Chapter 8

Jordan Canonical Forms

8.1 Jordan canonical forms

By Theorem 6.9, if A has an eigenvalue λ of multiplicity m such that $\dim E_\lambda(A) < m$, then A can not be diagonalizable. However, Theorem 8.1 below says that *any* matrix, even a non-diagonalizable matrix, is similar to a matrix very "close" to a diagonal matrix, called a *Jordan canonical form.* The columns of Q consists of a maximal set of linearly independent eigenvectors and some more vectors, so called *generalized eigenvectors.* Its proof may be beyond a beginning level, and can be found in advanced linear algebra books.

Theorem 8.1 *Let A be a square matrix of order n, and let $\sum_\lambda \dim E_\lambda(A) = s$. Then it is similar to a matrix J of the following form, called the* **Jordan canonical form**, *or a* **Jordan canonical matrix***:*

$$J = Q^{-1}AQ = \begin{bmatrix} J_1 & & & \mathbf{0} \\ & J_2 & & \\ & & \ddots & \\ \mathbf{0} & & & J_s \end{bmatrix},$$

in which each J_i, called a **Jordan block**, *is a triangular matrix of the form*

$$J_i = \begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix},$$

corresponding to a basis vector of $E_\lambda(A)$ for a single eigenvalue λ of A.

We will just illustrate how to find the Jordan canonical form and the transition matrix by observing the statements in the theorem.

Remark: (1) Each eigenvalue λ appears in $k = \dim E_\lambda(A)$ Jordan blocks. Thus the total number of Jordan blocks in a Jordan canonical form is the maximal number of linearly independent eigenvectors: *i.e.*, $\sum \dim E_{\lambda_i}(A)$.

(2) Thus, A has a full set of n linearly independent eigenvectors, if and only if the number of Jordan blocks of A is equal to n, if and only if every Jordan block is a 1×1 matrix. That is, a diagonal matrix is a particular case of the Jordan canonical form.

(3) The deficient eigenvectors in the transition matrix Q are filled up by, the so called, generalized eigenvectors, which will be discussed in the next section. Assuming Theorem 8.1, one can first determine the Jordan canonical form of A as shown in the following, and then try to find the generalized eigenvectors as discussed in the next section.

Example 8.1 Let A be a 5×5 matrix that has a single eigenvalue λ of multiplicity 5. Classify all possible Jordan canonical forms of A up to permutations of the Jordan blocks.

Solution: (1) If A has only one linearly independent eigenvector belonging to λ: *i.e.*, $\dim E_\lambda = 1$, then the Jordan canonical form of A has only one block of the form:

$$J^{(1)} = Q^{-1}AQ = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}.$$

(2) If it has two linearly independent eigenvectors belonging to λ: *i.e.*, $\dim E_\lambda = 2$, then its Jordan canonical form has two blocks, either one of the forms:

$$J^{(2)} = \begin{bmatrix} \lambda & 1 & & & \\ 0 & \lambda & & & \\ & & \lambda & 1 & 0 \\ & & 0 & \lambda & 1 \\ & & 0 & 0 & \lambda \end{bmatrix} \quad \text{or} \quad J^{(3)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & 1 & 0 & 0 \\ & 0 & \lambda & 1 & 0 \\ & 0 & 0 & \lambda & 1 \\ & 0 & 0 & 0 & \lambda \end{bmatrix}.$$

These two matrices $J^{(2)}$ and $J^{(3)}$ cannot be similar, because $(J^{(2)} - \lambda I)^3 = \mathbf{0}$, but $(J^{(3)} - \lambda I)^3 \neq \mathbf{0}$. (One can justify this by a direct computation).

(3) If it has three linearly independent eigenvectors belonging to λ: *i.e.*, $\dim E_\lambda = 3$, then its Jordan canonical form has three blocks, either one of the forms:

$$J^{(4)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & 1 & & \\ & 0 & \lambda & & \\ & & & \lambda & 1 \\ & & & 0 & \lambda \end{bmatrix} \quad \text{or} \quad J^{(5)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & 1 & 0 \\ & & 0 & \lambda & 1 \\ & & 0 & 0 & \lambda \end{bmatrix},$$

Again, these two matrices $J^{(4)}$ and $J^{(5)}$ are not similar, because $(J^{(4)} - \lambda I)^2 = \mathbf{0}$, but $(J^{(5)} - \lambda I)^2 \neq \mathbf{0}$.

(4) If it has four linearly independent eigenvectors belonging to λ: *i.e.*, $\dim E_\lambda = 4$, then its Jordan canonical form has four blocks of the form:

$$J^{(6)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & & \\ & & & \lambda & 1 \\ & & & 0 & \lambda \end{bmatrix}.$$

(5) If it has five linearly independent eigenvectors belonging to λ: *i.e.*, $\dim E_\lambda = 5$, then its Jordan canonical form has 5 blocks, each of which is just 1×1 matrix: *i.e.*, a diagonal matrix,

$$J^{(7)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & & \\ & & & \lambda & \\ & & & & \lambda \end{bmatrix}.$$

Thus, seven Jordan canonical forms are possible. Note that all of these seven possible Jordan canonical matrices have the same trace, determinant, characteristic polynomial, and eigenvalues as those of the matrix A, but no two of them are similar to each other. $\square$

As shown in the case (2) (also in (3)) of Example 8.1, when $\dim E_\lambda = 2$, two nonsimilar Jordan canonical matrices $J^{(2)}$ and $J^{(3)}$ were possible. To determine the right Jordan canonical form of the given matrix A, one has to look at the ranks of $(A - \lambda I)^k$ for $k = 1, 2, \ldots, n$.

In general, if λ is an eigenvalue of multiplicity m_λ of an $n \times n$ matrix A and J is its Jordan canonical form, then $\operatorname{rank}(A - \lambda I)^k = \operatorname{rank}\ (J - \lambda I)^k$

for all positive integer k, since they have the same eigenvalues and the same number of eigenvectors. Thus, to decide the order of each Jordan block in J, it is useful to examine the sequence $\{\text{rank}(J-\lambda I)^k \ : \ k = 1, \ldots, m_\lambda\}$.

Example 8.2 (*The orders of Jordan blocks*) Let A be a matrix with two distinct eigenvalues λ of multiplicity $m_\lambda = 7$ and μ of multiplicity $m_\mu = 1$, and similar to the Jordan canonical matrix J with four Jordan blocks:

$$J = \begin{bmatrix} J_1 & & & \mathbf{0} \\ & J_2 & & \\ & & J_3 & \\ \mathbf{0} & & & J_4 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} & & & \\ & \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} & & \\ & & \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} & \\ & & & [\mu] \end{bmatrix}.$$

This shows that λ has 3 linear independent eigenvectors. Let $c_\lambda = n - m_\lambda$ $(= 8 - 7 = 1)$ be the total multiplicities of the other eigenvalues of A. Then

(1) $\text{rank}(J-\lambda I) - c_\lambda = 5 - 1 = 4 =$ the total number of 1's off the diagonal in the Jordan blocks belonging to λ. The blocks of order 3 has two 1's and one in each block of order 2, so that there are one block of order 3 and two blocks of order 2.

(2) $\text{rank}(J-\lambda I)^2 - c_\lambda = 2 - 1 = 1 =$ the number of blocks of order 3.

$$(J-\lambda I)^2 = \begin{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & & & \\ & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & & \\ & & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \\ & & & [(\mu-\lambda)^2] \end{bmatrix}.$$

(3) $\text{rank}(J-\lambda I)^3 - c_\lambda = 1 - 1 = 0$. Thus, the number of blocks of order 1 for λ is $7 - (2 \times 2) - (3 \times 1) = 0$. $\square$

One can easily show that the power $(J_i - \lambda I)^k$ of the Jordan blocks J_i belonging to λ becomes a zero matrix for some $k \le m_\lambda$, and so $\text{rank}(J-\lambda I)^k$ decreases as k increases, and stops decreasing at $c_\lambda = n - m_\lambda$ for some

$k \le m_\lambda$, while all the other blocks belonging to the other eigenvalues remain as upper triangular matrices with nonzero diagonal entries, whose ranks will be summed up to $c_\lambda = n - m_\lambda$. This justifies that the sequence

$$\{\operatorname{rank}(A - \lambda I)^k - c_\lambda \ : \ k = 1, \ldots, m_\lambda\}$$

completely determines the orders of the blocks of J belonging to λ:

(1) If k is the smallest positive integer such that $\operatorname{rank}(A - \lambda I)^k - c_\lambda = \ell_0 = 0$ and $\operatorname{rank}(A - \lambda I)^{k-1} - c_\lambda = \ell_1 \neq 0$, then k is the order of the largest blocks belonging to λ and ℓ_1 is the number of such blocks.

(2) $\operatorname{rank}(A - \lambda I)^{k-2} - c_\lambda - 2\ell_1 = \ell_2$ is equal to the number of blocks of order $k - 1$.

(3) $\operatorname{rank}(A - \lambda I)^{k-3} - c_\lambda - 3\ell_1 - 2\ell_2 = \ell_3$ is equal to the number of blocks of order $k - 2$, and so on.

(4) If $\ell_1, \ldots, \ell_{i-1}$ are determined, then $\ell_i = \operatorname{rank}(A - \lambda I)^{k-i} - c_\lambda - i\ell_1 - (i-1)\ell_2 \cdots - 2\ell_{i-1}$ is the number of blocks of order $k - (i - 1)$ with $\ell_0 = 0$, for $i = 0, \ldots, k - 1$.

In summary, if $\lambda_1, \ \ldots, \ \lambda_t$ are all distinct eigenvalues of A with their multiplicities $m_{\lambda_1}, \ \ldots, \ m_{\lambda_t}$, respectively, so that $m_{\lambda_1} + \cdots + m_{\lambda_t} = n$, then one can determine the Jordan canonical form J of an $n \times n$ matrix A by the above procedure.

However, for a matrix of large order, the evaluation of $(A - \lambda I)^k$ might not be simple at all, while, for matrices of lower order or relatively simple matrices, the computations may be simple.

Example 8.3 Find the Jordan canonical form of the matrix

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

Solution: Clearly, the eigenvalues of A are $\lambda = 2$ and $\mu = 3$ of multiplicities 2 and 1, respectively. Hence, there are two possibilities of the Jordan canonical form of A:

$$J^{(1)} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{or} \quad J^{(2)} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

But, for $c_2 = 3 - 2 = 1$, $\operatorname{rank}(A - 2I)^2 - c_2 = 0$ and $\operatorname{rank}(A - 2I) - c_2 = 1$. Thus, there is one Jordan block belonging to $\lambda = 2$ of order 2, and so the Jordan canonical form of A must be $J^{(2)}$. □

Example 8.4 Determine the Jordan canonical form J of the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 4 & -6 & 4 \end{bmatrix}.$$

Solution: The characteristic polynomial of the matrix A is

$$\det(A - \lambda I) = \lambda^4 - 4\lambda^3 + 6\lambda^2 - 4\lambda + 1 = (\lambda - 1)^4.$$

Thus, the eigenvalue of A is $\lambda = 1$ of multiplicity 4. Now the rank of the matrix

$$A - I \ = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 4 & -6 & 3 \end{bmatrix}$$

is 3, since the first three columns are linearly independent and they span the last column. Hence, the $\dim \mathcal{N}(A - I) = \dim E_1 = 1$, *i.e.*, A has only one linearly independent eigenvector so that the Jordan canonical form J has only one block:

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

□

Problem 8.1 Let A be a 5×5 matrix with two distinct eigenvalues λ of multiplicity 3 and μ of multiplicity 2. Determine all possible Jordan canonical forms of A up to permutations of the Jordan blocks.

Problem 8.2 Find the Jordan canonical form for each of the following matrices:

$$(1) \begin{bmatrix} i & 0 \\ 1 & i \end{bmatrix}, \quad (2) \begin{bmatrix} 4 & 1 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 4 \end{bmatrix}, \quad (3) \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

8.2 Generalized eigenvectors

In the previous section, assuming Theorem 8.1, we have shown how to determine the Jordan canonical form of a matrix A by analyzing the sequence $\{\text{rank}(A-\lambda I)^k \ : \ k=1,2,\ldots,m_\lambda\}$, for each eigenvalue λ of A.

