Chapter 5 Structure of Square Matrices

5.1 Eigenvalues and Eigenvectors

Recall from Section 2.7 that a second order linear homogeneous differential equation with constant coefficients

$$y'' + a_1 y' + a_2 y = 0 (5.1)$$

is equivalent to a system of two first order differential equations

$$\mathbf{x}' = A\mathbf{x} \tag{5.2}$$

where

$$A = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} y \\ y' \end{bmatrix}$$

Also recall from from Section 2.4 that (5.1) has a complex solution of the form $y = \phi(t) = e^{st}$. Then (5.2) must have a solution of the form

$$\mathbf{x} = \boldsymbol{\phi}(t) = \begin{bmatrix} \phi(t) \\ \phi'(t) \end{bmatrix} = \begin{bmatrix} e^{st} \\ se^{st} \end{bmatrix} = e^{st} \begin{bmatrix} 1 \\ s \end{bmatrix} = e^{st} \mathbf{v}$$
(5.3)

where $\mathbf{v} = \operatorname{col}[1, s]$. Substituting $\mathbf{x} = e^{st}\mathbf{v}$ and $\mathbf{x}' = se^{st}\mathbf{v}$ into (5.2) and cancelling out the nonzero terms e^{st} , we obtain

$$A\mathbf{v} = s\mathbf{v} \tag{5.4}$$

(5.4) provides a necessary and sufficient condition to be satisfied by s in order for $\mathbf{x} = e^{st}\mathbf{v}$ to be solution of (5.2), and therefore, must be related to the characteristic equation of (5.1). Indeed, writing (5.2) in open form we get

$$\begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} 1 \\ s \end{bmatrix} = \begin{bmatrix} s \\ -a_1s - a_2 \end{bmatrix} = s \begin{bmatrix} 1 \\ s \end{bmatrix} = \begin{bmatrix} s \\ s^2 \end{bmatrix}$$

The first equation above is an identity, and the second is nothing but the characteristic equation of (5.1):

 $s^2 + a_1 s + a_2 = 0$

In the study above, (5.2) is derived from the second order differential equation in (5.1). This not only results in a coefficient matrix A having a special structure, but also constrains the vector **v** to the form in (5.3). Now suppose that we started directly

with (5.2) where no special structure was imposed on the coefficient matrix A. Then we could still assume a solution of the form $\mathbf{x} = e^{st}\mathbf{v}$, where \mathbf{v} is not restricted to a special form, and end up with (5.4).

Systems of differential equations is just one example where we come across the matrix equation (5.4). There are many other significant problems that lead to a similar equation. In fact, (5.4) is just a special case of a more general equation

$$\mathcal{A}(\mathbf{x}) = s\mathbf{x} \tag{5.5}$$

involving a linear operator \mathcal{A} defined on a vector space \mathbf{X} . This equation simply asks for such $\mathbf{x} \in \mathbf{X}$ that are not changed (except for a scaling) by \mathcal{A} , which is a significant question in many scientific and engineering problems.

Eigenvalues and Eigenvectors of A Square Matrix

Let $A \in \mathbf{C}^{n \times n}$. A complex scalar $s = \lambda$ that satisfies (5.4) for some nonzero vector $\mathbf{v} \in \mathbf{C}^{n \times 1}$ is called an **eigenvalue** of A, and \mathbf{v} is called an **eigenvector** of A associated with the eigenvalue λ .

Rewriting (5.4) as

$$(sI - A)\mathbf{v} = \mathbf{0}$$

we get an $n \times n$ homogeneous linear system. By Theorem 1.1 it has a nonzero solution if and only if the coefficient matrix is singular, or equivalently, if and only if

$$\det\left(sI - A\right) = 0$$

Treating s as a parameter, we observe that sI - A is a matrix whose elements are simple polynomials in s. The diagonal elements of sI - A are $s - a_{ii}$, and the offdiagonal elements are $-a_{ij}$. Therefore, det (sI - A) is also a polynomial in s, called the **characteristic polynomial** of A, denoted d(s). It is not too difficult to show that d(s) is an nth degree polynomial with unity leading coefficient (see Exercise 5.2). That is,

$$d(s) = \det(sI - A) = s^n + d_1 s^{n-1} + \dots + d_n$$

We thus reach the conclusion that equation (5.4) has a nonzero solution if and only if s is a root of the **characteristic equation**

$$d(s) = 0 \tag{5.6}$$

Since d(s) is an *n*th degree polynomial, by the fundamental theorem of algebra, the characteristic equation (5.6) has exactly *n* complex roots counting the multiplicities of the repeated roots (if any). Suppose that it has *k* distinct roots $\lambda_1, \ldots, \lambda_k$ with multiplicities n_1, \ldots, n_k , so that

$$d(s) = \prod_{i=1}^{\kappa} (s - \lambda_i)^{n_i}$$
(5.7)

where

 $\sum_{i=1}^{k} n_i = n$

Then each λ_i is an eigenvalue of A with **algebraic multiplicity** n_i . Any nonzero solution $\mathbf{v} = \mathbf{v}_i$ of

$$A\mathbf{v} = \lambda_i \mathbf{v}$$

is an eigenvector associated with λ_i . There are infinitely many eigenvectors associated with an eigenvalue. In fact, together with the zero vector, the set of all eigenvectors associated with λ_i form a subspace

$$\mathbf{K}_i = \ker\left(A - \lambda_i I\right)$$

which is called the **eigenspace** of A associated with λ_i . The dimension of \mathbf{K}_i

$$\nu_i = \dim (\mathbf{K}_i)$$

is called the **geometric multiplicity** of the eigenvalue λ_i .

