MATH36001 Theory of Eigensystems 2015

We will show how eigenvalues and eigenvectors are related to important canonical forms that display the structure of a matrix.

1 Basic Definitions

1.1 Eigenvalues and Eigenvectors

A vector $x \in \mathbb{C}^n$ is called an **eigenvector** of $A \in \mathbb{C}^{n\times n}$ if x is nonzero and Ax is a multiple of x, that is, there is a $\lambda \in \mathbb{C}$ such that

$$Ax = \lambda x, \quad x \neq 0. \tag{1}$$

The complex scalar λ is called the **eigenvalue** of A associated with the eigenvector x. The pair (λ, x) is called an **eigenpair** of A. The eigenvalue associated with a given eigenvector is unique. However, each eigenvalue has many eigenvectors associated with it since if x is an eigenvector corresponding to the eigenvalue λ and α is a nonzero scalar then αx is also an eigenvector with associated eigenvalue λ.

We can rewrite (1) in the form $(\lambda I - A)x = 0$, $x \neq 0$ showing that $\lambda I - A$ is a singular matrix and therefore any eigenvalue λ must satisfy the equation

$$\det(\lambda I - A) = 0.$$

By the cofactor expansion of determinants, it is easy to see that $p(\lambda) = \det(\lambda I - A)$ is a polynomial of degree n in λ. We call $p(\lambda)$ the **characteristic polynomial** of A. By the Fundamental Theorem of Algebra a polynomial of degree n has exactly n real or complex zeros, counting multiplicities. Hence an $n \times n$ matrix has exactly n eigenvalues, although they are not necessarily distinct. The set of all eigenvalues of A is called the **spectrum** of A and will be denoted by

$$\Lambda(A) = \{\lambda_1, \dots, \lambda_n\}.$$

We sometimes denote by $\lambda_i(A)$ the ith eigenvalue of A in some (usually arbitrary) ordering. The **algebraic multiplicity** of λ is its multiplicity as a zero of the characteristic polynomial $p(\lambda) = \det(\lambda I - A)$.

1.2 Invariant Subspaces

A subspace $\mathcal{X}$ of $\mathbb{C}^n$ is an **invariant subspace** for A if $A\mathcal{X} \subseteq \mathcal{X}$, that is, $x \in \mathcal{X}$ implies $Ax \in \mathcal{X}$.

Theorem 1 *Let the columns of $X \in \mathbb{C}^{n\times p}$, $p \leq n$, form a basis for a subspace $\mathcal{X}$ of $\mathbb{C}^n$. Then $\mathcal{X}$ is an invariant subspace for A if and only if $AX = XB$ for some $B \in \mathbb{C}^{p\times p}$. When the latter equation holds, the spectrum of B is contained within that of A.*

Proof. Let $X = [\, x_1 \ \ \dots \ \ x_p \,]$ and $Y = AX = [\, y_1 \ \ \dots \ \ y_p \,]$ be partitioned by columns. If $\mathcal{X}$ is an invariant subspace of A then $y_i \in \mathcal{X}$ and since X is a basis for $\mathcal{X}$, y_i can be expressed as a linear combination of columns of X, that is, $y_i = Xb_i$, for some vector $b_i \in \mathbb{C}^p$. If we let $B = [\, b_1 \ \ \dots \ \ b_p \,]$ then $AX = XB$.

Conversely, if $AX = XB$ for some $B \in \mathbb{C}^{p\times p}$, then equating the jth column of AX with the jth column of XB gives $Ax_j = \sum_{i=1}^{p} x_i b_{ij} \in \operatorname{span}\{x_1,\dots,x_p\} = \mathcal{X}$. Since $Ax_j \in \mathcal{X}$, $j = 1,\dots,p$ and $x_1,\dots,x_p$ span $\mathcal{X}$, it follows that $Ax \in \mathcal{X}$ for all $x \in \mathcal{X}$.

Let (λ, u) be an eigenpair of B. If $AX = XB$ for some $B \in \mathbb{C}^{p\times p}$ then $AXu = XBu = \lambda Xu$ and $Xu \neq 0$ since the columns of X are independent, so (λ, Xu) is an eigenpair of A. Since this is true for each eigenpair of B, we have that $\Lambda(B) \subset \Lambda(A)$. □

1.3 Similarity, Unitary Similarity

Let $A, B \in \mathbb{C}^{n\times n}$. The matrices A and B are **similar** if there exists a nonsingular matrix P such that

$$B = P^{-1}AP. \tag{2}$$

(2) is called a **similarity transformation** and P is the **transforming matrix**.