In this section, we discuss how to find a transition matrix Q, and provide a theoretical basis for the validity of the method.

Example 8.5 (*Transition matrix for a Jordan block*) Let A be a 3×3 matrix similar to a Jordan block of the form

$$Q^{-1}AQ = J = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}.$$

Determine the transition matrix $Q = [\mathbf{x}^1 \ \mathbf{x}^2 \ \mathbf{x}^3]$.

Solution: Clearly, λ is the only eigenvalue of the two similar matrices A and J of multiplicity 3. Since

$$\text{rank}(J-\lambda I) = \text{rank}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = 2,$$

$\dim \mathcal{N}(J-\lambda I) = \dim E_\lambda = 1$. That is, J has only one linearly independent eigenvector. To see what the columns of Q are, we expand $AQ = QJ$ as

$$[A\mathbf{x}^1 \ A\mathbf{x}^2 \ A\mathbf{x}^3] = [\lambda\mathbf{x}^1 \ \ \mathbf{x}^1 + \lambda\mathbf{x}^2 \ \ \mathbf{x}^2 + \lambda\mathbf{x}^3].$$

Thus, we get

$$\begin{array}{lll} A\mathbf{x}^3 = \mathbf{x}^2 + \lambda\mathbf{x}^3, & \text{or} & (A-\lambda I)\mathbf{x}^3 = (A-\lambda I)^1\mathbf{x}^3 = \mathbf{x}^2 \\ A\mathbf{x}^2 = \mathbf{x}^1 + \lambda\mathbf{x}^2, & \text{or} & (A-\lambda I)\mathbf{x}^2 = (A-\lambda I)^2\mathbf{x}^3 = \mathbf{x}^1 \\ A\mathbf{x}^1 = \lambda\mathbf{x}^1, & \text{or} & (A-\lambda I)\mathbf{x}^1 = (A-\lambda I)^3\mathbf{x}^3 = \mathbf{0}. \end{array}$$

This shows that $\mathbf{x}^1$ is an eigenvector of A, and if we find a vector $\mathbf{x}^3$ such that $(A-\lambda I)^3\mathbf{x}^3 = \mathbf{0}$ and $(A-\lambda I)^2\mathbf{x}^3 = \mathbf{x}^1 \neq \mathbf{0}$, then the other $\mathbf{x}^i$'s are obtained from $\mathbf{x}^3$ via $(A-\lambda I)^i\mathbf{x}^3 = \mathbf{x}^{3-i}$ for $i = 1,\ 2,\ 3$ with $\mathbf{x}^0 = \mathbf{0}$. Such a vector $\mathbf{x}^3$ is called a *generalized eigenvector* of rank 3, and the set $\{\mathbf{x}^1,\ \mathbf{x}^2,\ \mathbf{x}^3\}$ is called a *chain of the generalized eigenvectors* belonging to λ, and the columns of Q consist of chains of generalized eigenvectors. □

Remark: Note that, in the Example 8.5, if we set $Q' = [\mathbf{x}^3\ \mathbf{x}^2\ \mathbf{x}^1] = QP$ where P is an orthogonal permutation matrix that reverses the order of the columns of Q, then one can get

$$Q'^{-1}AQ' \;=\; P^{-1}(Q^{-1}AQ)P \;=\; P^{-1}JP = \begin{bmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{bmatrix} = J^T,$$

Thus, J and J^T are similar and represent essentially the same Jordan canonical form of A. Thus we say that any (complex) square matrix A is similar to a *unique* Jordan canonical matrix up to a permutation matrix. In this sense, it is called *the* Jordan canonical form of a matrix A.

In general, by expanding $AQ = QJ$, one can easily see that the columns of Q corresponding to the first columns of the Jordan blocks of J form a maximal set of linearly independent eigenvectors of A, and remaining columns of Q are generalized eigenvectors.

Definition 8.1 A nonzero vector $\mathbf{x}$ is said to be a **generalized eigenvector** of A of rank k belonging to an eigenvalue λ if

$$(A-\lambda I)^k\mathbf{x} = \mathbf{0} \quad \text{and} \quad (A-\lambda I)^{k-1}\mathbf{x} \neq \mathbf{0}.$$

Note that if $k = 1$, this is the usual definition of an eigenvector. For a generalized eigenvector $\mathbf{x}$ of rank $k \geq 1$ belonging to an eigenvalue λ, define

$$\begin{array}{lcllcl} \mathbf{x}^k & = & \mathbf{x}, & & & \\ \mathbf{x}^{k-1} & = & (A-\lambda I)\mathbf{x} & & = & (A-\lambda I)\mathbf{x}^k, \\ \mathbf{x}^{k-2} & = & (A-\lambda I)^2\mathbf{x} & & = & (A-\lambda I)\mathbf{x}^{k-1}, \\ \vdots & & & & & \\ \mathbf{x}^1 & = & (A-\lambda I)^{k-1}\mathbf{x} & & = & (A-\lambda I)\mathbf{x}^2. \end{array}$$

Thus, for each ℓ, $1 < \ell \leq k$, $(A-\lambda I)^\ell\mathbf{x}^\ell = (A-\lambda I)^k\mathbf{x} = \mathbf{0}$ and $(A-\lambda I)^{\ell-1}\mathbf{x}^\ell = (A-\lambda I)^{k-1}\mathbf{x} \neq \mathbf{0}$. Hence, the vector $\mathbf{x}^\ell = (A-\lambda I)^{k-\ell}\mathbf{x}$ is a generalized eigenvector of A of rank ℓ.

Definition 8.2 The set of vectors $\{\mathbf{x}^1,\ \mathbf{x}^2,\ \ldots,\ \mathbf{x}^k\}$ is called a **chain of generalized eigenvectors** belonging to the eigenvalue λ.

Note that, $\mathbf{x}^1 = (A-\lambda I)^{k-1}\mathbf{x}$ is always an eigenvector belonging to λ, called the **initial eigenvector** of the chain. Sometimes the length k of a chain is also called the **rank** of this initial eigenvector of the chain. Note also that $(A-\lambda I)^{\ell}\mathbf{x}^i = \mathbf{0}$ for $\ell \geq i$.

The following series of three theorems shows that a transition matrix Q may be constructed from the chains of linearly independent generalized eigenvectors of A, and justifies the invertibility of Q.

Theorem 8.2 *A chain of generalized eigenvectors* $S = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ *belonging to an eigenvalue* λ *is linearly independent.*

Proof: Let us solve $c_1\mathbf{x}^1+c_2\mathbf{x}^2+\cdots+c_k\mathbf{x}^k = \mathbf{0}$ for scalars c_i, $i = 1, \ldots, k$. If we multiply (on the left) both sides of this equation by $(A-\lambda I)^{k-1}$, then for $i = 1, \ldots, k-1$,

$$(A-\lambda I)^{k-1}\mathbf{x}^i = (A-\lambda I)^{k-(i+1)}(A-\lambda I)^i\mathbf{x}^i = \mathbf{0}.$$

Thus, $c_k(A-\lambda I)^{k-1}\mathbf{x}^k = \mathbf{0}$, and, hence, $c_k = 0$.

Do the same to the equation $c_1\mathbf{x}^1+\cdots+c_{k-1}\mathbf{x}^{k-1} = \mathbf{0}$ with $(A-\lambda I)^{k-2}$ and get $c_{k-1} = 0$. Proceeding successively, we can show that $c_i = 0$ for all $i = 1, \ldots, k$. That is, the equation has only the trivial solution. Hence, the set S is linearly independent. □

Theorem 8.3 *The union of chains of generalized eigenvectors of a square matrix* A *belonging to distinct eigenvalues is linearly independent.*

Proof: Let $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ and $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^\ell\}$ be the chains of generalized eigenvectors of A belonging to the eigenvalues λ and μ, respectively, and let $\lambda \neq \mu$. We wish to show that the set of vectors $\{\mathbf{x}^1, \ldots, \mathbf{x}^k, \mathbf{y}^1, \ldots, \mathbf{y}^\ell\}$ is linearly independent. To solve the linear dependence of them,

$$c_1\mathbf{x}^1+\cdots+c_k\mathbf{x}^k+d_1\mathbf{y}^1+\cdots+d_\ell\mathbf{y}^\ell = \mathbf{0},$$

for c_i's and d_j's, we multiply both sides of the equation by $(A-\lambda I)^k$ and note that $(A-\lambda I)^k\mathbf{x}^i = \mathbf{0}$ for all $i = 1, \ldots, k$. Thus we have

$$(A-\lambda I)^k(d_1\mathbf{y}^1+d_2\mathbf{y}^2+\cdots+d_\ell\mathbf{y}^\ell) = \mathbf{0}.$$

Again, multiply this equation by $(A-\mu I)^{\ell-1}$ and note that

$$\begin{array}{rcl}(A-\mu I)^{\ell-1}(A-\lambda I)^k & = & (A-\lambda I)^k(A-\mu I)^{\ell-1}, \\ (A-\mu I)^{\ell-1}\mathbf{y}^\ell & = & \mathbf{y}^1, \\ (A-\mu I)^{\ell-1}\mathbf{y}^i & = & \mathbf{0}\end{array}$$

for $i = 1, \ldots, \ell-1$. Thus we obtain $\mathbf{0} = d_\ell(A-\lambda I)^k\mathbf{y}^1$. Because $(A-\mu I)\mathbf{y}^1 = \mathbf{0}$ (or $A\mathbf{y}^1 = \mu\mathbf{y}^1$), this reduces to $d_\ell(\mu-\lambda)^k\mathbf{y}^1 = \mathbf{0}$, which implies that $d_\ell = 0$ by the assumption $\lambda \neq \mu$ and $\mathbf{y}^1 \neq \mathbf{0}$. Proceeding successively, we can show that $d_i = 0,\ i = \ell,\ \ell-1,\ \ldots,\ 2,\ 1$, so we are left with

$$c_1\mathbf{x}^1 + \cdots + c_k\mathbf{x}^k = \mathbf{0}.$$

Since $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$ is already linearly independent by Theorem 8.2, $c_i = 0$ for all $i = 1, \ldots, k$. Thus the set of generalized eigenvectors $\{\mathbf{x}^1, \ldots, \mathbf{x}^k, \mathbf{y}^1, \ldots, \mathbf{y}^\ell\}$ is linearly independent. □

The next step to produce Q such that $AQ = QJ$ is to choose chains of linearly independent generalized eigenvectors for each eigenvalue so that the union of the chains is linearly independent.

Definition 8.3 Let λ be an eigenvalue of A. The **generalized eigenspace** of A belonging to λ, denoted by K_λ, is the set

$$K_\lambda = \{\mathbf{x} \in \mathbb{C}^n \ :\ (A-\lambda I)^p\mathbf{x} = \mathbf{0} \quad \text{for some positive integer } p\}.$$

It turns out that dim K_λ is the multiplicity of λ, and it contains the usual eigenspace $\mathcal{N}(A-\lambda I)$. The following theorem enables us to choose a basis for K_λ, but we omit the proof even though it can be proved by induction on the number of vectors in $S \cup T$.

Theorem 8.4 *Let* $S = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ *and* $T = \{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^\ell\}$ *be two chains of generalized eigenvectors of* A *belonging to the same eigenvalue* λ. *If the initial vectors* $\mathbf{x}^1$ *and* $\mathbf{y}^1$ *are linearly independent, then the union* $S \cup T$ *is linearly independent.*

Note that this theorem easily extends to a finite number of chains of generalized eigenvectors of A belonging to an eigenvalue λ, and the union of such chains will form a basis for K_λ so that the matrix Q may be constructed from these bases for each eigenvalue as usual.