Some immediate results concerning eigenvalues can be derived from the definition and from properties of determinants, and are listed below.

- a) A is singular if and only if it has a zero eigenvalue.
- b) A and A^t have the same characteristic polynomial, and therefore, the same eigenvalues. (However, eigenvectors of A^t are, in general, different from those of A.)
- c) Eigenvalues of a lower (upper) triangular matrix are its diagonal elements.

Property (a) follows directly from the definition of an eigenvalue. Property (b) is a consequence of the equality

$$\det (sI - A^t) = \det (sI - A)^t = \det (sI - A)$$

Finally, property (c) follows from

$$\det(sI - A) = \det \begin{bmatrix} s - a_{11} & 0 & \cdots & 0 \\ -a_{21} & s - a_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & s - a_{nn} \end{bmatrix}$$
$$= (s - a_{11})(s - a_{22}) \cdots (s - a_{nn})$$

Note that property (c) is also true for a diagonal matrix. In particular, I_n has the only eigenvalue $\lambda_1 = 1$ with algebraic multiplicity $n_1 = n$. Also, since $I_n - \lambda_1 I_n = O$, $\nu_1 = n$, and any nonzero vector is an eigenvector of I.

Example 5.1

Let us find the eigenvalues and eigenvectors of the matrix

$$A = \left[\begin{array}{rrr} 2 & -1 \\ -2 & 3 \end{array} \right]$$

The characteristic polynomial of A is

$$d(s) = \det \begin{bmatrix} s-2 & 1\\ 2 & s-3 \end{bmatrix} = (s-2)(s-3) - 2 = s^2 - 5s + 4 = (s-1)(s-4)$$

Hence A has two simple eigenvalues:

 $\lambda_1 = 1 \,, \quad \lambda_2 = 4$

An eigenvector associated with $\lambda_1 = 1$ is obtained by solving the equation

$$(A - \lambda_1 I)\mathbf{v} = \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0}$$

to be

$$\mathbf{v}_1 = \left[\begin{array}{c} 1\\1 \end{array} \right]$$

Similarly, solving

$$(A - \lambda_2 I)\mathbf{v} = \begin{bmatrix} -2 & -1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0}$$

we obtain an eigenvector associated with $\lambda_2 = 4$ as

$$\mathbf{v}_2 = \left[\begin{array}{c} 1\\ -2 \end{array} \right]$$

Note that any scalar multiple of \mathbf{v}_1 is also an eigenvector associated with $\lambda_1 = 1$, and the same is true for any scalar multiple of \mathbf{v}_2 . Thus $\mathbf{K}_1 = \operatorname{span}(\mathbf{v}_1)$ and $\mathbf{K}_2 = \operatorname{span}(\mathbf{v}_2)$.

Example 5.2

The characteristic equation of the matrix

$$A = \left[\begin{array}{cc} 0 & 1 \\ -5 & 2 \end{array} \right]$$

is

$$d(s) = \det \begin{bmatrix} s & -1 \\ 5 & s-2 \end{bmatrix} = s^2 - 2s + 5 = 0$$

The eigenvalues of A are the complex conjugate roots

 $\lambda_1 = 1 + 2i, \quad \lambda_2 = \lambda_1^* = 1 - 2i$

of the characteristic equation. The reader can verify that

$$\mathbf{v}_1 = \left[\begin{array}{c} 1\\ 1+2i \end{array} \right], \quad \mathbf{v}_2 = \left[\begin{array}{c} 1\\ 1-2i \end{array} \right]$$

are associated eigenvectors. This example shows that even when A is real its eigenvalues and eigenvectors may be complex. However, they appear in conjugate pairs.

Eigenvalues and Eigenvectors of A Linear Operator

Since any linear operator \mathcal{A} on a finite dimensional vector space \mathbf{X} over a field \mathbf{F} is represented by a square matrix $A \in \mathbf{F}^{n \times n}$, the concept of eigenvalues and eigenvectors can be generalized to such operators as illustrated by the following example.

Example 5.3

Let \mathcal{A} denote the reflection of the xy plane in the line y = x. That is, $\mathcal{A}(\mathbf{v})$ is the mirror image of the vector \mathbf{v} with respect to the reflecting line. It is not difficult to show that \mathcal{A} is a linear operator on the xy plane.

If a vector \mathbf{r} lies in this line then it is mapped into itself, that is, $\mathcal{A}(\mathbf{r}) = \mathbf{r}$. It follows that $\lambda = 1$ is an eigenvalue of \mathcal{A} , with \mathbf{r} being an associated eigenvector. On the other hand, if a vector \mathbf{s} lies in the line y = -x, which is orthogonal to the line y = x, then $\mathcal{A}(\mathbf{s}) = -\mathbf{s}$ so that $\lambda = -1$ is also an eigenvalue of \mathcal{A} with \mathbf{s} being an associated eigenvector.