Theorem 2 *Let A and B be similar, say $B = P^{-1}AP$. Then A and B have the same eigenvalues, and x is an eigenvector of A with associated eigenvalue λ if and only if $P^{-1}x$ is an eigenvector of B with associated eigenvalue λ.*

Proof. $Ax = \lambda x \iff (P^{-1}AP)(P^{-1}x) = \lambda(P^{-1}x)$, so (λ, x) is an eigenpair of A iff $(\lambda, P^{-1}x)$ is an eigenpair of $P^{-1}AP = B$.

Note: we can also show that $\det(A - \lambda I) = \det(P^{-1}(A - \lambda I)P) = \det(P^{-1}AP - \lambda I)$. Thus A and B have the same eigenvalues, and the algebraic multiplicities are preserved. □

A and B are said to be **unitarily similar** if there is a unitary matrix U such that $B = U^*AU$. If A and B are real, then they are said to be **orthogonally similar** if there is a real, orthogonal matrix U such that $B = U^TAU$.

If a matrix A is similar to a diagonal matrix then A is said to be **diagonalizable** or **simple**.

1.4 How is Similarity Used in Solving Problems?

The **similarity method** is a strategy frequently used for solving problems. Here is an outline of the method.

Step 1 : Choose a matrix B similar to A for which the problem is easier to solve.

Step 2 : Solve the problem using the matrix B instead of A (the B-problem).

Step 3 : Interpret the solution to the B-problem in terms of the matrix A.

Example 1 Given $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$, find each entry in A^{1010}.

Step 1 (Find easier B similar to A): Choose $B = \begin{bmatrix} 5 & 0 \\ 0 & -2 \end{bmatrix}$, where $P = \begin{bmatrix} 1 & -4 \\ 1 & 3 \end{bmatrix}$; then $A = PBP^{-1}$.
Step 2 (Solve B-problem): $B^{1010} = \begin{bmatrix} 5^{1010} & 0 \\ 0 & (-2)^{1010} \end{bmatrix}$.
Step 3 (Interpret B-solution): $A^{1010} = (PBP^{-1})^{1010} = PB^{1010}P^{-1}$. Thus

$$A^{1010} = \frac{1}{7}\begin{bmatrix} 3(5^{1010}) + 2^{1012} & 4(5^{1010}) - 2^{1012} \\ 3(5^{1010}) - 3(2^{1010}) & 4(5^{1010}) + 3(2^{1010}) \end{bmatrix}.$$

Example 2 In certain problems in economics the state of a system is described by a matrix $S_n = I + A + A^2 + \cdots + A^n$ at time n where A is a given matrix. Use the similarity method to investigate the behaviour of the system in the "long run", i.e., when n is large.

Suppose there is a nonsingular matrix P and two scalars λ and μ such that $A = P\left[\begin{smallmatrix}\lambda & 0\\ 0 & \mu\end{smallmatrix}\right]P^{-1}$.
Step 1: Choose $B = \left[\begin{smallmatrix}\lambda & 0\\ 0 & \mu\end{smallmatrix}\right]$.
Step 2: Let $T_n = I + B + B^2 + \cdots + B^n$, then

$$T_n = \begin{bmatrix} 1+\lambda+\lambda^2+\cdots+\lambda^n & \\ & 1+\mu+\mu^2+\cdots+\mu^n \end{bmatrix} = \begin{bmatrix} \frac{1-\lambda^{n+1}}{1-\lambda} & \\ & \frac{1-\mu^{n+1}}{1-\mu} \end{bmatrix}$$

for all $n \geq 0$ and $\lambda, \mu \neq 1$. If $|\lambda|, |\mu| < 1$, T_n approximates $\left[\begin{smallmatrix}\frac{1}{1-\lambda} & 0\\ 0 & \frac{1}{1-\mu}\end{smallmatrix}\right]$ when n is large.
Step 3: Since $S_n = PT_nP^{-1}$, it follows that S_n is approximately PTP^{-1} for large enough n and so, even without knowing P we can predict that in the long run the system will reach a steady state. If we want a quantitative statement, we calculate P and can then say how large n has to be for S_n to approximate PTP^{-1} (the steady state matrix) within say 'x' decimal place accuracy.

2 Canonical Forms

We now consider the following question:

> *What is the simplest form a square matrix can take under similarity transformations?*

2.1 The Schur Decomposition: a Triangularization

We start by considering unitary similarities $A = UBU^*$. Unitary similarity is computationally simpler than similarity because the conjugate transpose is much easier to compute than the inverse. Schur's theorem asserts that we can reduce any $n \times n$ matrix to a triangular one by a unitary similarity transformation.

