal submatrix of A has an LU factorization in which the factors are the corresponding leading principal submatrices of L and U.

Exercise. Verify (3.5.2) by carrying out the partitioned multiplication.

3.5.3 Theorem. Let $A \in M_n$ be given. Then:
(a) A has an LU factorization in which L is nonsingular if and only if A has the *row inclusion property*: for each $i = 1, ..., n-1$, $A[\{i+1; 1, ..., i\}]$ is a linear combination of the rows of $A[\{1, ..., i\}]$.
(b) A has an LU factorization in which U is nonsingular if and only if A has the *column inclusion property*: for each $j = 1, ..., n-1$, $A[\{1, ..., j; j+1\}]$ is a linear combination of the columns of $A[\{1, ..., j\}]$.

Proof: If $A = LU$, then $A[\{1, ..., i+1\}] = L[\{1, ..., i+1\}]U[\{1, ..., i+1\}]$. Thus, to verify the necessity of the row inclusion property, it suffices to take $i = k = n-1$ in the partitioned presentation given in (3.5.2). Since L is nonsingular and triangular, L_{11} is also nonsingular and we have $A_{21} = L_{21}U_{11} = L_{21}L_{11}^{-1}L_{11}U_{11} = \left(L_{21}L_{11}^{-1}\right)A_{11}$, which verifies the row inclusion property.

Conversely, if A has the row inclusion property we may construct inductively an LU factorization with nonsingular L as follows (the cases $n = 1, 2$ are easily verified): Suppose that $A_{11} = L_{11}U_{11}$, L_{11} is nonsingular, and the row vector A_{21} is a linear combination of the rows of A_{11}. Then there is a vector y such that $A_{21} = y^T A_{11} = y^T L_{11}U_{11}$, and we may take $U_{12} = L_{11}^{-1}A_{12}$, $L_{21} = y^T L_{11}$, $L_{22} = 1$, and $U_{22} = A_{22} - L_{21}U_{12}$ to obtain an LU factorization of A in which L is nonsingular.

The assertions about the column inclusion property follow from considering an LU factorization of A^T. $\square$

Exercise. Consider the matrix $J_n \in M_n$, all of whose entries are 1. Find an LU factorization of J_n in which L is nonsingular. With this factorization in hand, $J_n = J_n^T = U^T L^T$ is an LU factorization of J_n in which the upper triangular factor is nonsingular.

Exercise. Show that the row inclusion property is equivalent to the following formally-stronger property: For each $i = 1, ..., n-1$, every row of $A[\{i+1, ..., n\}; \{1, ..., i\}]$ is a linear combination of the rows of $A[\{1, ..., i\}]$. What is the corresponding statement for column inclusion?

Exercise. Characterize the square matrices that have LU factorizations in which L may be taken to be nonsingular, or in which U may be taken to be nonsingular, but not necessarily both.

If $A \in M_n$, $\operatorname{rank} A = k$, and $\det A[\{1, ..., j\}] \neq 0$, $j = 1, ..., k$, then A has both the row inclusion and column inclusion properties. The following result follows from (3.5.3).

3.5.4 Corollary. Suppose that $A \in M_n$ and rank $A = k$. If $A[\{1, \ldots, j\}]$ is nonsingular for all $j = 1, \ldots, k$, then A has an LU factorization. Furthermore, either factor may be chosen to be nonsingular; both L and U are nonsingular if and only if $k = n$, that is, if and only if A and all of its leading principal submatrices are nonsingular.

3.5.5 Example. Not every matrix has an LU factorization. If $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ could be written as $A = LU = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$, then $l_{11}u_{11} = 0$ implies that one of L or U is singular; but $LU = A$ is nonsingular.

Exercise. Explain why a nonsingular matrix that has a singular leading principal submatrix cannot have an LU factorization.

Exercise. Verify that

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

has an LU factorization even though A has neither the row nor column inclusion property. However, A is a principal submatrix of a 4-by-4 matrix

$$\hat{A} = \begin{bmatrix} A & e_1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \hat{A}_{12} \\ \hat{A}_{21} & 0 \end{bmatrix}, \quad \hat{A}_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \hat{A}_{21} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

that *does not* have an LU factorization. Verify this by considering the block factorization in (3.5.2) with $k = 2$: $\hat{A}_{12} = L_{11}U_{12}$ implies that L_{11} is nonsingular, and hence $0 = L_{11}U_{11}$ implies that $U_{11} = 0$, which is inconsistent with $L_{21}U_{11} = \hat{A}_{21} \neq 0$.