Figure 5.1: Reflection of the xy plane in the line y = x.

The statements above, which are illustrated in Figure 5.1, can formally be shown by representing \mathcal{A} with a matrix. If we identify the xy plane with $\mathbf{R}^{2\times 1}$, then since the unit vectors along the x and y axes are mapped into each other, we have

$$\mathcal{A}(\mathbf{e}_1) = \mathbf{e}_2, \quad \mathcal{A}(\mathbf{e}_2) = \mathbf{e}_1$$

and \mathcal{A} is represented by the matrix

$$A = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right]$$

with respect to the canonical basis. The characteristic polynomial of A is

det
$$\begin{bmatrix} s & -1 \\ -1 & s \end{bmatrix} = s^2 - 1 = (s - 1)(s + 1)$$

Hence A has the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$. It is easy to verify that

$$\mathbf{v}_1 = \mathbf{r} = \begin{bmatrix} 1\\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \mathbf{s} = \begin{bmatrix} 1\\ -1 \end{bmatrix}$$

are associated eigenvectors.

Now let \mathcal{B} denote a rotation in the xy plane through an angle $0 < \theta < \pi$ counterclock-wise. Then there exists no vector in the plane whose image is a (real) multiple of itself. Hence \mathcal{B} has no eigenvectors in the plane. The reader can show that \mathcal{B} is represented by the matrix

$$B = \begin{bmatrix} \cos\theta & -\sin\theta\\ \sin\theta & \cos\theta \end{bmatrix}$$

with respect to the canonical basis. Calculating the characteristic polynomial of B as

$$\det \begin{bmatrix} s - \cos \theta & \sin \theta \\ -\sin \theta & s - \cos \theta \end{bmatrix} = s^2 - 2 \cos \theta s + 1$$

we observe that B has a pair of complex conjugate eigenvalues

$$\lambda_{1,2} = \cos\theta \mp i\sin\theta$$

Since B has no real eigenvalues, it has no real eigenvectors, verifying our observation. It does, however, have a pair of complex conjugate eigenvectors, which are

$$\mathbf{v}_{1,2} = \left[\begin{array}{c} 1\\ \mp i \end{array} \right]$$

The concept of eigenvalues and eigenvectors can also be generalized to linear operators on infinite dimensional vector spaces:

* Example 5.4

Let

$$\mathbf{X} = \{ \phi \, | \, \phi \in \mathcal{C}_{\infty}(\mathbf{R}, \mathbf{R}), \, \phi(0) = \phi(\pi) = 0 \}$$

Clearly, **X** is a subspace of $\mathcal{C}_{\infty}(\mathbf{R}, \mathbf{R})$. Let $\mathcal{A} : \mathbf{X} \to \mathbf{X}$ be defined as

$$\mathcal{A}(\phi) = \phi''$$

Then the eigenvalue problem

$$\mathcal{A}(\phi) = s\phi$$

is equivalent to determining those s for which the boundary value problem

$$y'' = sy$$
, $y(0) = y(\pi) = 0$

has a nonzero solution.

This boundary value problem, which has already been considered in Exercise 2.33, has a nonzero solution if and only if

$$s = \lambda_n = -n^2, \quad n = 1, 2, \dots$$

in which case a solution is

 $y = \phi_n(t) = \sin nt, \quad n = 1, 2, \dots$

We thus conclude that the operator \mathcal{A} has infinitely many eigenvalues $\lambda_n = -n^2$ with associated eigenvectors $\phi_n(t) = \sin nt$.

When \mathcal{A} is a linear operator defined on a function space as in this example, its eigenvectors are also called **eigenfunctions**.

In the rest of this chapter, we will deal with linear operators defined by square matrices only.

5.2 The Cayley-Hamilton Theorem

Let $A \in \mathbf{F}^{n \times n}$, and let

 $p(s) = p_0 s^m + p_1 s^{m-1} + \dots + p_{m-1} s + p_m$

be a polynomial in s with $p_i \in \mathbf{F}, i = 1, ..., m$. We define the matrix $p(A) \in \mathbf{F}^{n \times n}$ as

 $p(A) = p_0 A^m + p_1 A^{m-1} + \dots + p_{m-1} A + p_m I$

From the definition it follows that for arbitrary polynomials p and q

$$p(A) + q(A) = (p+q)(A)$$

and

p(A)q(A) = q(A)p(A) = (pq)(A)

We now state and prove one of the key theorems of matrix algebra.

Theorem 5.1 (Cayley-Hamilton) Let the characteristic polynomial of A be d(s). Then d(A) = O.