Theorem 3 (Schur's theorem) *Let $A \in \mathbb{C}^{n\times n}$. Then there exists a unitary matrix U and an upper triangular matrix T such that*

$$T = U^{-1}AU = U^*AU.$$

Proof. The proof is by induction on n. The result clearly holds for $n = 1$. Let us show that it holds for $n = k$, given that it holds for $n = k-1$. Let $A \in \mathbb{C}^{k\times k}$. Let λ be an eigenvalue of A and x an associated eigenvector normalized such that $x^*x = 1$. Let U_1 be any unitary matrix having x as its first column (there are many such matrices: just take any orthonormal basis of $\mathbb{C}^k$ whose first member is x and let U_1 be the matrix whose columns are the members of the basis). Write $U_1 = [\, x \quad W\,]$. Since the columns of W are orthogonal to x, $W^*x = 0$. Let $A_1 = U^*AU$. Then Hojo use Gram Schmidt

$$A_1 = \begin{bmatrix} x^* \\ W^* \end{bmatrix} A \,[\, x \quad W\,] = \begin{bmatrix} x^*Ax & x^*AW \\ W^*Ax & W^*AW \end{bmatrix}.$$

Since $Ax = \lambda x$, it follows that $x^*Ax = \lambda$ and $W^*Ax = \lambda W^*x = 0$. Let $\widehat{A} = W^*AW$. Then

$$A_1 = \begin{bmatrix} \lambda & x^*AW \\ 0 & \widehat{A} \end{bmatrix}.$$

But $\widehat{A} \in \mathbb{C}^{(k-1)\times(k-1)}$, so by the induction hypothesis there exists a unitary matrix $\widehat{U}_2$ and an upper triangular matrix $\widehat{T}$ such that $\widehat{T} = \widehat{U}_2^*\widehat{A}\widehat{U}_2$. Define

$$U_2 = \begin{bmatrix} 1 & 0_{1\times(k-1)} \\ 0_{(k-1)\times 1} & \widehat{U}_2 \end{bmatrix} = \begin{bmatrix} 1 & \\ & \widehat{U}_2 \end{bmatrix}.$$

Then U_2 is unitary and

$$U_2^*A_1U_2 = \begin{bmatrix} \lambda & x^*AW\widehat{U}_2 \\ 0 & \widehat{U}_2^*\widehat{A}\widehat{U}_2 \end{bmatrix} = \begin{bmatrix} \lambda & x^*AW\widehat{U}_2 \\ 0 & \widehat{T} \end{bmatrix},$$

which is upper triangular. Let us call this matrix T, and let $U = U_1U_2$. Then $T = U_2^*A_1U_2 = U_2^*U_1^*AU_1U_2 = U^*AU$. $\square$

The unitary similarity transformation $T = U^{-1}AU$ can also be written

$$A = UTU^*.$$

Written in this way, we call it a **Schur decomposition** of A. Note that since $\det(U)\det(U^*) = \det(UU^*) = 1$,

$$\det(\lambda I - A) = \det(\lambda I - T) = \prod_{i=1}^{n}(\lambda - t_{ii}),$$

the diagonal elements of T are the eigenvalues of A. The Schur decomposition is not unique (the eigenvalues can be made to appear in any order on the diagonal of T). The columns of U are called **Schur vectors**. With the exception of u_1, the Schur vectors are not, in general, eigenvectors of A.

The most widely used methods for numerically computing the eigenvalues of a matrix consist of approximating a Schur decomposition of that matrix; the QR algorithm is an example of such method.

2.2 Diagonalizable Matrices

Is any matrix A diagonalizable?

In other words can we always find a nonsingular P such that $P^{-1}AP$ is diagonal?

2.2.1 The Class of Matrices Unitary Similar to a Diagonal Matrix

A matrix A is **normal** if $AA^* = A^*A$. The class of normal matrices contains the important subclasses of Hermitian and unitary matrices. For normal matrices Schur's theorem takes a particularly nice form: the triangular matrix T turns out to be diagonal. This special form of Schur's theorem is called the *spectral theorem*.

Theorem 4 (spectral theorem) *Let $A \in \mathbb{C}^{n\times n}$. Then A is normal if and only if there is a unitary matrix U and a diagonal matrix Λ such that*

$$A = U\Lambda U^*.$$

Proof. Let $A = UTU^*$ be the Schur decomposition of A. If A is normal then it is easy to show that T is normal (see Exercise 4). Since a normal and triangular matrix is diagonal (see Exercise 8), T is diagonal.

Conversely, if A admits the decomposition $A = U\Lambda U^*$ with U unitary and Λ diagonal then, since diagonal matrices commute,

$$AA^* = (U\Lambda U^*)(U\Lambda^* U^*) = U\Lambda\Lambda^* U^* = U\Lambda^*\Lambda U^* = (U\Lambda^* U^*)(U\Lambda U)^* = A^*A. \qquad \square$$

The next theorem gives another characterization of normal matrices.