However, finding chains of generalized eigenvectors are not simple in general. One may try to find them in two ways as follows:

Method 1: (1) For each eigenvalue λ of multiplicity m_λ, first find a maximal set of linearly independent eigenvectors $\mathbf{x}, \mathbf{y}, \ldots, \mathbf{z}$ which form a basis for the eigenspace E_λ, and then solve

$$(A-\lambda I)\mathbf{x}^2 = a\mathbf{x} + b\mathbf{y} + \cdots + c\mathbf{z} = \mathbf{x}^1$$

for suitable a, b, …, c that make the systems (including those systems in step (2) below) consistent.

(2) Inductively, solve $(A-\lambda I)\mathbf{x}^i = \mathbf{x}^{i-1}$ for $i = 2, \ldots$, until the equation becomes inconsistent. When $(A-\lambda I)\mathbf{x}^{k+1} = \mathbf{x}^k$ becomes inconsistent, k is the rank of the generalized eigenvector $\mathbf{x}^k$.

(3) If $k < m_\lambda$, then, by the same method in step (1) above, find another suitable vector $\mathbf{y}^1$ in E_λ which is linearly independent from $\mathbf{x}^1$, and then repeat the above process.

(4) Do this until the ranks sum up to m_λ.

(5) An inconvenience of this method is that one does not know what the initial eigenvectors of the chains and their ranks are at the outset.

Example 8.6 The characteristic polynomial of

$$A = \begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix},$$

is $\det(\lambda I - A) = \lambda^3 - 3\lambda^2 + 3\lambda - 1 = (\lambda - 1)^3$, and the eigenvalue of A is $\lambda = 1$ of multiplicity 3. Its eigenvectors are the solutions of

$$(A - I)\mathbf{x} = \begin{bmatrix} 4 & -3 & -2 \\ 8 & -6 & -4 \\ -4 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since the three equations are identical, one gets two linearly independent eigenvectors $\mathbf{u}^1 = (1, 0, 2)$ and $\mathbf{u}^2 = (0, 2, -3)$. A generalized eigenvector is the solution of $(A - I)\mathbf{x}^2 = c_1\mathbf{u}^1 + c_2\mathbf{u}^2 = \mathbf{x}^1$ for some constants c_i's. In fact, the system

$$\begin{bmatrix} 4 & -3 & -2 \\ 8 & -6 & -4 \\ -4 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} c_1 \\ 2c_2 \\ 2c_1 - 3c_2 \end{bmatrix}$$

has a solution if and only if $c_1 = c_2$, and a solution is $\mathbf{x}^2 = (0, 0, -1)$ for $\mathbf{x}^1 = (2, 4, -2)$. Since $(A - I)\mathbf{x}^3 = \mathbf{x}^2$ is inconsistent, the rank of the generalized eigenvector $\mathbf{x}^2$ is 2 so that $\{\mathbf{x}^1, \mathbf{x}^2\}$ is a chain. Since $\mathbf{u}^1 = (1, 0, 2)$ is linearly independent to $\mathbf{x}^1$, we may take $\mathbf{x}^3 = \mathbf{u}^1$ and so

$$Q = [\mathbf{x}^1\ \mathbf{x}^2\ \mathbf{x}^3] = \begin{bmatrix} 2 & 0 & 1 \\ 4 & 0 & 0 \\ -2 & -1 & 2 \end{bmatrix}$$

is a transition matrix so that $Q^{-1}AQ = J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Method 2: (1) Determine the Jordan canonical form together with the order of each block as explained in the previous section.

(2) For each eigenvalue λ and each block of order k belonging to λ, find the solution $(A - \lambda I)^k\mathbf{x} = \mathbf{0}$, but $(A - \lambda I)^{k-1}\mathbf{x} \neq \mathbf{0}$, and then construct a chain of this generalized vector.

(3) An inconvenience of this method is that one has to compute $(A-\lambda I)^k$ for each eigenvalue λ and k, which is not simple if the order of A is large.

Example 8.7 For the same matrix A as in Example 8.6, we know that A has only two Jordan blocks for only one eigenvalue 1. By direct computation we find that $A - I \neq \mathbf{0}$ and $(A - I)^2 = \mathbf{0}$. Thus, we can take any vector $\mathbf{x}^2$ such that $(A - I)\mathbf{x}^2 \neq \mathbf{0}$, which becomes a generalized eigenvector of rank 2. Take $\mathbf{x}^2 = (1, 1, 1)$, then $\mathbf{x}^1 = (A - I)\mathbf{x}^2 = (-1, -2, 1)$. Now find another eigenvector of A belonging to $\lambda = 1$ by solving $(A - I)\mathbf{x}^3 = \mathbf{0}$, and get $\mathbf{x}^3 = (1, 0, 2)$. Then for $Q = [\mathbf{x}^1 \quad \mathbf{x}^2 \quad \mathbf{x}^3] = \begin{bmatrix} -1 & 1 & 1 \\ -2 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix}$, we get

$$Q^{-1}AQ = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad \square$$

Example 8.8 For a matrix $A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}$, find a transition matrix Q so that $Q^{-1}AQ$ is the Jordan canonical matrix.

Solution: Write $Q = [\mathbf{x}^1\ \mathbf{x}^2\ \mathbf{x}^3]$, and we try to find the columns $\mathbf{x}^i$, $i = 1$, 2, 3.

Method 1: In Example 8.3, the matrix J of A is determined as

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

From $AQ = QJ$, by comparing the column vectors of both sides, one can get

$$A\mathbf{x}^1 = 2\mathbf{x}^1,\ A\mathbf{x}^2 = 2\mathbf{x}^2 + \mathbf{x}^1,\ A\mathbf{x}^3 = 3\mathbf{x}^3.$$

Thus $\mathbf{x}^1$ and $\mathbf{x}^3$ are the eigenvectors of A belonging to $\lambda = 2$ and $\lambda = 3$, respectively. By a direct computation, they are found to be $\mathbf{x}^1 = (1,\ 0,\ 0)$ and $\mathbf{x}^3 = (3,\ -1,\ 1)$. By solving the equation $(A - 2I)\mathbf{x}^2 = \mathbf{x}^1$, one gets $\mathbf{x}^2 = (a,\ 1,\ 0)$ with any constant a, so that

$$Q = \begin{bmatrix} 1 & a & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is not hard to see that $\mathbf{x}^1$, $\mathbf{x}^2$, $\mathbf{x}^3$ are linearly independent, so that $Q^{-1}AQ = J$.

Method 2: From Example 8.3, the Jordan block belonging to 2 is of order 2. Thus we need to find a generalized eigenvector $\mathbf{x}^2$ of rank 2 belonging to the eigenvalue 2, which is a solution of the following systems:

$$\begin{aligned} (A - 2I)\mathbf{x} &= \begin{bmatrix} 0 & 1 & 4 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{x} \neq \mathbf{0}, \\ (A - 2I)^2\mathbf{x} &= \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{x} = \mathbf{0}. \end{aligned}$$

From the second equation, $\mathbf{x}^2$ has to be of the form $(a,\ b,\ 0)$, and from the first equation we must have $b \neq 0$. Let us take $\mathbf{x}^2 = (0,\ 1,\ 0)$ for a generalized eigenvector of rank 2. Thus we have

$$\begin{aligned} (A - 2I)\mathbf{x}^2 &= \mathbf{x}^1 = (1,\ 0,\ 0), \\ (A - 2I)^2\mathbf{x}^2 &= (A - 2I)\mathbf{x}^1 = \mathbf{0}. \end{aligned}$$

Thus $\mathbf{x}^1 = (1,\ 0,\ 0)$ is the only one linearly independent eigenvector belonging to 2. An eigenvector belonging to $\lambda_3 = 3$ is easily found to be $\mathbf{x}^3 = (3,\ -1,\ 1)$. Clearly, the set of vectors $\{\mathbf{x}^1,\ \mathbf{x}^2,\ \mathbf{x}^3\}$ is linearly independent, and so

$$Q = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{so} \quad Q^{-1} = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then

$$Q^{-1}AQ = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix},$$

where $J_1 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ and $J_2 = [3]$. □

Example 8.9 Find Q so that $Q^{-1}AQ = J$ is the Jordan canonical form of the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 4 & -6 & 4 \end{bmatrix}.$$

Solution: *Method 1*: In Example 8.4, we have found that

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Write a transition matrix as $Q = [\mathbf{x}^1\ \mathbf{x}^2\ \mathbf{x}^3\ \mathbf{x}^4]$. From the expansion of $AQ = QJ$, one gets:

$$A\mathbf{x}^1 = \mathbf{x}^1,\ A\mathbf{x}^2 = \mathbf{x}^2 + \mathbf{x}^1,\ A\mathbf{x}^3 = \mathbf{x}^3 + \mathbf{x}^2,\ A\mathbf{x}^4 = \mathbf{x}^4 + \mathbf{x}^3.$$

There is only one linearly independent eigenvector belonging to $\lambda = 1$, which is $\mathbf{x}^1 = (1,\ 1,\ 1,\ 1)$. Now a solution of $(A-I)\mathbf{x}^2 = \mathbf{x}^1$ is $\mathbf{x}^2 = (a,\ a+1,\ a+2,\ a+3)$ for any a. Take $\mathbf{x}^2 = (0,\ 1,\ 2,\ 3)$. Inductively, the solutions of $(A-I)\mathbf{x}^3 = \mathbf{x}^2$ and $(A-I)\mathbf{x}^4 = \mathbf{x}^3$ are $\mathbf{x}^3 = (b,\ b,\ b+1,\ b+3)$ for any b and set $\mathbf{x}^3 = (0,\ 0,\ 1,\ 3)$, and successively one can take $\mathbf{x}^4 = (0,\ 0,\ 0,\ 1)$. Clearly, they are linearly independent and so

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

One may check $Q^{-1}AQ = J$ by a direct matrix multiplication.

Method 2: As we saw in Example 8.4, A has only one Jordan block. Therefore, one has to find a generalized eigenvector of rank 4, which is a solution $\mathbf{x}$ of the following equations:

$$\begin{aligned} (A-I)^4\mathbf{x} &= \mathbf{0} \\ (A-I)^3\mathbf{x} &= \begin{bmatrix} -1 & 3 & -3 & 1 \\ -1 & 3 & -3 & 1 \\ -1 & 3 & -3 & 1 \\ -1 & 3 & -3 & 1 \end{bmatrix}\mathbf{x} \neq \mathbf{0}. \end{aligned}$$

But, a direct computation shows that the matrix $(A-I)^4 = \mathbf{0}$. Hence, one can take any vector that satisfies the second equation as a generalized eigenvector of rank 4: Take $\mathbf{x}^4 = (-1,\ 0,\ 0,\ 0)$, and then

$$\begin{aligned}
\mathbf{x}^3 &= (A-I)\mathbf{x}^4 = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 4 & -6 & 3 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \\
\mathbf{x}^2 &= (A-I)\mathbf{x}^3 = (-1,\ 0,\ 1,\ 2), \\
\mathbf{x}^1 &= (A-I)\mathbf{x}^2 = (1,\ 1,\ 1,\ 1).
\end{aligned}$$

Therefore, these vectors $\{\mathbf{x}^1,\ \mathbf{x}^2,\ \mathbf{x}^3,\ \mathbf{x}^4\}$ form the chain of linearly independent generalized eigenvectors. Therefore,

$$Q = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ -1 & 3 & -3 & 1 \end{bmatrix},$$

$$\text{and} \quad Q^{-1}AQ = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = J.$$

□

Problem 8.3 Find a full set of generalized eigenvectors of the following matrices:

$$(1)\ \begin{bmatrix} -2 & 0 & -2 \\ -1 & 1 & -2 \\ 0 & 1 & -1 \end{bmatrix}, \qquad (2)\ \begin{bmatrix} -6 & 31 & -14 \\ -1 & 6 & -2 \\ 0 & 2 & 1 \end{bmatrix}.$$

8.3 Applications of Jordan canonical form

The Jordan canonical form of any square matrix A enables us to compute the power A^k and the exponential matrix e^A, and to solve many other problems related to the matrix A. Let J be the Jordan canonical form of an arbitrary $n \times n$ square matrix A such that

$$Q^{-1}AQ = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{bmatrix},$$

where Q is made of generalized eigenvectors of A and J_i's are Jordan blocks.