Proof Consider

$$(sI - A)^{-1} = \frac{1}{\det(sI - A)} \operatorname{adj}(sI - A) = \frac{1}{d(s)} B(s)$$
(5.8)

Since the elements of $B(s) = \operatorname{adj} (sI - A)$ are cofactors of the elements of sI - A, they are polynomials of degree not exceeding n - 1 (see Exercise 5.2). Thus we can write

 $B(s) = s^{n-1}B_1 + \dots + sB_{n-1} + B_n$

for some $B_i \in \mathbf{F}^{n \times n}$. Premultiplying both sides of (5.8) with d(s)(sI - A) we obtain

$$d(s)I = (sI - A)B(s)$$

or in open form as

$$s^{n}I + \sum_{i=1}^{n-1} s^{n-i} d_{i}I + d_{n}I = s^{n}B_{1} + \sum_{i=1}^{n-1} s^{n-i}(B_{i+1} - AB_{i}) - AB_{n}$$

Equating the coefficient matrices of the like powers of s, we get

$$B_{1} = I$$

$$B_{2} = AB_{1} + d_{1}I$$

$$\vdots$$

$$B_{n} = AB_{n-1} + d_{n-1}I$$

$$O = AB_{n} + d_{n}I$$

Substituting B_1 into the equation for B_2 , the resulting expression for B_2 into the equation for B_3 , and so on, we obtain

$$B_2 = A + d_1 I$$

$$B_3 = A^2 + d_1 A + d_2 I$$

$$\vdots$$

$$O = A^n + d_1 A^{n-1} + \dots + d_n I = d(A)$$

completing the proof.

Example 5.5

The characteristic polynomial of

$$A = \left[\begin{array}{rrrr} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{array} \right]$$

is

$$d(s) = s^{3} - 2s^{2} - s + 2 = (s - 1)^{2}(s - 2)$$

Forming

$$A - I = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$
$$(A - I)^2 = A - I$$
$$A - 2I = \begin{bmatrix} 0 & -1 & 1 \\ 1 & -2 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

it is easily verified that

$$d(A) = (A - I)^{2}(A - 2I) = O$$

Now consider the polynomial $\alpha(s) = (s-1)(s-2) = s^2 - 3s + 2$, which has a smaller degree than the characteristic polynomial. Since $(A - I)^2 = A - I$, it follows that we also have

$$\alpha(A) = (A - I)(A - 2I) = O$$

The least degree polynomial $\alpha(s)$ with a unity leading coefficient for which $\alpha(A) = O$ is called the **minimum polynomial** of A. Properties of the minimum polynomial are discussed in Exercise 5.9.

The Cayley-Hamilton theorem implies that the *n*th power of an $n \times n$ matrix can be expressed as a linear combination of smaller powers as

$$A^n = -d_1 A^{n-1} - \dots - d_n I$$

It then follows by induction that any power of A, and therefore, any polynomial p(A) can be written as a linear combination of powers up to n - 1. If p is any polynomial of degree higher than n - 1, then dividing p with the characteristic polynomial d, we obtain

$$p(s) = d(s)q(s) + r(s)$$

where q is the quotient polynomial and r is the remainder polynomial with degree not exceeding n-1. Then

$$p(A) = d(A)q(A) + r(A) = r(A)$$

where the last equality follows from the fact that d(A) = O. Thus p(A) can be evaluated by calculating r(A), which involves powers of A up to at most n - 1.

Example 5.6

The reflection matrix

$$A = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right]$$

in Example 5.3 has the property that $A^2 = I$. Hence $A^{2k} = I$, k = 1, 2, ..., that is, reflecting a vector an even number of times in the line y = x we end up with the original vector. Let us verify this observation for A^{100} using the Cayley-Hamilton theorem.

The characteristic polynomial of A has already been found as $d(s) = s^2 - 1$. Dividing $p(s) = s^{100}$ with d(s) we get

$$p(s) = s^{100} = (s^2 - 1)(s^{98} + s^{96} + \dots + s^2 + 1) + 1 = d(s)q(s) + r(s)$$

Hence r(s) = 1, and $p(A) = A^{100} = r(A) = I$.

Example 5.7

Let us evaluate A^q and find $\lim_{q \to \infty} A^q$, if it exists, for

$$A = \left[\begin{array}{cc} 2/3 & 1/2\\ 1/3 & 1/2 \end{array} \right]$$

The characteristic polynomial of A is obtained as

$$d(s) = s^{2} - \frac{7}{6}s + \frac{1}{6} = (s - \frac{1}{6})(s - 1)$$

Let

$$s^{q} = (s - \frac{1}{6})(s - 1)q(s) + r_{0}s + r_{1}, \quad q \ge 2$$

Evaluating both sides at the eigenvalues s = 1/6 and s = 1, we get

$$\frac{1}{6^q} = \frac{1}{6}r_0 + r_1 \\ 1 = r_0 + r_1$$

from which we obtain

$$r_{0} = \frac{6}{5} \left(1 - \frac{1}{6^{q}}\right)$$

$$r_{1} = \frac{6}{5} \left(\frac{1}{6^{q}} - \frac{1}{6}\right)$$

Thus

$$A^q = r_0 A + r_1 I$$

 $\quad \text{and} \quad$

$$\lim_{q \to \infty} A^q = (\lim_{q \to \infty} r_0) A + (\lim_{q \to \infty} r_1) I = \frac{6}{5} A - \frac{1}{5} I = \begin{bmatrix} 3/5 & 3/5\\ 2/5 & 2/5 \end{bmatrix}$$