Theorem 5 *$A \in \mathbb{C}^{n\times n}$ is normal if and only if it has n orthogonal eigenvectors.*

Proof. From Theorem 4 we have that normal matrices have an orthonormal basis of eigenvectors. Conversely, suppose $u_1, \dots, u_n$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A: $Au_j = \lambda_j u_j$, $j = 1, \dots, n$. Let $U = [\, u_1 \ \cdots \ u_n \,]$ and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Then $AU = U\Lambda$ or equivalently $A = U\Lambda U^*$ and A is normal by Theorem 4. $\square$

2.2.2 Matrices Similar to a Diagonal Matrix

We now consider similarity transformations $P^{-1}AP$ where P is not necessarily unitary. The next theorem identifies the large class of diagonalizable matrices.

Theorem 6 *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.*

Proof. Suppose A is diagonalizable, that is, there exist $P = [\, p_1, \dots, p_n \,]$ nonsingular and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ such that $\Lambda = P^{-1}AP$ or equivalently,

$$AP = P\Lambda \iff [\, Ap_1, \dots, Ap_n \,] = [\, \lambda_1 p_1, \dots, \lambda_n p_n \,] \iff Ap_i = \lambda_i p_i, \ i = 1, \dots, n.$$

Thus the columns of P are eigenvectors of A and are linearly independent since P is nonsingular.

Conversely, if A has n linearly independent eigenvectors $p_1, \dots, p_n$ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$, then with the notation above the matrix P is nonsingular and we have that $AP = P\Lambda$ which shows that A is similar to Λ. $\square$

We now show that when A's eigenvalues $\lambda_1, \dots, \lambda_n$ are distinct the corresponding eigenvectors $x_1, \dots, x_n$ are linearly independent. Suppose the eigenvectors x_i, $i = 1, \dots, n$ are linearly dependent so that $\sum_{i=1}^n \alpha_i x_i = 0$ with not all the constants $\alpha_1, \dots, \alpha_n$ equal to zero. We may assume that $\alpha_n \neq 0$ by renumbering the eigenvalues if necessary. Multiplying $\sum_{i=1}^n \alpha_i x_i = 0$ by $(A - \lambda_1 I)$ gives

$$0 = (A - \lambda_1 I)\sum_{i=1}^n \alpha_i x_i = \sum_{i=2}^n \alpha_i(\lambda_i - \lambda_1)x_i.$$

Multiplying the above expression by $A - \lambda_2 I$ gives

$$0 = (A - \lambda_2 I) \sum_{i=2}^{n} \alpha_i (\lambda_i - \lambda_1) x_i = \sum_{i=3}^{n} \alpha_i (\lambda_i - \lambda_1)(\lambda_i - \lambda_2) x_i.$$

Continuing similar multiplications with $A - \lambda_3 I$, then $A - \lambda_4 I$ and so on, we obtain that

$$0 = \alpha_n (\lambda_n - \lambda_{n-1})(\lambda_n - \lambda_{n-2}) \ldots (\lambda_n - \lambda_1) x_n.$$

But this is a contradiction when the λ_i are distinct, since $\alpha_n \neq 0$ and $x_n \neq 0$. Hence a sufficient condition for $A \in \mathbb{C}^{n\times n}$ to have n linearly independent eigenvectors is to have n distinct eigenvalues. We conclude from Theorem 6 that A is diagonalizable.

Theorem 7 *A matrix with distinct eigenvalues is diagonalizable.*

Theorems 6 and 7 show that matrices not similar to a diagonal matrix necessarily have multiple eigenvalues and less than n linearly independent eigenvectors. Note that the matrix $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ has 0 as an eigenvalue of multiplicity 2 and that there is only one eigenvector (any multiple of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$) associated with 0. Hence $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalizable.

2.3 The Jordan Canonical Form

If not all matrices are diagonalizable, what is the simplest form that a matrix, in general, can take under a similarity transformation?

Theorem 8 (Jordan canonical form) *Any matrix $A \in \mathbb{C}^{n\times n}$ can be expressed in the Jordan canonical form*

$$X^{-1}AX = J = \begin{bmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_p(\lambda_p) \end{bmatrix}, \tag{3a}$$

$$J_k = J_k(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k \times m_k}, \tag{3b}$$

where X is nonsingular and $m_1 + m_2 + \cdots + m_p = n$.

Proof. For a proof of this theorem (which is not easy) see [1], [2] or [3]. □

The $m_k \times m_k$ matrices J_k are called **Jordan blocks**. The Jordan matrix J is unique up to the ordering of the blocks J_k, but the transforming matrix X is not unique. The Jordan block J_k has only one linearly independent eigenvector. For example for $m_k = 3$,

$$(J_k(\lambda_k) - \lambda_k I)x = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \;\Rightarrow\; x_2 = x_3 = 0$$

so that x is a multiple of e_1.

The Jordan matrix J displays several important properties.