8.3.1 Computations of A^n and e^A I

Note that we have

$$A^k = QJ^kQ^{-1} = Q\begin{bmatrix} J_1^k & & \\ & \ddots & \\ & & J_s^k \end{bmatrix}Q^{-1},$$

for $k = 1,\ 2,\ \ldots$. Hence, for A^k, it is good enough to compute J^k for each Jordan block J. Now an $m \times m$ Jordan block J belonging to an eigenvalue λ of A may be written as

$$\begin{aligned} J = \begin{bmatrix} \lambda & 1 & & 0 \\ 0 & \ddots & \ddots & \\ & & \lambda & 1 \\ 0 & & 0 & \lambda \end{bmatrix} &= \lambda\begin{bmatrix} 1 & 0 & & 0 \\ 0 & \ddots & \ddots & \\ \vdots & & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & & 0 & 1 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \\ &= \lambda I + N. \end{aligned}$$

Since I is the identity matrix, clearly $IN = NI$ and

$$J^k = (\lambda I + N)^k = \sum_{j=0}^{k}\binom{k}{j}\lambda^{k-j}N^j.$$

But $N^k = \mathbf{0}$ for $k \geq m$. Thus, by assuming $\binom{k}{\ell} = 0$ if $k < \ell$,

$$\begin{aligned} J^k &= \sum_{j=0}^{m-1}\binom{k}{j}\lambda^{k-j}N^j \\ &= \lambda^k I + \binom{k}{1}\lambda^{k-1}N + \cdots + \binom{k}{m-1}\lambda^{k-(m-1)}N^{m-1} \\ &= \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots & \binom{k}{m-1}\lambda^{k-m+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \binom{k}{m-2}\lambda^{k-m+2} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ 0 & \cdots & \cdots & 0 & \lambda^k \end{bmatrix}. \end{aligned}$$

Example 8.10 Compute A^k, $k = 1,\ 2,\ \ldots$, for

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 1 & 0 & 3 \end{bmatrix}.$$

Solution: The characteristic polynomial of A is $\det(\lambda I - A) = \lambda^4 - 8\lambda^3 + 24\lambda^2 - 32\lambda + 16 = (\lambda - 2)^4$, and $\lambda = 2$ is an eigenvalue of multiplicity 4. By a direct computation, one can see that $\operatorname{rank}(A - 2I) = 2$, $\operatorname{rank}(A - 2I)^2 = 1$ and $\operatorname{rank}(A - 2I)^3 = 0$. Thus, the Jordan canonical form J of A must be of the form

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

Also one can easily find a transition matrix Q to be

$$Q = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}, \quad \text{and then} \quad Q^{-1} = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 2 & -1 & 0 & -2 \end{bmatrix},$$

such that $Q^{-1}AQ = J$. Therefore,

$$A^k = QJ^kQ^{-1} = Q \begin{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}^k & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} 2 \end{bmatrix}^k \end{bmatrix} Q^{-1},$$

where

$$\begin{aligned} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}^k &= \left(\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \right)^k \\ &= \begin{bmatrix} 2^k & 0 & 0 \\ 0 & 2^k & 0 \\ 0 & 0 & 2^k \end{bmatrix} + \binom{k}{1} 2^{k-1} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} + \binom{k}{2} 2^{k-2} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2^k & \binom{k}{1} 2^{k-1} & \binom{k}{2} 2^{k-2} \\ 0 & 2^k & \binom{k}{1} 2^{k-1} \\ 0 & 0 & 2^k \end{bmatrix}. \end{aligned}$$

Hence,

$$A^k = QJ^kQ^{-1} = Q \begin{bmatrix} \begin{bmatrix} 2^k & \binom{k}{1} 2^{k-1} & \binom{k}{2} 2^{k-2} \\ 0 & 2^k & \binom{k}{1} 2^{k-1} \\ 0 & 0 & 2^k \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & 2^k \end{bmatrix} Q^{-1}$$

$$= \begin{bmatrix} 2^k - k2^{k-1} & k2^{k-1} & \frac{k(k-1)}{2}2^{k-2} + k2^{k-1} & k2^{k-1} \\ 0 & 2^k & 2k2^{k-1} & 0 \\ 0 & 0 & 2^k & 0 \\ -k2^{k-1} & k2^{k-1} & \frac{k(k-1)}{2}2^{k-2} & k2^{k-1} + 2^k \end{bmatrix}. \qquad \square$$

Problem 8.4 Compute A^k, $k = 1,\ 2,\ \ldots,$ for

$$(1)\ A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad (2)\ B = \begin{bmatrix} 0 & -3 & 1 & 2 \\ -2 & 1 & -1 & 2 \\ -2 & 1 & -1 & 2 \\ -2 & -3 & 1 & 4 \end{bmatrix}.$$

For the computations of the exponential matrix e^A, note that

$$\begin{aligned} e^A &= e^{QJQ^{-1}} = Qe^JQ^{-1} \\ &= Q \begin{bmatrix} e^{J_1} & & & \mathbf{0} \\ & e^{J_2} & & \\ & & \ddots & \\ \mathbf{0} & & & e^{J_s} \end{bmatrix} Q^{-1}, \end{aligned}$$

where J_i's are the Jordan blocks. Thus, it is enough to compute e^J for a simple Jordan block J. Let

$$J = \lambda I + N,$$

where I and N are as in Section 8.3.1. Then, $N^k = \mathbf{0}$ for $k \geq n$, and

$$e^J = e^{\lambda I}e^N = e^\lambda \sum_{k=0}^{n-1} \frac{N^k}{k!} = e^\lambda \begin{bmatrix} 1 & 1 & \frac{1}{2!} & \cdots & \frac{1}{(n-1)!} \\ 0 & 1 & 1 & \ddots & \frac{1}{(n-2)!} \\ & & 1 & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & 1 \end{bmatrix}.$$

Example 8.11 Compute e^A for

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 1 & 0 & 3 \end{bmatrix}.$$

Solution: With the same notations as in Example 8.10,

$$e^A = e^{QJQ^{-1}} = Qe^J Q^{-1} = Q \begin{bmatrix} e^{J_1} & \mathbf{0} \\ \mathbf{0} & e^{J_2} \end{bmatrix} Q^{-1},$$

where

$$J_1 = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} = 2I + \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = 2I + N \quad \text{and} \quad J_2 = [2]\,.$$

Hence,

$$e^{J_1} = e^{2I} e^N = e^2 \sum_{k=0}^{2} \frac{N^k}{k!} = e^2 \begin{bmatrix} 1 & 1 & \dfrac{1}{2!} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

It gives that

$$e^A = Qe^J Q^{-1} = e^2 Q \begin{bmatrix} 1 & 1 & \frac{1}{2} & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} Q^{-1} = \begin{bmatrix} 0 & e^2 & \frac{3}{2}e^2 & e^2 \\ 0 & e^2 & 2e^2 & 0 \\ 0 & 0 & e^2 & 0 \\ -e^2 & e^2 & \frac{1}{2}e^2 & 2e^2 \end{bmatrix}. \qquad \square$$

8.3.2 Linear difference equations II

Let A be the companion matrix of order k of a linear difference equation $\mathbf{x}_n = A\mathbf{x}_{n-1}$, and let λ_0 be an eigenvalue of A with multiplicity $m > 1$. Then, by Lemma 7.2, λ_0 has only one linearly independent eigenvector $\mathbf{v}$ of the form $[\lambda_0^{k-1} \ \cdots \ \lambda_0 \ 1]^T$, which means that the size of the Jordan block corresponding to this eigenvector is the multiplicity m of λ_0. Therefore, if A has s distinct eigenvalues each of multiplicity m_j, $j = 1, 2, \ldots, s$ so that $m_1 + m_2 + \cdots + m_s = k$, then the Jordan Canonical form of A has s Jordan blocks each of order m_i: *i.e.*,

$$Q^{-1}AQ = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{bmatrix}, \quad \text{where} \ J_j = \begin{bmatrix} \lambda_j & 1 & \cdots & 0 \\ & \ddots & & 1 \\ & & & \lambda_j \end{bmatrix}.$$

The chain of generalized eigenvectors containing $\mathbf{v}_{j1} = [\lambda_j^{k-1} \ \cdots \ \lambda_j \ 1]^T$ consists of m_j vectors: $\mathbf{v}_{ji}$, $i = 1, 2, \ldots, m_j$, which, from $AQ = QJ$, satisfy

$$(A - \lambda_j I)\mathbf{v}_{j1} = \mathbf{0}, \ (A - \lambda_j I)\mathbf{v}_{j2} = \mathbf{v}_{j1}, \ \ldots, \ (A - \lambda_j I)\mathbf{v}_{jm_j} = \mathbf{v}_{j(m_j - 1)}.$$

Beginning with $\mathbf{v}_{j1} = [\lambda_j^{k-1} \ \cdots \ \lambda_j \ 1]^T$, one can easily compute inductively a chain of generalized eigenvectors: *i.e.*, if we denote a generalized eigenvector by $\mathbf{v} = [x_{k-1} \ \cdots \ x_1 \ x_0]^T$, then

$$\begin{array}{llll}
\text{for } \ \mathbf{v}_{j1}, & x_n = \lambda_j^n & = \binom{n}{0}\lambda_j^n, & \text{for } \ n \geq 0, \\
\text{for } \ \mathbf{v}_{j2}, & x_n = n\lambda_j^{n-1} & = \binom{n}{1}\lambda_j^{n-1}, & \text{for } \ n \geq 1, \\
\text{for } \ \mathbf{v}_{j3}, & x_n = \frac{n(n-1)}{2}\lambda_j^{n-2} & = \binom{n}{2}\lambda_j^{n-2}, & \text{for } \ n \geq 2, \\
 & \vdots & & \\
\text{for } \ \mathbf{v}_{jm_j}, & x_n = \quad \cdots & = \binom{n}{m_j-1}\lambda_j^{n-(m_j-1)}, & \text{for } \ n \geq m_j - 1.
\end{array}$$

Thus the columns of the transition matrix Q corresponding to J_j looks like:

$$\begin{bmatrix}
\lambda_j^{k-1} & \binom{k-1}{1}\lambda_j^{(k-1)-1} & \binom{k-1}{2}\lambda_j^{(k-1)-2} & \cdots & \binom{k-1}{m_j-1}\lambda_j^{(k-1)-(m_j-1)} \\
\vdots & \vdots & \vdots & & \vdots \\
\lambda_j^{m_j-1} & \binom{m_j-1}{1}\lambda_j^{m_j-2} & \binom{m_j-1}{2}\lambda_j^{m_j-3} & \cdots & 1 \\
\vdots & \vdots & \vdots & & 0 \\
\lambda_j^2 & \lambda_j & 1 & & \vdots \\
\lambda_j & 1 & 0 & & \\
1 & 0 & 0 & \cdots & 0
\end{bmatrix}.$$

Notice that the coefficients form the Pascal triangle. Therefore, the general solution of a linear difference equation $\mathbf{x}_n = A\mathbf{x}_{n-1}$ is of the form,

$$x_n \ = \ \sum_{j=1}^{s}\left(c_{j1}\binom{n}{0}\lambda_j^n + c_{j2}\binom{n}{1}\lambda_j^{n-1} + \ \cdots + c_{jm_j}\binom{n}{m_j-1}\lambda_j^{n-m_j+1}\right),$$

for $n \geq 0$. Here we assume that $\binom{n}{m} = 0$ if $n < m$.

8.3.3 Discrete dynamical systems II

In Section 7.2, the discrete dynamical systems have been introduced, and it is shown that 1 is always an eigenvalue of a stochastic matrix. We now show that the magnitudes of all the eigenvalues of a stochastic matrix are bounded by 1, and moreover, if a stochastic matrix has positive entries, then its eigenvalues λ are all either $|\lambda| < 1$ or $\lambda = 1$, and so no one is such that $|\lambda| = 1$ but $\lambda \neq 1$.