5.3 The Diagonal Form

Recall from Section 4.3 that two matrices $A, A' \in \mathbf{F}^{n \times n}$ are said to be similar if they are related as

$$A' = P^{-1}AP$$

for some nonsingular matrix $P \in \mathbf{F}^{n \times n}$ which represents a change of basis in $\mathbf{F}^{n \times 1}$. Since similar matrices represent the same linear operator with respect to different bases, they are expected to share some common characteristics. From

$$det (sI - P^{-1}AP) = det [P^{-1}(sI - A)P]$$

= $(det P^{-1}) \cdot det (sI - A) \cdot (det P)$
= $det (sI - A)$

it follows that similar matrices have the same characteristic polynomial, and therefore, the same eigenvalues. It is then natural to look for a canonical form that displays the eigenstructure of similar matrices. Such a canonical form can conveniently represent the whole equivalence class of similar matrices.

Consider a diagonal matrix

$$D = \operatorname{diag}\left[d_1, \ldots, d_n\right] \in \mathbf{F}^{n \times n}$$

where **F** is either **R** or **C**. Clearly, the eigenvalues of D are $\lambda_i = d_i, i = 1, ..., n$, and D looks like a good candidate to characterize the class of matrices that have the same eigenvalues. As we proved above, all matrices that are similar to D have the same eigenvalues $\lambda_i = d_i$. The difficult problem is to determine whether a matrix having the eigenvalues $\lambda_i = d_i$ is similar to D.

Example 5.8

Let $D_1 = \text{diag}[1,2]$. We can construct infinitely many matrices that are similar to D. For example,

$$A_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}$$

and

$$A_{2} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -2 \\ 6 & -2 \end{bmatrix}$$

are both similar to D. In fact, all 2×2 matrices with eigenvalues $\lambda_1 = d_1 = 1, \lambda_2 = d_2 = 2$ are similar to D as we will prove shortly.

On the other hand, there is no other matrix that is similar to $D_2 = I_2$, simply because $P^{-1}IP = I$ for any nonsingular P. Although the matrix

$$A = \left[\begin{array}{rrr} 1 & 1 \\ 0 & 1 \end{array} \right]$$

has the same eigenvalues $\lambda_1 = \lambda_2 = 1$ as D_2 does, it is not similar to D_2 .

Example 5.8 shows that eigenvalues alone do not provide enough information to determine if a matrix is similar to a diagonal matrix. That information is in the eigenvectors.

Since

 $D\mathbf{e}_i = d_i \mathbf{e}_i$

for a diagonal matrix $D = \text{diag}[d_1, \ldots, d_n]$, it follows that $\mathbf{v}_i = \mathbf{e}_i$ is an eigenvector of D associated with the eigenvalue $\lambda_i = d_i$. Thus eigenvectors of a diagonal matrix are linearly independent in $\mathbf{F}^{n \times 1}$.

Suppose A is similar to D. Then $P^{-1}AP = D$, or equivalently, AP = PD for some nonsingular matrix P. Partitioning P into its columns as

$$P = [\mathbf{v}_1 \cdots \mathbf{v}_n] \tag{5.9}$$

we have

$$AP = [A\mathbf{v}_1 \cdots A\mathbf{v}_n] = PD = [d_1\mathbf{v}_1 \cdots d_n\mathbf{v}_n]$$

or columnwise

$$A\mathbf{v}_i = d_i \mathbf{v}_i, \quad i = 1, \dots, n$$

Thus each diagonal element d_i of D is an eigenvalue of A and the corresponding column \mathbf{v}_i of P is an eigenvector of A associated with it. Since P is nonsingular, \mathbf{v}_i are linearly independent.

Conversely, if A has the linearly independent eigenvectors \mathbf{v}_i associated with the eigenvalues $\lambda_i, i = 1, ..., n$, then the matrix P in (5.9) is nonsingular, and

$$AP = [A\mathbf{v}_1 \cdots A\mathbf{v}_n] = [\lambda_1\mathbf{v}_1 \cdots \lambda_n\mathbf{v}_n] = PD$$

Thus $P^{-1}AP = D$, where

 $D = \operatorname{diag}[\lambda_1, \ldots, \lambda_n]$

We summarize this important result as a theorem:

Theorem 5.2 A matrix $A \in \mathbf{F}^{n \times n}$ is similar to a diagonal matrix D if and only if it has n linearly independent eigenvectors in $\mathbf{F}^{n \times 1}$, in which case the diagonal elements of D are the eigenvalues of A.

The diagonal matrix D in Theorem 5.2 is called the **diagonal form**, and the matrix P consisting of the eigenvectors is called a **modal matrix** of A. D represents the equivalence class of matrices that are similar to A in the sense that every matrix that is similar to A has the same diagonal form D up to a reordering of the diagonal elements.

When an $n\times n$ matrix A has n linearly independent eigenvectors, the MATLAB command

[P,D]=eig(A)

returns a modal matrix and the diagonal form of A.