(i) The number p of Jordan blocks is the number of linearly independent eigenvectors of A (or equivalently of J). Thus the matrix A is diagonalizable if and only if $p = n$.

(ii) The algebraic multiplicity (defined in section 1.1) of a given eigenvalue λ is the sum of the dimensions of the Jordan blocks in which λ appears.

(iii) The **geometric multiplicity** of λ is the number of Jordan blocks associated with λ. Thus the geometric multiplicity of λ is the number of linearly independent eigenvectors associated with λ or, equivalently, $\dim(\mathrm{null}(A - \lambda I))$.

(iv) An eigenvalue λ is **defective** if it appears in a Jordan block of size greater than 1, or, equivalently, if its algebraic multiplicity exceeds its geometric multiplicity. A matrix is **defective** if it has a defective eigenvalue, or, equivalently, if it does not have a complete set of linearly independent eigenvectors.

Example 3 Find a Jordan matrix J of a matrix A having as characteristic polynomial $p(\lambda) = (\lambda - 1)^3(\lambda - 2)^4$ if the geometric multiplicities are also known: $\dim(\mathrm{null}(A - I)) = 2$ and $\dim(\mathrm{null}(A - 2I)) = 3$.

$\dim(\mathrm{null}(A - I)) = 2$ implies that there are two Jordan blocks associated with $\lambda_1 = 1$ and since λ_1 has algebraic multiplicity 3, one of the block is of order 2, the other is of order 1. Reasoning in a similar way for the second eigenvalue we obtain

$$J = \mathrm{diag}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, [\,1\,], \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, [\,2\,], [\,2\,] \right).$$

2.3.1 The Structure of a Jordan Matrix

A Jordan matrix is not completely determined in general by the knowledge of the eigenvalues and their algebraic and geometric multiplicities. (E.g. an eigenvalue λ of algebraic multiplicity 6 and geometric multiplicity 3 could appear in Jordan blocks of sizes $2, 2, 2$ or $4, 1, 1$ or $3, 2, 1$.) One must also know the sizes of the Jordan blocks corresponding to each eigenvalue. For a given matrix $A \in \mathbb{C}^{n\times n}$, the Jordan canonical form of A (but not the similarity that transforms A to Jordan canonical form) can be determined by the following procedure:

1. Find all the distinct eigenvalues of A, perhaps by finding the roots of the characteristic polynomial.

2. For each distinct eigenvalue λ_i of A form $(A - \lambda_i I), (A - \lambda_i I)^2, \dots,$ and analyse the sequence of ranks of these matrices as follows:

   - the smallest value of k_i for which $\mathrm{rank}(A - \lambda_i I)^{k_i}$ attains its minimum value is the order of the largest Jordan block corresponding to λ_i. This minimum value is called the **index** of the eigenvalue λ_i.
   - The number of Jordan blocks of size k in J with eigenvalue λ_i is

$$\mathrm{rank}(A - \lambda_i I)^{k-1} + \mathrm{rank}(A - \lambda_i I)^{k+1} - 2\,\mathrm{rank}(A - \lambda_i I)^k. \qquad (4)$$

Example 4 Determine the Jordan canonical form of the matrix

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

The characteristic polynomial is $p(\lambda) = (\lambda - 2)^4$ so $\lambda = 2$ is an eigenvalue of multiplicity 4. Since $\dim(\text{null}(A - 2I)) = 2$, the Jordan form consists of two blocks. To determine their order (2 and 2 or 3 and 1), we find that $(A - 2I)^2 = O$ and therefore $k_1 = k_2 = 2$ so that the order of the largest Jordan block corresponding to $\lambda = 2$ is 2. Consequently,

$$J = \text{diag}\left(\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \right).$$

Example 5 Determine the Jordan canonical form of a 14×14 matrix A having the following eigenvalues and sequences of ranks:

	$\text{rank}(A - \lambda I)^k$				
k	1	2	3	4	5
$\lambda = 1$	11	10	9	9	9
$\lambda = 2$	12	10	10	10	10
$\lambda = 3$	12	11	10	9	9

$\underline{\lambda = 1}$: $\dim(\text{null}(A - I)) = n - \text{rank}(A - I) = 14 - 11 = 3$ so there are 3 Jordan blocks with eigenvalue $\lambda = 1$. Also since $\text{rank}(A - I)^3 = \text{rank}(A - I)^4$, the index of $\lambda = 1$ is 3 and therefore the size of the largest Jordan block with eigenvalue $\lambda = 1$ is 3. The formula in (4) gives

number of blocks of size 1: $14 + 10 - 2 \times 11 = 2$.

number of blocks of size 2: $11 + 9 - 2 \times 10 = 0$.

number of blocks of size 3: $10 + 9 - 2 \times 9 = 1$.