For this, we first try to estimate the bound for the magnitude of the eigenvalues of any square matrix A (including complex matrix) in terms of the absolute values of the entries of A. For any square matrix $A = [{a_i}^j]$ of order k, let

$$\begin{aligned} R(A) &= \max\{R_i(A) = \sum_{j=1}^{k} |{a_i}^j| \ : \ 1 \le i \le k\}, \\ C(A) &= \max\{C_j(A) = \sum_{i=1}^{k} |{a_i}^j| \ : \ 1 \le j \le k\}, \\ s_i &= R_i(A) - |{a_i}^i|. \end{aligned}$$

Theorem 8.5 (Gerschgorin's Theorem) *For any square matrix A of order k, every eigenvalue λ of A satisfies $|\lambda - a_{\ell\ell}| \le s_\ell$ for some $1 \le \ell \le k$.*

Proof: Let λ be an eigenvalue with eigenvector $\mathbf{x} = [x_1\ x_2\ \cdots\ x_k]^T$. Then $\sum_{j=1}^{k} {a_i}^j x_j = \lambda x_i$, for $i = 1, \ldots, k$. Take a coordinate x_ℓ of $\mathbf{x}$ with the largest absolute value. Then clearly $x_\ell \ne 0$, and

$$|\lambda - {a_\ell}^\ell||x_\ell| = |\lambda x_\ell - {a_\ell}^\ell x_\ell| = \left|\sum_{j\ne\ell} {a_\ell}^j x_j\right| \le \sum_{j\ne\ell} |{a_\ell}^j||x_\ell| \ = \ s_\ell |x_\ell|.$$

Since $|x_\ell| > 0$, $|\lambda - {a_\ell}^\ell| \le s_\ell$. □

Corollary 8.6 *For any square matrix A of order k, every eigenvalue λ of A satisfies $|\lambda| \le \min\{R(A), C(A)\}$.*

Proof: Note that $|\lambda| \le |\lambda - {a_\ell}^\ell| + |{a_\ell}^\ell| \le s_\ell + |{a_\ell}^\ell| = R_\ell(A) \le R(A)$. Moreover, since λ is also an eigenvalue of A^T, $|\lambda| \le R(A^T) = C(A)$. □

Furthermore, if A is a stochastic matrix, then $C(A) = 1$.

Corollary 8.7 *If λ is an eigenvalue of a stochastic matrix A, then $|\lambda| \le 1$.*

We next give a quick explanation of the key properties of a positive matrix.

Theorem 8.8 (Perron-Frobenius Theorem) *Let A be a matrix with positive entries. If λ_0 is an eigenvalue of A such that $|\lambda_0|$ is the largest, then λ_0 is real and positive, and so are the components of its eigenvector $\mathbf{x}$.*

Proof: Let λ and $\mathbf{z}$ be such that $A\mathbf{z} = \lambda\mathbf{z}$. Then $|\lambda||\mathbf{z}| = |A\mathbf{z}| \le A|\mathbf{z}|$, where $\mathbf{x} = |\mathbf{z}|$ is a nonnegative vector. Thus the following number exists:

$$\lambda_0 = \max\{\ t \in \mathbb{R}\ :\ A\mathbf{x} \ge t\mathbf{x},\quad \text{for some}\ \ \mathbf{x} \ge 0,\ \mathbf{x} \ne 0\}.$$

Then clearly, $\lambda_0 \ge |\lambda| > 0$. We claim that $A\mathbf{x} = \lambda_0\mathbf{x}$, which shows that λ_0 is the largest eigenvalue with a nonnegative eigenvector: Indeed, suppose not: $A\mathbf{x} \ge \lambda_0\mathbf{x}$, *i.e.*, some components may be in equalities and some are in strict inequalities. Then, since A is positive, one can easily see that we have strict inequality:

$$A^2\mathbf{x} > \lambda_0 A\mathbf{x},$$

$$A\mathbf{y} > \lambda_0\mathbf{y},\ \ \mathbf{y} = A\mathbf{x},$$

which contradicts the maximality of λ_0. □

Note that a stochastic matrix A has 1 as an eigenvalue by (1) of Theorem 7.8. In the following, we will further show that, if a stochastic matrix A has all positive entries, then no eigenvalue of A other than $\lambda = 1$ satisfies $|\lambda| = 1$. Thus 1 is the largest eigenvalue with a nonnegative eigenvector.

Theorem 8.9 *Let A be a matrix with positive entries. If λ is an eigenvalue of A such that $|\lambda| = R(A)$, then $\lambda = R(A)$ and $\dim E_\lambda = 1$ with a basis $\mathbf{u} = [1\ 1\ \cdots 1]^T$.*

Proof: Let $\mathbf{x} = [x_1\ x_2\ \cdots\ x_k]^T$ be an eigenvector belonging to λ. Take a coordinate x_ℓ of $\mathbf{x}$ with the largest absolute value. Then

$$\begin{aligned} |\lambda||x_\ell| = |\lambda x_\ell| = \left|\sum_{j=1}^{k} a_\ell{}^j x_j\right| &\le \sum_{j=1}^{k} |a_\ell{}^j||x_j| \\ &\le \sum_{j=1}^{k} |a_\ell{}^j||x_\ell| = R_\ell(A)|x_\ell| \\ &\le R(A)|x_\ell|. \end{aligned}$$

Since $|\lambda| = R(A)$, the three inequalities are actually equalities, *i.e.*,

$$\begin{aligned} \left|\sum_{j=1}^{k} a_\ell{}^j x_j\right| &= \sum_{j=1}^{k} |a_\ell{}^j x_j| \\ \sum_{j=1}^{k} |a_\ell{}^j||x_j| &= \sum_{j=1}^{k} |a_\ell{}^j||x_\ell| \\ R_\ell(A) &= R(A). \end{aligned}$$

It is an easy exercise (see Problem 8.5 below) to show that the first equality means that all terms ${a_\ell}^j x_j$, $j = 1, \ldots, k$, are nonnegative multiple of some complex number z with $|z| = 1$. Thus ${a_\ell}^j x_j = c_j z$, or $x_j = \frac{c_j z}{{a_\ell}^j}$, for some nonnegative real numbers $c_1, \ldots, c_k$. The second equality means that ${a_\ell}^j = 0$ or $|x_j| = |x_\ell|$ for all $j = 1, \ldots, k$. Since ${a_\ell}^j > 0$ by assumption, we have the second possibility: $|x_j| = |x_\ell| = M$ for all $j = 1, \ldots, k$. This means that, for each $j = 1, \ldots, k$, the above computation is valid: *i.e.*, $R_j(A) = R(A)$ for all j. Thus one gets $M = |x_j| = \left|\frac{c_j z}{{a_\ell}^j}\right| = \frac{c_j}{{a_\ell}^j}$ or $x_j = Mz$ for all $j = 1, \ldots, k$, and so $\mathbf{x} = [x_1 \ \cdots \ x_k]^T = Mz[1\ 1\ \cdots 1]^T$. That is, $\mathbf{u} = [1\ 1\ \cdots 1]^T$ is a basis for E_λ, so that $\dim E_\lambda = 1$. Moreover,

$$A\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = \begin{bmatrix}\sum_{j=1}^k {a_1}^j\\ \vdots\\ \sum_{j=1}^k {a_k}^j\end{bmatrix} = \begin{bmatrix}R_1(A)\\ \vdots\\ R_k(A)\end{bmatrix} = \begin{bmatrix}R(A)\\ \vdots\\ R(A)\end{bmatrix} = R(A)\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix},$$

or $A\mathbf{x} = R(A)\mathbf{x}$. Thus if $\mathbf{x}$ is an eigenvector belonging to λ with $|\lambda| = R(A)$, then it is an eigenvector belonging also to $R(A)$ so that $\lambda = R(A)$. □

Problem 8.5 **(1)** Prove that $|u+v| \le |u| + |v|$, for any complex numbers u and v, and the equality holds if and only if $u = az$ and $v = bz$ where a, b are nonnegative real numbers and z is a complex number such that $|z| = 1$.

(2) Use induction to show $|\sum_{i=1}^k z_i| \le \sum_{i=1}^k |z_i|$, and equality holds if and only if $z_i = c_i z$ for some nonnegative real numbers c_i and a complex number z such that $|z| = 1$.

Note that $|\lambda| = C(A)$ also implies $\lambda = C(A)$ since $C(A) = R(A^T)$. If A is a stochastic matrix, then $C(A) = 1$, and so we have the following.

Corollary 8.10 *Let A be a stochastic matrix with positive entries, and let λ be an eigenvalue of A other then 1. Then $|\lambda| < 1$. Moreover,* $\dim E_1 = 1$.

Problem 8.6 Prove that the dimensions of the eigenspaces of a common eigenvalue λ of A and A^T are equal.

Therefore, if A is diagonalizable stochastic matrix with positive entries, then $\lim_{n\to\infty} A^n = L$ always exists by Theorem 7.6 and Corollary 8.10. For a general (nondiaginalizable) matrix, we have the following criteria.

Theorem 8.11 *Let A be any square matrix. Then $\lim_{n\to\infty} A^n$ exists if and only if the following conditions hold:*

(1) *An eigenvalue* λ *of* A *satisfies* $|\lambda| < 1$*, or* $\lambda = 1$.

(2) *If* 1 *is an eigenvalue of* A*, then* $\dim E_1$ *is the multiplicity of* 1.

Proof: Let J be the Jordan canonical form of A. Then $\lim\limits_{k\to\infty} A^k$ exists if and only if $\lim\limits_{k\to\infty} J^k$ exists. Let J_i be a Jordan block of order m in J of the form as **(1)** in page 348. Then, for $1 \le i < m$, $\lim\limits_{k\to\infty} \binom{k}{i}\lambda^{k-i} = 0$ if $|\lambda| < 1$, and $\lim\limits_{k\to\infty} \binom{k}{i}|\lambda|^{k-i} = \infty$ if $|\lambda| \ge 1$. It follows that $\lim\limits_{k\to\infty} J_i^k$ exists if and only if either $|\lambda| < 1$ or $\lambda = 1$ and $m = 1$. In the former case, $\lim\limits_{k\to\infty} J_i^k = \mathbf{0}$, and in the latter case, $\lim\limits_{k\to\infty} J_i^k = [1]$ and so each Jordan block of $\lambda = 1$ is a 1×1 matrix, which means $\dim E_1 =$ multiplicity of 1. □

In general, it may happen that $\dim E_1$ is less then the multiplicity of the eigenvalue $\lambda = 1$, as the following example shows. In fact, we have seen in Example 6.4 that the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not diagonalizable since 1 is an eigenvalue of multiplicity 2, while $\dim E_1 = 1$. Thus the second condition of Theorem 8.11 does not holds, and one can easily see that $\lim\limits_{n\to\infty} A^n$ does not exist since $A^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$.

The following theorem summarizes what have been discussed so far about a stochastic matrix.

Theorem 8.12 *Let* A *be a stochastic matrix with positive entries. Then*

(1) *The multiplicity of the eigenvalue* $\lambda = 1$ *of* A *is* 1.

(2) $\lim\limits_{k\to\infty} A^k = L$ *exists* (*that is, the diagonalizability of* A *is not necessary for this*)*, and* L *is also a stochastic matrix.*

(3) $AL = LA = L$.

(4) *The columns of* L *are identical. In fact, each column of* L *is equal to the unique eigenvector* $\mathbf{u}$ *of* A *belonging to* $\lambda = 1$ *which is also a probability vector.*

(5) *For any probability vector* $\mathbf{x}$, $\lim\limits_{k\to\infty} A^k\mathbf{x} = \mathbf{u}$.

Proof: (1) By the proof of Theorem 8.11, each Jordan block of 1 is 1×1, while $\dim E_1 = 1$ by Corollary 8.10: *i.e.*, 1 has only one Jordan block.