Example 5.9

The eigenvectors of the real matrix A in Example 5.1 can easily be shown to be linearly independent in $\mathbf{R}^{2\times 1}$. The modal matrix formed from the eigenvectors

$$P_A = \left[\begin{array}{rrr} 1 & 1 \\ 1 & -2 \end{array} \right]$$

is nonsingular. Calculating

$$P_A^{-1} = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & -1/3 \end{bmatrix}$$

the reader can verify that

$$P_A^{-1}AP_A = D = \left[\begin{array}{cc} 1 & 0\\ 0 & 4 \end{array} \right]$$

MATLAB returns a different modal matrix

$$P = \begin{bmatrix} -0.7071 & 0.4472\\ -0.7071 & -0.8944 \end{bmatrix}$$

but the same diagonal form D.

Consider a nonsingular matrix

$$Q = \begin{bmatrix} 1 & 2\\ 2 & 5 \end{bmatrix} \quad \text{with} \quad Q^{-1} = \begin{bmatrix} 5 & -2\\ -2 & 1 \end{bmatrix}$$

and let

$$B = Q^{-1}AQ = \begin{bmatrix} -8 & -27\\ 4 & 13 \end{bmatrix}$$

Since B is similar to A, it must have the same diagonal form. Indeed, forming the characteristic polynomial of B as

$$\det (sI - B) = \det \begin{bmatrix} s+8 & 27\\ -4 & s-13 \end{bmatrix} = (s+8)(s-13) + 108 = s^2 - 5s + 4$$

we observe that the eigenvalues of B are the same as those of A: $\lambda_1 = 1$ and $\lambda_2 = 4$. The reader can verify that columns of

$$P_B = \left[\begin{array}{cc} 3 & -9 \\ -1 & 4 \end{array} \right]$$

are eigenvectors of B so that P_B is a modal matrix of B. The diagonal form of B

$$P_B^{-1}BP_B = D = \left[\begin{array}{cc} 1 & 0\\ 0 & 4 \end{array} \right]$$

is the same as the diagonal form of A as expected.

$$C = \left[\begin{array}{cc} 0 & -1 \\ 4 & 5 \end{array} \right]$$

which also has the same eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 4$. The reader can verify that

$$P_C = \left[\begin{array}{rrr} 1 & -1 \\ -1 & 4 \end{array} \right]$$

is a modal matrix of C, and that

$$P_C^{-1}CP_C = D = \left[\begin{array}{cc} 1 & 0\\ 0 & 4 \end{array} \right]$$

Since C has the same diagonal form, it must be similar to A. Indeed,

$$C = P_C D P_C^{-1} = P_C P_A^{-1} A P_A P_C^{-1} = P^{-1} A P$$

where

$$P = P_A P_C^{-1} = \begin{bmatrix} 5/3 & 2/3 \\ 1/3 & -1/3 \end{bmatrix}$$

Example 5.10

The characteristic polynomial of the complex matrix

$$A = \left[\begin{array}{cc} 2+i & -1-i \\ 2 & -1 \end{array} \right]$$

is

$$d(s) = s^{2} - (1+i)s + i = (s-1)(s-i)$$

The eigenvectors associated with $\lambda_1 = 1$ and $\lambda_2 = i$ can be found as

$$\mathbf{v}_1 = \left[\begin{array}{c} 1\\1 \end{array} \right], \quad \mathbf{v}_2 = \left[\begin{array}{c} 1+i\\2 \end{array} \right]$$

 \mathbf{v}_1 and \mathbf{v}_2 are linearly independent in $\mathbf{C}^{2 \times 1}$. Constructing a modal matrix

$$P = \left[\begin{array}{rrr} 1 & 1+i \\ 1 & 2 \end{array} \right]$$

from the eigenvectors and computing P^{-1} , we obtain the diagonal form of A as

$$P^{-1}AP = D = \begin{bmatrix} 1 & 0\\ 0 & i \end{bmatrix}$$

Example 5.11

The real matrix A in Example 5.2 has no real eigenvectors, and therefore, it can not be diagonalized by a change of basis in $\mathbb{R}^{2\times 1}$. However, it has two linearly independent complex eigenvectors as given in Example 5.2. Treating A as a complex matrix, we can construct a modal matrix

$$P = \left[\begin{array}{cc} 1 & 1\\ 1+2i & 1-2i \end{array} \right]$$

from the eigenvectors. The transformation

$$P^{-1}AP = \begin{bmatrix} 1+2i & 0\\ 0 & 1-2i \end{bmatrix} = D$$

results in a diagonal matrix that is similar to A in $\mathbb{C}^{2\times 1}$. Note that although A is real, its diagonal form D is complex.

From Example 5.11 we observe that some real matrices that can not be diagonalized when considered as a linear operator on $\mathbf{R}^{n\times 1}$ may be diagonalized when considered as a linear operator on $\mathbf{C}^{n\times 1}$. For this reason, we will first concentrate on complex matrices, and treat all matrices as linear operators on $\mathbf{C}^{n\times 1}$ even if they are real.