Hence $\lambda = 1$ has algebraic multiplicity $2 \times 1 + 1 \times 3 = 5$.

$\underline{\lambda = 2}$: $\dim(\text{null}(A - 2I)) = n - \text{rank}(A - 2I) = 14 - 12 = 2$ so there are 2 Jordan blocks with eigenvalue $\lambda = 2$. Also since $\text{rank}(A - I)^2 = \text{rank}(A - I)^3$, the index of $\lambda = 2$ is 2 and therefore the size of the largest Jordan block with eigenvalue $\lambda = 2$ is 2. The formula in (4) gives

number of blocks of size 1: $14 + 10 - 2 \times 12 = 0$.

number of blocks of size 2: $12 + 10 - 2 \times 10 = 2$.

Hence $\lambda = 2$ has algebraic multiplicity $2 \times 2 = 4$.

$\underline{\lambda = 3}$: $\dim(\text{null}(A - I)) = n - \text{rank}(A - 3I) = 14 - 12 = 2$ so there are 3 Jordan blocks with eigenvalue $\lambda = 3$. Also since $\text{rank}(A - I)^4 = \text{rank}(A - I)^5$, the size of the largest Jordan block with eigenvalue $\lambda = 3$ is 4. The formula in (4) gives

number of blocks of size 1: $14 + 11 - 2 \times 12 = 1$.

number of blocks of size 2: $12 + 10 - 2 \times 11 = 0$.

number of blocks of size 3: $11 + 9 - 2 \times 10 = 0$.

number of blocks of size 4: $10 + 9 - 2 \times 9 = 1$.

Hence $\lambda = 1$ has algebraic multiplicity $1 \times 1 + 1 \times 4 = 5$.

From this information,

$$J = \operatorname{diag}\left(\begin{bmatrix} 1 & 1 & \\ & 1 & 1 \\ & & 1 \end{bmatrix}, [\,1\,], [\,1\,], \begin{bmatrix} 2 & 1 \\ & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ & 2 \end{bmatrix}, \begin{bmatrix} 3 & 1 & & \\ & 3 & 1 & \\ & & 3 & 1 \\ & & & 3 \end{bmatrix}, [\,3\,] \right).$$

2.3.2 Eigenvectors and Generalized Eigenvectors

Let $X^{-1}AX = J$ be the Jordan canonical form of $A \in \mathbb{C}^{n \times n}$. Then

$$AX = XJ \tag{5}$$

and the columns of X in positions $1, m_1 + 1, m_1 + m_2 + 1, \ldots, m_1 + m_2 + \cdots + m_{p-1} + 1$ are eigenvectors of A and these are linearly independent since X is nonsingular. The other columns of X are called **generalized eigenvectors**. Equating the first m_1 columns of (5) corresponding to the first Jordan block J_1 yields

$$Ax_1 = \lambda_1 x_1, \quad Ax_i = \lambda_1 x_i + x_{i-1},\ i = 2, \ldots, m_1. \tag{6}$$

The vectors $x_1, x_2, \ldots, x_{m_1}$ are called a **Jordan chain**. The columns of X form p Jordan chains

$$\{x_1, \ldots, x_{m_1}\}, \{x_{m_1+1}, \ldots, x_{m_1+m_2}\}, \ldots, \{x_{n-m_p+1}, \ldots, x_n\}.$$

Setting $i = 2$ in (6) gives $(A - \lambda_1 I)x_2 = x_1$ so that

$$(A - \lambda_1 I)^2 x_2 = (A - \lambda_1 I)x_1 = 0. \tag{7}$$

Also, (6) with $i = 3$ and (7) yield $(A - \lambda_1 I)^3 x_3 = (A - \lambda_1 I)^2 x_2 = 0$. More generally,

$$(A - \lambda_1 I)^j x_j = 0, \quad j = 1, \ldots, m_1.$$

If λ_1 is distinct from the other eigenvalues, this shows that the linearly independent vectors $x_1, \ldots, x_i$, $i \le m_1$ form a basis for null$\big((A - \lambda_1 I)^i\big)$. If several Jordan blocks have the same eigenvalue, then the corresponding chains must be combined to obtain the basis for the null spaces.