Exercise. Consider $A = \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 2-a \end{bmatrix}$ and explain why an LU factorization need not be unique even if the diagonal entries of L are required to be 1.

It is now clear that an LU factorization of a given matrix need not be unique, and it may or may not exist. Much of the trouble arises from singularity, either of A or of its leading principal submatrices. Using the tools of (3.5.2) and (3.5.3), however, we can give a full description in the nonsingular case, and we can impose a normalization that makes the factorization unique.

3.5.6 Corollary. Suppose that $A \in M_n$ is nonsingular. Then A has an LU factorization $A = LU$ if and only if $A[\{1, \ldots, j\}]$ is nonsingular for all $j = 1, \ldots, n$. Moreover, A may be factored as $A = L'DU'$ in which $L' \in M_n$ is lower triangular and $U' \in M_n$ is upper triangular, every diagonal entry of L' and U' is equal to 1, and D is a nonsingular diagonal matrix determined by

$$\det D[\{1, \ldots, j\}] = \det A\,[\{1, \ldots, j\}], \quad j = 1, \ldots, n$$

The factors L', U', and D are uniquely determined by A.

Exercise. Use (3.5.2), (3.5.3), and prior exercises to provide details for a proof of the preceding corollary.

Exercise. If $A \in M_n$ has an LU factorization with $L = [\ell_{ij}]$ and $U = [u_{ij}]$, show that $\ell_{11} u_{11} = \det[A\{1\}]$ and $\ell_{ii} u_{ii} \det A[\{1, ..., i{-}1\}] = \det A[\{1, ..., i\}]$, $i = 2, ..., n$.

Returning to the solution of the linear system $Ax = b$, suppose that $A \in M_n$ cannot be factored as LU, but can be factored as PLU, in which $P \in M_n$ is a permutation matrix, and L and U are lower and upper triangular, respectively. This amounts to a reordering of the equations in the linear system prior to factorization. In this event, solution of $Ax = b$ is still quite simple via $Ly = P^T b$ and $Ux = y$. It is worth knowing that any nonsingular $A \in M_n$ may be so factored and that L may be taken to be nonsingular. The solutions of $Ax = b$ are the same as those of $Ux = L^{-1} P^T b$.

3.5.7 Lemma. Let $A \in M_k$ be nonsingular. Then there is a permutation matrix $P \in M_k$ such that $\det(P^T A)[\{1, \ldots, j\}] \neq 0$, $j = 1, \ldots, k$.

Proof: The proof is by induction on k. If $k = 1$ or 2, the result is clear by inspection. Suppose that it is valid up to and including $k - 1$. Consider a nonsingular $A \in M_k$ and delete its last column. The remaining $k - 1$ columns are linearly independent and hence they contain $k - 1$ linearly independent rows. Permute these rows to the first $k - 1$ positions and apply the induction hypothesis to the nonsingular upper ($k - 1$)-by-($k - 1$) submatrix. This determines a desired overall permutation P, and $P^T A$ is nonsingular. □

3.5.8 Theorem. For each $A \in M_n$ there is a permutation matrix $P \in M_n$, a nonsingular lower triangular $L \in M_n$, and an upper triangular $U \in M_n$ such that $A = PLU$.

Proof: If we show that there is a permutation matrix Q such that QA has the row inclusion property, then (3.5.3) ensures that $QA = LU$ with L nonsingular, so $A = PLU$ for $P = Q^T$.

If A is nonsingular, the desired permutation is guaranteed by (3.5.7). If $\operatorname{rank} A = k < n$, first permute the rows of A so that the first k are linearly independent. It follows that $A[\{i+1\};\{1,...,i\}]$ is a linear combination of the rows of $A[\{1,...,i\}]$, $i=k,...,n-1$. If $A[\{1,...,k\}]$ is nonsingular, apply (3.5.7) again to further permute the rows so that $A[\{1,...,k\}]$, and thus A, has the row inclusion property. If $\operatorname{rank} A[\{1,...,k\}] = \ell < k$, treat it in the same way that we have just treated A, and obtain row inclusion for the indices $i=\ell,...,n-1$. Continue in this manner until either the upper left block is 0, in which case we have row inclusion for all indices, or it is nonsingular, in which case one further permutation completes the argument. $\square$

Exercise. Show that each $A \in M_n$ may be written $A = LUQ$, in which L is lower triangular, U is upper triangular and nonsingular, and Q is a permutation matrix.