5.3.1 Complex Diagonal Form

Let $A \in \mathbf{C}^{n \times n}$ have k distinct eigenvalues $\lambda_i \in \mathbf{C}$ with multiplicities n_i . Consider the eigenspaces $\mathbf{K}_i \subset \mathbf{C}^{n \times 1}$ associated with λ_i . We claim that these subspaces are linearly independent. To prove the claim, pick arbitrary vectors $\mathbf{v}_i \in \mathbf{K}_i$ and set

 $\mathbf{v}_1 + \dots + \mathbf{v}_k = \mathbf{0} \tag{5.10}$

Let us define the polynomials

$$p_q(s) = \prod_{\substack{j=1\\j \neq q}}^k (s - \lambda_j), \quad q = 1, \dots, k$$

Since all factors $(s - \lambda_i)$ except $(s - \lambda_q)$ are included in $p_q(s)$, $p_q(\lambda_i) = 0$ for $i \neq q$ but $p_q(\lambda_q) \neq 0$. Consider the products

$$p_q(A)\mathbf{v}_i = (A - \lambda_1 I) \cdots (A - \lambda_{q-1} I)(A - \lambda_{q+1} I) \cdots (A - \lambda_k I)\mathbf{v}_i$$

Since the matrices $(A - \lambda_i I)$ and $(A - \lambda_j I)$ commute for all (i, j) and

$$(A - \lambda_j I)\mathbf{v}_i = (\lambda_i - \lambda_j)\mathbf{v}_i$$

we have

$$p_q(A)\mathbf{v}_i = p_q(\lambda_i)\mathbf{v}_i = \begin{cases} p_q(\lambda_q)\mathbf{v}_q, & i = q \\ \mathbf{0}, & i \neq q \end{cases}$$

Hence premultiplying both sides of (5.10) with $p_a(A)$ we obtain

$$p_q(\lambda_q)\mathbf{v}_q = \mathbf{0}, \quad q = 1, \dots, k$$

Since $p_q(\lambda_q) \neq 0$ the last equality implies

$$\mathbf{v}_q = \mathbf{0}, \quad q = 1, \dots, k$$

proving the claim.

An immediate consequence of this result is that if A has n distinct eigenvalues (in which case they are simple zeros of the characteristic polynomial, that is, $n_i = 1, ..., n$) then it has n linearly independent eigenvectors, and hence it can be diagonalized by a similarity transformation. This has already been illustrated in Examples 5.9-5.11. Since almost all square matrices with independent elements have simple eigenvalues, in practice almost all matrices can be diagonalized.¹ However, some problems may lead to matrices whose elements are so interdependent that they

¹The reader may try to verify this statement by generating several random matrices by MATLAB and computing their eigenvalues.

have repeated eigenvalues. For completeness of the theory, in the rest of this section and in the following section we will study such matrices.

We first answer the question of under what conditions matrices with repeated eigenvalues can be diagonalized:

Corollary 5.2.1 Let $A \in \mathbf{C}^{n \times n}$ have the characteristic polynomial in (5.7), where $\lambda_i \neq \lambda_i$ for $i \neq j$. Then the following are equivalent.

- a) A is similar (in $\mathbf{C}^{n \times 1}$) to a diagonal matrix D.
- b) $\nu_i = \dim(\mathbf{K}_i) = n_i, \quad i = 1, ..., k.$
- c) $\mathbf{C}^{n \times 1} = \mathbf{K}_1 \oplus \cdots \oplus \mathbf{K}_k$.

Proof We will show that (a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a).

(a) \Rightarrow (b):

Let $P^{-1}AP = D$ for some diagonal matrix D. Since D has the same characteristic polynomial as A, we can assume without loss of generality that

$$D = \begin{bmatrix} \lambda_1 I_{n_1} & & \\ & \ddots & \\ & & \lambda_k I_{n_k} \end{bmatrix}$$
(5.11)

Let P be partitioned into its columns as

$$P = [P_1 \cdots P_k] \tag{5.12}$$

where P_i is an $n \times n_i$ block of P corresponding to the *i*th diagonal block of D. From

 $AP = [AP_1 \cdots AP_k] = PD = [\lambda_1 P_1 \cdots \lambda_k P_k]$

we observe that each column of P_i is an eigenvector of A associated with the eigenvalue $\lambda_i, i = 1, \ldots, k$. Hence $\mathbf{K}_i \supset \operatorname{cs}(P_i)$, which implies that

 $\dim \left(\mathbf{K}_{i} \right) \geq n_{i} \,, \quad i = 1, \dots, k$

However, since \mathbf{K}_i are linearly independent and $\bigoplus \mathbf{K}_i \subset \mathbf{C}^{n \times 1}$, we have

$$\sum n_i \leq \sum \dim (\mathbf{K}_i) = \dim (\bigoplus \mathbf{K}_i) \leq n$$

Then the only possibility is that $\dim(\mathbf{K}_i) = n_i, i = 1, \dots, k$.

(b) \Rightarrow (c):

This follows directly from the facts that \mathbf{K}_i are linearly independent and that

$$\sum \dim \left(\mathbf{K}_{i} \right) = \sum n_{i} = n$$

(c) \Rightarrow (a):

Let the columns of $n \times n_i$ matrices

$$P_i = [\mathbf{v}_{i1} \cdots \mathbf{v}_{in_i}]$$

form bases for \mathbf{K}_i , i = 1, ..., k, and let P be constructed from P_i as in (5.12). Then columns of P form a basis for $\mathbf{C}^{n \times 1}$, and therefore, P is nonsingular. Also,

$$A\mathbf{v}_{ij} = \lambda_i \mathbf{v}_{ij} \implies AP_i = \lambda_i P_i \implies AP = PD$$

where D is as given by (5.11), and the result follows.