(2) The first part follows directly from part (1), Corollary 8.10, and Theorem 8.11. For the second part, since A^k are stochastic matrices, each entry of A^k is nonnegative for $k = 1, 2, \ldots$. Thus $[L]_i{}^j = \lim_{k\to\infty} [A^k]_i{}^j \geq 0$ for $0 \leq i, j \leq n$. Moreover, for $1 \leq j \leq n$,

$$\sum_{i=1}^{n} [L]_i{}^j = \sum_{i=1}^{n} \left[\lim_{k\to\infty} [A^k]_i{}^j\right] = \lim_{k\to\infty} \left[\sum_{i=1}^{n} [A^k]_i{}^j\right] = \lim_{k\to\infty} 1 = 1.$$

(3) Trivial, since $A(\lim_{k\to\infty} A^k) = (\lim_{k\to\infty} A^k)A = \lim_{k\to\infty} A^{k+1} = L$.

(4) Since $AL = L$, each column of L is an eigenvector of A belonging to $\lambda = 1$. But $\dim E_1 = 1$ means they columns of L are the same eigenvector belonging to $\lambda = 1$, and each of them is a probability vector by part (2).

(5) The vector $\mathbf{y} = L\mathbf{x}$ is also a probability vector by (2) and Lemma 7.7, and $A\mathbf{y} = AL\mathbf{x} = L\mathbf{x} = \mathbf{y}$. That is, $\mathbf{y}$ is also an eigenvector of A belonging to $\lambda = 1$, and so $\mathbf{y} = \mathbf{u}$. $\square$

Note that (5) in Theorem 8.12 means that the eventual distribution of the objects depends only on the sum of all the components of the probability vector, but not the initial distribution.

The vector $\mathbf{u}$ in Theorem 8.12 is called the **stationary vector** or the steady state of A.

Remark: Actually, in Theorem 8.12, the positiveness of the entries of a stochastic matrix may be weakened: That is, Theorem 8.12 still holds if some power of the stochastic matrix A has only positive entries. For example, for the matrix $A = \begin{bmatrix} 0.8 & 0 & 0.6 \\ 0 & 0.3 & 0.4 \\ 0.2 & 0.7 & 0 \end{bmatrix}$, A^2 has positive entries. To prove this assertion, it is good enough to reprove Corollary 8.10.

In fact, suppose that there is $d > 1$ such that A^d has only positive entries. Then it is clear that the entries of $A^{d+1} = (A^d)A$ are all positive since A is a stochastic matrix. If λ is an eigenvalue of A such that $|\lambda| = 1$, then λ^d and λ^{d+1} are eigenvalues of A^d and A^{d+1}, respectively, with absolute value 1. Thus, by Corollary 8.10, $\lambda^d = \lambda^{d+1} = 1$, which means $\lambda = 1$. Note that $E_1(A) \subseteq E_1(A^d)$, and $\dim E_1(A^d) = 1$, which means $E_1(A) = E_1(A^d)$ and $\dim E_1(A) = 1$.

8.3.4 Linear differential equations II

Now, we go back to a system of linear differential equations

$$\mathbf{y}' = A\mathbf{y} \quad \text{with an initial condition} \quad \mathbf{y}(0) = \mathbf{y}_0.$$

Its solution is known as $\mathbf{y}(t) = e^{tA}\mathbf{y}_0$ (see Theorem 7.16). In particular, if A is diagonalizable with n eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^n$ belonging to the eigenvalues λ_i's, this solution can be written as

$$\mathbf{y}(t) = e^{tA}\mathbf{y}_0 = c_1 e^{\lambda_1 t}\mathbf{v}^1 + c_2 e^{\lambda_2 t}\mathbf{v}^2 + \cdots + c_n e^{\lambda_n t}\mathbf{v}^n.$$

For an arbitrary square matrix A (not necessarily diagonalizable), find its Jordan canonical form $Q^{-1}AQ = J$. Then, the solution $\mathbf{y}(t) = e^{tA}\mathbf{y}_0$ is

$$\begin{aligned}
e^{tA}\mathbf{y}_0 &= Qe^{tJ}Q^{-1}\mathbf{y}_0 \\
&= \begin{bmatrix}\mathbf{u}^1 \ \mathbf{u}^2 \ \cdots \ \mathbf{u}^n\end{bmatrix} \begin{bmatrix} e^{tJ_1} & 0 & \cdots & 0 \\ 0 & e^{tJ_2} & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & & & e^{tJ_s} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\
&= \begin{bmatrix}\mathbf{u}^1 \ \mathbf{u}^2 \ \cdots \ \mathbf{u}^n\end{bmatrix} \begin{bmatrix} e^{\lambda_1 t}e^{tN_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t}e^{tN_2} & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & & & e^{\lambda_s t}e^{tN_s} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix},
\end{aligned}$$

where $Q^{-1}\mathbf{y}_0 = (c_1, \ldots, c_n)$ and the $\mathbf{u}^i$'s are generalized eigenvectors of A. In particular, if J is just a Jordan block with corresponding generalized eigenvectors $\mathbf{u}^i$ of order k, then the solution becomes:

$$\begin{aligned}
e^{tA}\mathbf{y}_0 &= e^{\lambda t}Qe^{tN}Q^{-1}\mathbf{y}_0 \\
&= e^{\lambda t}\begin{bmatrix}\mathbf{u}^1 \ \mathbf{u}^2 \ \cdots \ \mathbf{u}^n\end{bmatrix} \begin{bmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \ddots & \frac{t^{n-2}}{(n-2)!} \\ & & 1 & \ddots & \\ & & & \ddots & t \\ 0 & & & & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\
&= e^{\lambda t}\left(\left(\sum_{k=0}^{n-1} c_{k+1}\frac{t^k}{k!} \right)\mathbf{u}^1 + \left(\sum_{k=0}^{n-2} c_{k+2}\frac{t^k}{k!} \right)\mathbf{u}^2 + \cdots + c_n\mathbf{u}^n \right).
\end{aligned}$$

Example 8.12 Solve the linear differential equation $\mathbf{y}' = A\mathbf{y}$ with initial condition $\mathbf{y}(0) = \mathbf{y}_0$, where

$$A = \begin{bmatrix} 4 & -3 & -1 \\ 1 & 0 & -1 \\ -1 & 2 & 3 \end{bmatrix}, \quad \mathbf{y}_0 = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}.$$

Solution: (1) The characteristic polynomial of A is $\det(\lambda I - A) = \lambda^3 - 7\lambda^2 + 16\lambda - 12 = (\lambda - 3)(\lambda - 2)^2$. Since $\mathbf{x}^1 = (-1, -1, 1)$ and $\mathbf{x}^3 = (2, 1, -1)$ are linearly independent eigenvectors belonging to $\lambda = 2$ and $\lambda = 3$, respectively, one can compute the Jordan canonical form of A as follows:

$$J = Q^{-1}AQ = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix},$$

where

$$J_1 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \quad J_2 = [3], \quad \text{and } Q = \begin{bmatrix} -1 & 1 & 2 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

(2) Let $\mathbf{y} = Q\mathbf{x}$. Then the given system changes to $\mathbf{x}' = J\mathbf{x}$ with

$$\mathbf{x}(0) = Q^{-1}\mathbf{y}(0) = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 1 \end{bmatrix},$$

and the solution of this new system is given by

$$\mathbf{x}(t) = e^{tJ}\mathbf{x}(0) = \begin{bmatrix} e^{tJ_1} & \mathbf{0} \\ \mathbf{0} & e^{tJ_2} \end{bmatrix} \begin{bmatrix} 5 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} e^{2t} & te^{2t} & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 5 \\ 5 \\ 1 \end{bmatrix},$$

since

$$e^{tJ_1} = e^{2t} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad e^{tJ_2} = e^{3t}.$$

(3) Thus, we get

$$\begin{aligned} \mathbf{y}(t) &= Q\mathbf{x}(t) = \begin{bmatrix} -1 & 1 & 2 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} e^{2t} & te^{2t} & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 5 \\ 5 \\ 1 \end{bmatrix} \\ &= e^{2t} \left((5 + 5t) \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right) + e^{3t} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}. \end{aligned}$$

□

Example 8.13 Solve a system of linear differential equations $\mathbf{y}'(t) = A\mathbf{y}(t)$, where

$$A = \begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix}.$$

Solution: In Example 8.6 or 8.7, we have found a transition matrix $Q = [\mathbf{x}^1\ \ \mathbf{x}^2\ \ \mathbf{x}^3] = \begin{bmatrix} 2 & 0 & 1 \\ 4 & 0 & 0 \\ -2 & -1 & 2 \end{bmatrix}$ so that

$$Q^{-1}AQ = J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} & \\ & [1] \end{bmatrix} = \begin{bmatrix} J_1 & \mathbf{0} \\ \mathbf{0} & J_2 \end{bmatrix}.$$

Thus, the solution $\mathbf{y}(t) = e^{tA}\mathbf{y}_0$ is

$$\begin{aligned} e^{tA}\mathbf{y}_0 &= Q\begin{bmatrix} e^{J_1} & \mathbf{0} \\ \mathbf{0} & e^{J_2} \end{bmatrix} Q^{-1}\mathbf{y}_0 = e^tQ\begin{bmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} Q^{-1}\mathbf{y}_0 \\ &= \begin{bmatrix} 2 & 0 & 1 \\ 4 & 0 & 0 \\ -2 & -1 & 2 \end{bmatrix} \begin{bmatrix} e^t & te^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^t \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ 2 & -\frac{3}{2} & -1 \\ 1 & -\frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \\ &= c_1\begin{bmatrix} 4te^t + e^t \\ 8te^t \\ -4te^t \end{bmatrix} + c_2\begin{bmatrix} -3te^t \\ e^t - 6te^t \\ 3te^t \end{bmatrix} + c_3\begin{bmatrix} -2te^t \\ -4te^t \\ 2te^t + e^t \end{bmatrix}. \end{aligned}$$

Or, if we set $Q^{-1}\mathbf{y}_0 = (d_1, d_2, d_3)$, then

$$\begin{aligned} \mathbf{y}(t) &= e^t[\mathbf{x}^1\ \ t\mathbf{x}^1 + \mathbf{x}^2\ \ \mathbf{x}^3]\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \\ &= e^t((d_1 + d_2t)\mathbf{x}^1 + d_2\mathbf{x}^2 + d_3\mathbf{x}^3). \end{aligned}$$

□

Problem 8.7 Solve the system of linear differential equations $\mathbf{y}' = A\mathbf{y}$ with the initial condition $\mathbf{y}(0) = \mathbf{y}_0$, where

$$A = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 1 \\ 9 & 3 & -4 \end{bmatrix}, \quad \mathbf{y}^0 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

8.4 Cayley-Hamilton theorem

As we saw in earlier chapters, the association of the characteristic polynomial with each matrix is very useful in studying matrices. In this section, using this association of the polynomials with matrices we prove one more useful theorem, called the *Cayley-Hamilton theorem*, which makes the calculation of matrix polynomials simple, and has many applications to real problems.

Let $f(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0$ be a polynomial, and let A be an $n \times n$ square matrix. The matrix defined by

$$f(A) = a_m A^m + a_{m-1}A^{m-1} + \cdots + a_1 A + a_0 I_n$$

is called a **matrix polynomial** of A. For example, if $f(x) = x^2 - 2x + 2$ and $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, then

$$\begin{aligned} f(A) &= A^2 - 2A + 2I_2 \\ &= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} - 2\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}. \end{aligned}$$

Problem 8.8 Let λ be an eigenvalue of A and $\mathbf{x}$ an eigenvector belonging to λ. If $f(x)$ is any polynomial, then $f(\lambda)$ is an eigenvalue of the matrix polynomial $f(A)$.