Example 6 Determine the Jordan canonical form, the eigenvectors and generalized eigenvectors of

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

Since $\det(A-\lambda I) = (2-\lambda)(\lambda-4)^2$, it follows that $\Lambda(A) = \{2,4\}$. It is easily seen that the matrix A has one (linearly independent) eigenvector say $x_1 = [\,0 \quad -1 \quad 1\,]^T$, corresponding to $\lambda = 2$ and one (linearly independent) eigenvector say $x_2 = [\,2 \quad -2 \quad 0\,]^T$ associated with $\lambda_2 = 4$. From the equation $(A-4I)x_3 = x_2$, it is found that $x_3 = [\,1 \quad 0 \quad 0\,]^T$ is a generalized eigenvector of A associated with $\lambda_2 = 4$. Hence a Jordan basis for A is $\{x_1, x_2, x_3\}$ and the matrix

$$X = [\,x_1 \quad x_2 \quad x_3\,] = \begin{bmatrix} 0 & 2 & 1 \\ -1 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

transforms A in to the Jordan form

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

There is no numerically stable way to compute Jordan canonical forms. A simple example makes this clear. If $A_\epsilon = \left[\begin{smallmatrix} \epsilon & 0 \\ 1 & 0 \end{smallmatrix}\right]$ and $\epsilon \neq 0$, then A_ϵ has the Jordan form $J_\epsilon = \left[\begin{smallmatrix} 0 & 0 \\ 0 & \epsilon \end{smallmatrix}\right]$ with $X_\epsilon = \left[\begin{smallmatrix} 0 & \epsilon \\ 1 & 1 \end{smallmatrix}\right]$. Then $J_\epsilon \to \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ as $\epsilon \to 0$, which cannot be the Jordan form of the nonzero matrix $A_0 = \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]$. In fact, A_0 has $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ as its Jordan form. Since the Jordan form of a matrix need not be a continuous function of the entries of the matrix, small variations in the entries of the matrix can result in large variations in the Jordan form. There is no hope of computing such an object in a stable way.

3 Polynomials and Matrices

As an application of the similarity method we shall prove the Cayley–Hamilton theorem.

Theorem 9 (Cayley–Hamilton theorem) *If p is the characteristic polynomial of an $n \times n$ matrix A, then $p(A) = O$.*

Proof. Let $A = XJX^{-1}$ be the Jordan canonical decomposition of A as in (3) and let $p(\lambda) = (\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_p)^{m_p}$ be its characteristic polynomial. Then,

$$p(A) = p(XJX^{-1}) = Xp(J)X^{-1} = X \begin{bmatrix} p(J_1) & & & \\ & p(J_2) & & \\ & & \ddots & \\ & & & p(J_p) \end{bmatrix} X^{-1}$$

with $p(J_k) = (J_k - \lambda_1 I)^{m_1}(J_k - \lambda_2 I)^{m_2}\cdots(J_k - \lambda_p I)^{m_p}$, $k = 1,\dots,p$. Now

$$(J_k - \lambda_k I)^{m_k} = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}^{m_k} = 0$$

since $J_k - \lambda_k I$ is an $m_k \times m_k$ nilpotent matrix with index of nilpotency m_k. Hence $p(J_k) = O$, $k = 1, \dots, p$ and the theorem is proved. $\square$

For example $p(\lambda) = (\lambda - 1)^n$ is the characteristic polynomial of the $n \times n$ identity matrix I_n and clearly p satisfies $p(I) = O$. Notice that the polynomials $q_k(\lambda) = (\lambda - 1)^k$ of degree k less than n also satisfy $q_k(I) = (I - I)^k = O$. We call the monic (leading coefficient equal to one) polynomial q of lowest degree such that $q(A) = O$, the **minimal polynomial** of $A \in \mathbb{C}^{n\times n}$. The following result follows from the proof of the Cayley–Hamilton theorem.

Theorem 10 (Minimal polynomial) *Let A be an $n \times n$ matrix with s distinct eigenvalues $\lambda_1, \dots, \lambda_s$. The minimal polynomial of A is*

$$q(\lambda) = \prod_{i=1}^{s} (\lambda - \lambda_i)^{n_i},$$

where n_i is the dimension of the largest Jordan block in which λ_i appears.

A key property is that the minimal polynomial divides any other polynomial p for which $p(A) = O$ (see Exercise 15).

One important use of the Cayley–Hamilton theorem is to write powers A^k of $A \in \mathbb{C}^{n\times n}$, for $k \geq n$, as linear combinations of $I, A, A^2, \dots, A^{n-1}$ (see Exercise 13).

Exercises

1. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of A. Show that

$$\begin{aligned} \det(A) &= \lambda_1 \lambda_2 \dots \lambda_n, \\ \operatorname{trace}(A) &= \lambda_1 + \lambda_2 + \cdots + \lambda_n, \\ \operatorname{trace}(A^k) &= \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k, \;\; k = 1, 2, \dots. \end{aligned}$$

2. (a) Let $x_1, x_2, \dots, x_k$ be eigenvectors of A. Show that $\mathcal{S} = \operatorname{span}\{x_1, x_2, \dots, x_k\}$ is invariant under A.