Problems

1. The theory developed in this section deals with a factorization $A = LU$, with L lower triangular and U upper triangular. Discuss a parallel theory of $A = UL$ factorization, noting that the factors may be different.

2. Describe how $Ax = b$ may be solved if A is presented as $A = QR$, in which Q is unitary and R is upper triangular (2.1.14).

3. Show that $A \in M_n$ may be written as $A = LP_0U$, in which $L \in M_n$ is nonsingular and lower triangular, $U \in M_n$ is nonsingular and upper triangular, and P_0 is a sub-permutation matrix [a permutation matrix with as many of the 1's replaced by 0's as the rank of A is less than n]. *Hint:* Use elementary row and column operations.

4. If the leading principal minors of $A \in M_n$ are all nonzero, describe how an LU factorization of A may be obtained by using type 3 elementary row operations to zero out entries below the diagonal.

5. (Lanczos tridiagonalization algorithm.) Let $A \in M_n$ and $x \in \mathbf{C}^n$ be given. Define $X = [x\ Ax\ A^2x\ \dots\ A^{n-1}x]$. The columns of X are said to form a *Krylov sequence*. Assume that X is nonsingular. (a) Show that $X^{-1}AX$ is a companion matrix (3.3.12) for the characteristic polynomial of A. (b) If $R \in M_n$ is any given nonsingular upper triangular matrix and $S \equiv XR$, show that $S^{-1}AS$ is in upper Hessenberg form. (c) Let $y \in \mathbf{C}^n$ and define $Y = [y\ A^*y\ (A^*)^2y\ \dots\ (A^*)^{n-1}y]$. Suppose that Y is nonsingular and that Y^*X can be written as LDU, in which L is lower triangular and U is upper triangular and nonsingular, and D is diagonal and nonsingular. Show that there exist nonsingular upper triangular matrices R and T such that $(XR)^{-1} = T^*Y^*$ and such that T^*Y^*AXR is tridiagonal and similar to A.

(d) If $A \in M_n$ is Hermitian, use these ideas to describe an algorithm that produces a tridiagonal Hermitian matrix that is similar to A.

6. Explain why the n, n entry of a given $A \in M_n$ has no influence on whether it has an LU factorization, or has one with L nonsingular, or has one with U nonsingular.

7. Show that $C_n = [1/\max\{i,j\}] \in M_n(\mathbf{R})$ has an LU decomposition of the form $C_n = L_n L_n^T$, in which the entries of the lower triangular matrix L_n are $\ell_{ij} = 1/\max\{i,j\}$ for $i \geq j$. Conclude that $\det L_n = (1/n!)^2$.

8. Show that the condition "$A[\{1,\ldots,j\}]$ is nonsingular for all $j = 1,\ldots,n$" in (3.5.6) may be replaced with the condition "$A[\{j,\ldots,n\}]$ is nonsingular for all $j = 1,\ldots,n$".

9. Let $A \in M_n(\mathbf{R})$ be the symmetric tridiagonal matrix (0.9.10) with all main diagonal entries equal to $+2$ and all entries in the first superdiagonal and subdiagonal equal to -1. Consider

$$L = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & \ddots & & \\ & & \ddots & 1 & \\ & & & -\frac{n-1}{n} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 2 & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \ddots & \ddots & \\ & & & \frac{n}{n-1} & -1 \\ & & & & \frac{n+1}{n} \end{bmatrix}$$

Show that $A = LU$ and $\det A = n + 1$. The eigenvalues of A are $\lambda_k = 4\sin^2 \frac{k\pi}{2(n+1)}$, $k = 1,\ldots,n$ (see Problem 17 in (1.4)). Notice that $\lambda_1(A) \to 0$ and $\lambda_n(A) \to 4$ as $n \to \infty$, and $\det A = \lambda_1 \cdots \lambda_n \to \infty$.

10. Suppose that $A \in M_n$ is symmetric and that all its leading principal submatrices are nonsingular. Show that there is a nonsingular lower triangular L such that $A = LL^T$, that is, A has an LU factorization in which $U = L^T$.

Further Reading. Problem 5 is adapted from [Ste], where additional information about the numerical applications of LU factorizations may be found.

Index