Theorem 8.13 (Cayley-Hamilton) *For any $n \times n$ matrix A, if $f(\lambda) = \det(\lambda I - A)$ is the characteristic polynomial of A, then $f(A) = \mathbf{0}$.*

Proof: By Theorem 8.1, any square matrix A is similar to the Jordan canonical form $J = \begin{bmatrix} J_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & J_s \end{bmatrix} = Q^{-1}AQ$, or $A = QJQ^{-1}$. Then clearly $f(A) = Qf(J)Q^{-1}$, and

$$J^k = \begin{bmatrix} J_1^k & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & J_s^k \end{bmatrix} \quad \text{and} \quad f(J) = \begin{bmatrix} f(J_1) & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & f(J_s) \end{bmatrix}.$$

On the other hand, the characteristic polynomial f of A is factorized by the characteristic polynomials f_j's of the Jordan block J_j's: In fact, by Exercise **4.21**,

$$\begin{aligned} \det(\lambda I - A) = \det(\lambda I - J) &= \det(\lambda I - J_1) \cdots \det(\lambda I - J_s) \\ \text{or} \quad f(\lambda) &= f_1(\lambda) \ \cdots \ f_s(\lambda), \end{aligned}$$

where $f_j(\lambda) = \det(\lambda I - J_j)$ is the characteristic polynomial of J_j. Thus, it is good enough to show that $g(J) = \mathbf{0}$ for the characteristic polynomial $g(\lambda)$ of a single Jordan block $J = aI + N$ with an eigenvalue a of multiplicity m. Note that $N^m = \mathbf{0}$ and $g(\lambda) = \det(\lambda I - J) = (\lambda - a)^m$. Hence

$$g(J) = (J - aI)^m = (aI + N - aI)^m = N^m = \mathbf{0}.$$

Thus, $f(J_j) = f_1(J_j)\ \cdots\ f_j(J_j)\ \cdots\ f_s(J_j) = \mathbf{0}$ for each $j = 1, \ldots, s$. $\square$

Example 8.14 The characteristic polynomial of

$$A = \begin{bmatrix} 3 & 6 & 6 \\ 0 & 2 & 0 \\ -3 & -12 & -6 \end{bmatrix}$$

is $f(\lambda) = \det(\lambda I - A) = \lambda^3 + \lambda^2 - 6\lambda$, and

$$\begin{aligned} f(A) &= A^3 + A^2 - 6A \\ &= \begin{bmatrix} 27 & 78 & 54 \\ 0 & 8 & 0 \\ -27 & -102 & -54 \end{bmatrix} + \begin{bmatrix} -9 & -42 & -18 \\ 0 & 4 & 0 \\ 9 & 30 & 18 \end{bmatrix} \\ &\quad - 6\begin{bmatrix} 3 & 6 & 6 \\ 0 & 2 & 0 \\ -3 & -12 & -6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

$\square$

Remark: It is interesting that the last part of the proof of the Cayley-Hamilton Theorem 8.13 can be replaced by a direct computation of $f(J)$ for a Jordan block J of order m as follows: Recall that

$$\begin{aligned} J^k &= \sum_{j=0}^{m-1} \binom{k}{j} \lambda^{k-j} N^j \\ &= \lambda^k I + \binom{k}{1}\lambda^{k-1} N + \cdots + \binom{k}{m-1}\lambda^{k-(m-1)} N^{m-1} \\ &= \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots & \binom{k}{m-1}\lambda^{k-m+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \binom{k}{m-2}\lambda^{k-m+2} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ 0 & \cdots & \cdots & 0 & \lambda^k \end{bmatrix}. \end{aligned}$$

Therefore, for any polynomial $p(\lambda) = b_\ell \lambda^\ell + b_{\ell-1}\lambda^{\ell-1} + \cdots + b_1\lambda + b_0$,

$$\begin{aligned} p(J) &= b_\ell J^\ell + b_{\ell-1}J^{\ell-1} + \cdots + b_1 J + b_0 \\ &= \begin{bmatrix} p(\lambda) & p'(\lambda) & \cdots & \frac{p^{(m-1)}(\lambda)}{(m-1)!} \\ 0 & p(\lambda) & \cdots & \frac{p^{(m-2)}(\lambda)}{(m-2)!} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & p(\lambda) \end{bmatrix}, \end{aligned}$$

where $p^{(i)}(\lambda)$ is the i-th derivative of $p(\lambda)$. In particular, if $p(\lambda)$ is the characteristic polynomial $f(\lambda)$ of A, and a is an eigenvalue of multiplicity m, then $f(\lambda) = (\lambda - a)^m$ and so $f^{(i)}(a) = 0$ for $i = 0, 1, \ldots, m-1$. Thus $f(J) = \mathbf{0}$.

8.4.1 Application to inverse matrices

The Cayley-Hamilton theorem can be used to find the inverse of a nonsingular matrix. If $f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$ is the characteristic polynomial of a matrix A, then

$$\begin{aligned} \mathbf{0} = f(A) &= A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I, \\ \text{or} \qquad -a_0 I &= (A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_1 I)A. \end{aligned}$$

Since $a_0 = f(0) = \det(0I - A) = \det(-A) = (-1)^n \det A$, A is nonsingular if and only if $a_0 = (-1)^n \det A \neq 0$. Therefore, if A is nonsingular,

$$A^{-1} = -\frac{1}{a_0}(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_1 I).$$

Example 8.15 The characteristic polynomial of the matrix

$$A = \begin{bmatrix} 4 & 2 & -2 \\ -5 & 3 & 2 \\ -2 & 4 & 1 \end{bmatrix}$$

is $f(\lambda) = \det(\lambda I_3 - A) = \lambda^3 - 8\lambda^2 + 17\lambda - 10$, and the Cayley-Hamilton theorem yields

$$A^3 - 8A^2 + 17A = 10I_3.$$

Hence

$$\begin{aligned} A^{-1} &= \frac{1}{10}(A^2 - 8A + 17I_3) \\ &= \frac{1}{10}\begin{bmatrix} 10 & 6 & -6 \\ -39 & 7 & 18 \\ -30 & 12 & 13 \end{bmatrix} - \frac{8}{10}\begin{bmatrix} 4 & 2 & -2 \\ -5 & 3 & 2 \\ -2 & 4 & 1 \end{bmatrix} + \frac{17}{10}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{10}\begin{bmatrix} -5 & -10 & 10 \\ 1 & 0 & 2 \\ -14 & -20 & 22 \end{bmatrix}. \end{aligned}$$

□

Problem 8.9 Let A and B be square matrices, not necessarily of the same size, and let $f(\lambda) = \det(\lambda I - A)$ be the characteristic polynomial of A. Show that $f(B)$ is invertible if and only if A has no eigenvalue in common with B.

8.4.2 Application to matrix polynomials

The Cayley-Hamilton theorem can also be used to simplify the calculation of matrix polynomials. Let $p(\lambda)$ be any polynomial and let $f(\lambda)$ be the characteristic polynomial of an $n \times n$ matrix A. By the Euclidean division algorithm of polynomials, one can find two polynomials $q(\lambda)$ and $r(\lambda)$ such that

$$p(\lambda) = q(\lambda)f(\lambda) + r(\lambda),$$

where the degree of $r(\lambda)$ less than that of $f(\lambda)$. Then

$$p(A) = q(A)f(A) + r(A).$$

By the Cayley-Hamilton theorem, $f(A) = \mathbf{0}$ and

$$p(A) = r(A).$$

Thus the problem of evaluating a polynomial of an $n \times n$ matrix A, in particular A^k, can be reduced to the problem of evaluating a polynomial of degree less than n.

Example 8.16 The characteristic polynomial of the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is $f(\lambda) = \lambda^2 - 2\lambda - 3$. Let $p(\lambda) = \lambda^4 - 7\lambda^3 - 3\lambda^2 + \lambda + 4$ be a polynomial. A straightforward calculation shows that

$$p(\lambda) = (\lambda^2 - 5\lambda - 10)f(\lambda) - 34\lambda - 26.$$

Therefore

$$\begin{aligned} p(A) &= (A^2 - 5A + 10)f(A) - 34A - 26I \\ &= -34A - 26I \\ &= -34\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} - 26\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -60 & -68 \\ -68 & -60 \end{bmatrix}. \end{aligned}$$

□

Problem 8.10 For the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, evaluate the matrix polynomial $A^5 + 3A^4 + A^3 - A^2 + 4A + 6I$.

8.4.3 Computations of A^k and e^A II

Once we have the Cayley-Hamilton Theorem, we can use this to compute A^k or e^A without finding the Jordan canonical form of A or a transition matrix Q. Let $f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$ be the characteristic polynomial of A. Then $f(A) = \mathbf{0}$ implies that

$$A^n = -a_{n-1}A^{n-1} - \cdots - a_1A - a_0I.$$

Thus, for any $k \geq n$, the power A^k can be computed by a matrix polynomial of degree less than n. This fact implies that the computation of $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$, an infinite series of powers of A, can also be reduced to that of a matrix polynomial $g(\lambda)$ of degree at most $n-1$. However, the computation of the coefficients of g might not be easy. This problem can be handled if we consider a matrix polynomial $g(\lambda)$ of the smallest degree such that $g(A) = \mathbf{0}$. If we require the coefficient of the highest power of λ in g to be 1, such a polynomial exists uniquely, and is called the **minimal polynomial** of A. By the Cayley-Hamilton Theorem, the degree of g is less than or equal to that of f. Moreover, by the definition of the minimal polynomial g, it is easy to see that $g(\lambda)$ is a factor of $f(\lambda)$. The following facts are well known:

(1) If the eigenvalues of A are all distinct, the minimal polynomial is the characteristic polynomial:

$$g(\lambda) = f(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_n),$$

where λ_i's are all distinct n eigenvalues of A.

(2) For each eigenvalue λ_i with multiplicity $m_i > 1$, $g(\lambda)$ has a factor $(\lambda - \lambda_i)^{p_i}$ of highest power p_i for λ_i, where p_i is the order of the largest Jordan block among the Jordan blocks with λ_i on the diagonal. Thus, if $\dim E_{\lambda_i} = m_i$: *i.e.*, λ_i has m_i linearly independent eigenvectors so that $p_i = 1$, then $(\lambda - \lambda_i)$ is a factor of the highest power in λ_i of $g(\lambda)$. In particular, if A is diagonalizable, then the minimal polynomial is $g(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_s)$, where λ_i's are distinct eigenvalues of A.

(3) In general, if A has s distinct eigenvalues $\lambda_1, \ldots, \lambda_s$ with multiplicities $m_1, \ldots, m_s$, respectively, and $1 \le p_i \le m_i$, $i = 1, \ldots, s$, denote the orders of the largest Jordan blocks belonging to λ_i, then the minimal polynomial is

$$g(\lambda) = (\lambda - \lambda_1)^{p_1} \cdots (\lambda - \lambda_s)^{p_s}$$

so that the degree of $g(\lambda)$ is $q = \sum_{i=1}^{s} p_i \le n$.