(b) Let $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Show that the space $\mathcal{S} = \operatorname{span}\{e_1, e_2\}$ is invariant under A and is not spanned by eigenvectors of A. Here $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

3. Let $A = UTU^*$ be the Schur decomposition of $A \in \mathbb{C}^{n\times n}$. Let $u_1, \dots, u_n$ denote the columns of U and let $\mathcal{S}_j = \operatorname{span}\{u_1, \dots, u_j\}$ for $j = 1, \dots, n$. Show that each $\mathcal{S}_j$ is invariant under A. [Hint: Use Theorem 1].

4. Suppose $B = U^*AU$, where U is unitary.

(a) Prove that B is normal if A is and that B is Hermitian if A is.

(b) Show by example that neither of the properties in (a) is preserved under arbitrary similarity transformation $B = S^{-1}AS$, where S is nonsingular.

5. Let $A \in \mathbb{C}^{n\times n}$ be normal. Use the spectral theorem to prove the following.

(a) A is Hermitian ($A^* = A$) if and only if the eigenvalues of A are real.

(b) A is skew-Hermitian ($A^* = -A$) if and only if the eigenvalues of A are purely imaginary.

(c) A is unitary ($A^* = A^{-1}$) if and only if the eigenvalues of A lie on the unit circle in the complex plane.

6. This exercise shows that a matrix is nilpotent if and only if its eigenvalues are all zero.

(a) Show that if N is nilpotent then all its eigenvalues are zero.

(b) Show that a strictly upper or lower triangular matrix (the main diagonal is zero) is nilpotent.

(c) Show that if N is a matrix whose eigenvalues are all zero, then N is unitarily similar to a strictly upper triangular matrix. Deduce that N is nilpotent.

7. Let $A \in \mathbb{C}^{n\times n}$ be normal and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denotes its eigenvalues with corresponding eigenvectors $u_1, u_2, \ldots u_n$ normalized such that $u_i^* u_i = 1$. Show that

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^*.$$

This is called the **spectral representation** of A. Illustrate with the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

8. Show that a normal and triangular matrix is diagonal.

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

$$(a) \begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}, \quad (b) \begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & 2 & \\ & & & 2 \end{bmatrix}, \quad (c) \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 2 & \\ & & & 2 \end{bmatrix}.$$

10. Ascertain if the following matrices are similar to a diagonal matrix:

$$(a) \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad (b) \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \quad (c) \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad (d) \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}.$$

11. Let A be a 4×4 matrix with an eigenvalue $\lambda = 2$ of algebraic multiplicity 4.

(a) List all the possible Jordan forms for this matrix up to permutations of the Jordan blocks.

(b) What is the geometric multiplicity of $\lambda = 2$ for each of the Jordan forms listed in (a).

(c) Compute eigenvectors for each of the Jordan forms in (a).

(d) Determine the minimal polynomials of the matrices in (a). The Jordan form of a matrix determines its minimal polynomial. Show that the converse is in general not true.

12. Determine the Jordan canonical form, the eigenvectors and generalized eigenvectors of

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

13. Let $A = \left[\begin{smallmatrix} 3 & 1 \\ -2 & 0 \end{smallmatrix}\right]$. Use the Cayley–Hamilton theorem to express A^4 as a linear combination of I and A.

14. Let A be an $n \times n$ nonsingular matrix. Use the Cayley–Hamilton theorem to express A^{-1} as a polynomial in A.

15. Show that the minimal polynomial q of a matrix A divides any other polynomial p for which $p(A) = 0$.

16. Show that if $A \in \mathbb{C}^{n\times n}$ has minimal polynomial $q(A) = A^2 - A - I$ then $(I - \frac{1}{3}A)^{-1} = \frac{3}{5}(A+2I)$.

17. Find the characteristic polynomial and the minimal polynomial of the rank-1 matrix $uv^* \in \mathbb{C}^{n\times n}$.

18.* Let p be a polynomial and $A \in \mathbb{C}^{n\times n}$. Show that $p(A) = O$ if and only if $p(t)(tI - A)^{-1}$ is a polynomial in t. Deduce the Cayley–Hamilton theorem.

19. Let $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{m\times m}$ be given and suppose A and B have no eigenvalues in common, that is, $\Lambda(A) \cap \Lambda(B) = \emptyset$. Consider the matrix equation

$$AX - XB = O, \qquad X \in \mathbb{C}^{n\times m}. \tag{8}$$

(a) Show inductively that if (8) holds then $A^kX = XB^k$ for all $k = 1, 2, \ldots$.

(b) Use the Cayley–Hamilton theorem to show that (8) has only the solution $X = O$.

References

[1] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[2] James M. Ortega. *Matrix Theory: A Second Course*. Plenum Press, New York, 1987.

[3] Gilbert Strang. *Linear Algebra and Its Applications*. Harcourt Brace Jovanovich, San Diego, third edition, 1988.