Lat $g(\lambda) = \lambda^q + a_{q-1}\lambda^{q-1} + \cdots a_1\lambda + a_0$ be the minimal polynomial of A. Then $g(A) = \mathbf{0}$ implies that, for any $k \ge q$, we may set up

$$\begin{aligned} A^k &= x_0 I + x_1 A + \cdots + x_{q-1}A^{q-1}, \text{ or} \\ e^A &= x_0 I + x_1 A + \cdots + x_{q-1}A^{q-1}, \end{aligned}$$

for some coefficients x_i's. We can now determine those coefficients x_i's as follows: Let $\{\mathbf{v}^1, \ldots, \mathbf{v}^p\}$ be a maximal chain of generalized eigenvectors belonging to an eigenvalue λ. Then

$$A\mathbf{v}^p = \lambda\mathbf{v}^p + \mathbf{v}^{p-1}, \ \ldots, \ A\mathbf{v}^2 = \lambda\mathbf{v}^2 + \mathbf{v}^1, \ A\mathbf{v}^1 = \lambda\mathbf{v}^1.$$

From the last equation, $A^k\mathbf{v}^1 = \lambda^k\mathbf{v}^1 = (x_0 I + x_1 A + \cdots + x_{q-1}A^{q-1})\mathbf{v}^1$. Thus, we get

$$\lambda^k = x_0 + \lambda x_1 + \cdots + \lambda^{q-1}x_{q-1}.$$

Similarly from the second to the last equation, $A^k\mathbf{v}^2 = \lambda^k\mathbf{v}^2 + k\lambda^{k-1}\mathbf{v}^1$. On the other hand, by a direct computation we get:

$$\begin{aligned} A^k\mathbf{v}^2 &= (x_0 I + x_1 A + \cdots + x_{q-1}A^{q-1})\mathbf{v}^2 \\ &= (x_0 + \lambda x_1 + \quad \cdots \quad + \lambda^{q-1}x_{q-1})\mathbf{v}^2 \\ &\quad\ + (x_1 + 2\lambda x_2 + \cdots + (q-1)\lambda^{q-2}x_{q-1})\mathbf{v}^1. \end{aligned}$$

Since $\mathbf{v_2}$ and $\mathbf{v_1}$ are linearly independent, we have

$$\begin{aligned} \lambda^k &= x_0 + \lambda x_1 + \lambda^2 x_2 + \cdots + \qquad \lambda^{q-1}x_{q-1} \\ k\lambda^{k-1} &= \qquad 1x_1 + 2\lambda x_2 + \cdots + (q-1)\lambda^{q-2}x_{q-1}. \end{aligned}$$

Note that the second equation is just the derivative of the first one with respect to λ. By the same computations with $A^k\mathbf{v}^3, \ldots, A^k\mathbf{v}^p$, one gets the following additional $p-2$ equations:

$$\begin{array}{rcl}\binom{k}{2}\lambda^{k-2} & = & \frac{2\cdot 1}{2!}x_2 + \frac{3\cdot 2}{2!}\lambda x_3 + \cdots + \binom{q-1}{2}\lambda^{q-3}x_{q-1} \\ & \vdots & \\ \binom{k}{p-1}\lambda^{k-(p-1)} & = & \frac{(p-1)!}{(p-1)!}x_{p-1} + \frac{p!}{(p-1)!}\lambda x_p + \cdots + \binom{q-1}{p-1}\lambda^{q-p}x_{q-1}.\end{array}$$

Note also that these equations are the higher derivatives of the first equation. Thus, for each eigenvalue λ of multiplicity m, we get a p equations, and so, all together, a system of $q = \sum_{i=1}^{s} p_i$ equations in q unknowns x_j's. It is easy to see that the coefficient matrix is invertible so that the unknowns $x_0, x_1, \ldots, x_{q-1}$ are uniquely determined.

The same computation can be applied to

$$e^A = x_0 I + x_1 A + \cdots + x_{q-1}A^{q-1} = h(A).$$

For a maximal chain of generalized eigenvectors $\{\mathbf{v}^1, \ldots, \mathbf{v}^p\}$ belonging to an eigenvalue λ, by computing both sides of

$$e^A\mathbf{v}^j = (x_0 I + x_1 A + \cdots + x_{q-1}A^{q-1})\mathbf{v}^j,$$

for $j = 1, \ldots, p$, separately, we get the following p equations

$$\begin{array}{rcl} e^\lambda & = & h(\lambda) = x_0 + \lambda x_1 + \cdots + \lambda^{q-1}x_{q-1} \\ e^\lambda & = & h'(\lambda) = \qquad x_1 + \cdots + (q-1)\lambda^{q-2}x_{q-1} \\ & \vdots & \\ \frac{1}{(p-1)!}e^\lambda & = & h^{(p-1)}(\lambda) = \binom{p-1}{p-1}\lambda x_{p-1} + \cdots + \binom{q-1}{p-1}\lambda^{q-p}x_{q-1}.\end{array}$$

With the same argument as the case of A^k, one can determine the coefficients x_j's uniquely.

Example 8.17 Compute A^{-1}, A^k and e^A for

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 1 & 0 & 3 \end{bmatrix}.$$

Solution: The characteristic polynomial of A is $f(\lambda) = (\lambda - 2)^4$. Since, by a direct computation, $(A-2I)^3 = \mathbf{0}$ and $(A-2I)^2 \neq \mathbf{0}$, the Jordan canonical form J of A must be of the form

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

Thus the minimal polynomial of A is $g(\lambda) = (\lambda - 2)^3 = \lambda^3 - 6\lambda^2 + 12\lambda - 8$, and so

$$A^{-1} = \frac{1}{8}(A^2 - 6A + 12I) = \frac{1}{8}\begin{bmatrix} 6 & -2 & -1 & -2 \\ 0 & 4 & -4 & 0 \\ 0 & 0 & 4 & 0 \\ 2 & -2 & 1 & 2 \end{bmatrix}.$$

Let $A^k = x_0 I + x_1 A + x_2 A^2$. Then, for $\lambda = 2$, we get

$$\begin{aligned} 2^k &= x_0 + 2x_1 + 2^2 x_2, \\ k2^{k-1} &= x_1 + 2 \cdot 2x_2, \\ k(k-1)2^{k-2} &= 2x_2. \end{aligned}$$

The solution is

$$\begin{aligned} x_0 &= 2^{k-1}(k-1)(k-2), \\ x_1 &= k(2-k)2^{k-1}, \\ x_2 &= k(k-1)2^{k-3}, \end{aligned}$$

so that

$$\begin{aligned} A^k &= 2^{k-1}(k-1)(k-2)I + k(2-k)2^{k-1}A + k(k-1)2^{k-3}A^2 \\ &= \begin{bmatrix} (2-k)2^{k-1} & k2^{k-1} & 2^{k-3}(3k+k^2) & k2^{k-1} \\ 0 & 2^k & k2^k & 0 \\ 0 & 0 & 2^k & 0 \\ -k2^{k-1} & k2^{k-1} & k(k-1)2^{k-3} & (k+2)2^{k-1} \end{bmatrix}. \end{aligned}$$

Similarly, let $e^A = x_0 I + x_1 A + x_2 A^2$. Then, for $\lambda = 2$,

$$\begin{aligned} e^2 &= x_0 + 2x_1 + 2^2 x_2 \\ e^2 &= x_1 + 2 \cdot 2x_2 \\ e^2 &= 2x_2. \end{aligned}$$

The solution is $x_0 = e^2$, $x_1 = -e^2$, $x_2 = \frac{1}{2}e^2$, so that

$$e^A = e^2(I - A + \frac{1}{2}A^2) = e^2 \begin{bmatrix} 0 & 1 & \frac{3}{2} & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & \frac{1}{2} & 2 \end{bmatrix}.$$

□

8.5 Exercises

8.1. Show that if A nonsingular, then A^{-1} has the same block structure in its Jordan canonical form as A does.

8.2. Find the number of linearly independent eigenvectors for each of the following matrices:

$$(1) \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}, \ (2) \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}, \ (3) \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}.$$

8.3. Solve the system of linear equations

$$\begin{cases} (1-i)x & + & (1+i)y & = & 2-i \\ (1+i)x & + & (1+i)y & = & 1+3i. \end{cases}$$

8.4. Find the Jordan-canonical form for $A = \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}$, and compute e^A.

8.5. Let $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ and $\mathbf{y}_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

(1) Solve $\mathbf{y}^n = A\mathbf{y}^{n-1}$ with $\mathbf{y}_0$.

(2) Solve $\mathbf{y}' = A\mathbf{y}$ with $\mathbf{y}(0) = \mathbf{y}_0$.

8.6. Let $A = \begin{bmatrix} -6 & 24 & 8 \\ -1 & 8 & 4 \\ 2 & -12 & -6 \end{bmatrix}$ and $\mathbf{y}_0 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

(1) Solve $\mathbf{y}^n = A\mathbf{y}^{n-1}$ with $\mathbf{y}_0$.

(2) Solve $\mathbf{y}' = A\mathbf{y}$ with $\mathbf{y}(1) = \mathbf{y}_0$.

8.7. Solve the initial value problem

$$\begin{cases} y_1' & = & -y_1 & & +2y_3, & & y_1(0) = & -2 \\ y_2' & = & 2y_1 & +y_2 & -2y_3, & & y_2(0) = & 0 \\ y_3' & = & -2y_1 & & +3y_3, & & y_3(0) = & -1\,. \end{cases}$$

8.8. Consider a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

(1) Find a necessary and sufficient condition for A to be diagonalizable.
(2) The characteristic polynomial for A is $f(t) = t^2 - (a+d)t + (ad - bc)$. Show that $f(A) = 0$.

8.9. For each of the following matrices, find a polynomial of which the matrix is a root.

$$(1) \begin{bmatrix} 2 & 5 \\ 1 & -3 \end{bmatrix}, \quad (2) \begin{bmatrix} 2 & -3 \\ 7 & -4 \end{bmatrix}, \quad (3) \begin{bmatrix} 1 & 4 & -3 \\ 0 & 3 & 1 \\ 0 & 2 & -1 \end{bmatrix}.$$

8.10. Verify that each of the matrices below satisfies its own characteristic polynomial and from these results compute A^{-1}, if it exists.

$$(1) \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix}, \quad (2) \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad (3) \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

8.11. For $f(x) = 3x^3 + x^2 - 2x + 3$, compute $f(A)$ for

$$(1)\ A = \begin{bmatrix} 1 & 2 & 0 \\ -3 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \qquad (2)\ A = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}.$$

8.12. Show that a Jordan block J is similar to its transpose, $J^T = P^{-1}JP$, by the permutation matrix $P = [\mathbf{e}^n \ \cdots \ \mathbf{e}^1]$. Deduce that every matrix is similar to its transpose.

8.13. For any square matrix $A = [a_i{}^j]$, define $\|A\| = \max\{|a_i{}^j| \ : \ 1 \le i, j \le n\}$. Prove the followings for any square matrices A, B and $c \in \mathbb{C}$:

(1) $\|A\| \ge 0$ and equality holds if and only if A is the zero matrix.
(2) $\|cA\| = |c|\|A\|$.
(3) $\|A + B\| \le \|A\| + \|B\|$.
(4) $\|AB\| \le n\|A\| \cdot \|B\|$.

8.14. Let A be a stochastic matrix, and $Q^{-1}AQ = J$ be the Jordan canonical form of A.

(1) Prove $\|A^k\| \le 1$ for any positive integer k.
(2) Deduce that $\{\|J^k\| \ : \ k = 1, 2, \ldots\}$ is bounded.
(3) Prove that each Jordan block belonging to the eigenvalue $\lambda = 1$ of A is 1×1.
(4) Prove that $\lim_{k \to \infty} A^k$ exists if and only if any eigenvalue λ of A such that $|\lambda| = 1$ is in fact $\lambda = 1$.

8.15. Compute A^{-1}, A^k and e^A for

$$(1) \begin{bmatrix} i & 0 \\ 1 & i \end{bmatrix}, \quad (2) \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad (3) \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

8.16. Determine whether the following statements are true or false, in general, and justify your answers.

(1) Any square matrix is similar to a triangular matrix.

(2) If a matrix A has exactly k linearly independent eigenvectors, then the Jordan canonical form of A has k Jordan blocks.

(3) If a matrix A has k distinct eigenvalues, then the Jordan canonical form of A has k Jordan blocks.

(4) If a 4×4 matrix A has eigenvalues 1 and 2, each of multiplicity 2, such that $\dim E_1 = 2$ and $\dim E_2 = 1$, then the Jordan canonical form of A has three Jordan blocks.

(5) If $\lambda_1, \ldots, \lambda_k$ are k distinct eigenvalues of A with multiplicities m_i and $\dim E_{\lambda_i} \neq m_i$, then A is not diagonalizable.

(6) For any Jordan block J with eigenvalue λ, $\det e^J = e^\lambda$.

(7) For any square matrix A, A and A^T have the same Jordan canonical forms.

(8) The inverse of any invertible matrix A can be written as a polynomial in A.