Although proved indirectly, the key result of Corollary 5.2.1 is the implication (b) \Rightarrow (a), which simply states that A can be diagonalized if the geometric multiplicity of each eigenvalue equals its algebraic multiplicity, in which case there exist sufficiently many linearly independent eigenvectors associated with each multiple eigenvalue. Thus Corollary 5.2.1 provides a necessary and sufficient condition for diagonalizability of a matrix in terms of the dimensions of its eigenspaces. We illustrate this with the following examples.

Example 5.12

The matrix

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

has the characteristic polynomial

$$d(s) = s^{4} - 2s^{3} + 2s^{2} - 2s + 1 = (s - 1)^{2}(s^{2} + 1)$$

Hence $\lambda_1 = 1$ with $n_1 = 2$, $\lambda_2 = i$ with $n_2 = 1$, and $\lambda_3 = -i$ with $n_3 = 1$. From

$$(A - \lambda_1 I) = \begin{bmatrix} 0 & 2 & -2 & -2 \\ 0 & -2 & 2 & 2 \\ 0 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix} \xrightarrow{\text{e.r.o.}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we find that $\dim(\mathbf{K}_1) = 2 = n_1$. Also, since \mathbf{K}_2 and \mathbf{K}_3 each contain at least one eigenvector, $\dim(\mathbf{K}_2) = 1 = n_2$ and $\dim(\mathbf{K}_3) = 1 = n_3$. Then by Corollary 5.2.1 A is similar to a diagonal matrix.

To construct a modal matrix for A, we need to find four linearly independent eigenvectors; two associated with λ_1 , and one associated with each of λ_2 and λ_3 . Eigenvectors associated with λ_1 can be found from the reduced row echelon form of $A - \lambda_1 I$ above as

$$\mathbf{v}_{11} = \begin{bmatrix} 0\\1\\1\\0 \end{bmatrix}, \quad \mathbf{v}_{12} = \begin{bmatrix} 0\\1\\0\\1 \end{bmatrix}$$

An eigenvector associated with λ_2 can be found from

$$(A - \lambda_2 I) = \begin{bmatrix} 1-i & 2 & -2 & -2 \\ 0 & -1-i & 2 & 2 \\ 0 & -1 & 2-i & 1 \\ 1 & 1 & -1 & -i \end{bmatrix} \stackrel{\text{e.r.o.}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 & -2i \\ 0 & 1 & 0 & 2i \\ 0 & 0 & 1 & i \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

as

$$\mathbf{v}_2 = \begin{bmatrix} -2\\2\\1\\i \end{bmatrix}$$

Finally, since $\lambda_3 = \lambda_2^*$, we choose $\mathbf{v}_3 = \mathbf{v}_2^*$.

Thus

$$P = \begin{bmatrix} 0 & 0 & -2 & -2 \\ 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & i & -i \end{bmatrix}, \quad P^{-1} = \frac{1}{4} \begin{bmatrix} 2 & 0 & 4 & 0 \\ 2 & 4 & -4 & 0 \\ -1+i & 2i & -2i & -2i \\ -1-i & -2i & 2i & 2i \end{bmatrix}$$

and

$$P^{-1}AP = \begin{bmatrix} 1 & & \\ & 1 & \\ & & i \\ & & -i \end{bmatrix}$$

Example 5.13

The matrix

$$A = \begin{bmatrix} -5 & 3 & 0 & 3 \\ -3 & 1 & 3 & 3 \\ 3 & -3 & -2 & 0 \\ -3 & 0 & -3 & -2 \end{bmatrix}$$

has the characteristic polynomial

$$d(s) = s^{4} + 8s^{3} + 42s^{2} + 104s + 169 = (s^{2} + 4s + 13)^{2}$$

and hence the complex conjugate eigenvalues

 $\lambda_{1,2} = -2 \mp 3i \,, \quad n_{1,2} = 2$

Obtaining the reduced row echelon form of $A-\lambda_1 I$ as

$$A - \lambda_1 I = \begin{bmatrix} -3 - 3i & 3 & 0 & 3\\ -3 & 3 - 3i & 3 & 3\\ 3 & -3 & -3i & 0\\ -3 & 0 & -3 & -3i \end{bmatrix} \stackrel{\text{e.r.o.}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 & i\\ 0 & 1 & 1 + i & i\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we find two linearly independent eigenvectors associated with λ_1 :

$$\mathbf{v}_{11} = \begin{bmatrix} 1\\1+i\\-1\\0 \end{bmatrix}, \quad \mathbf{v}_{12} = \begin{bmatrix} i\\i\\0\\-1 \end{bmatrix}$$

Since $\lambda_2 = \lambda_1^*$ we choose eigenvectors associated with λ_2 to be the complex conjugates of \mathbf{v}_{11} and \mathbf{v}_{12} :

$$\mathbf{v}_{21} = \mathbf{v}_{11}^* = \begin{bmatrix} 1\\ 1-i\\ -1\\ 0 \end{bmatrix}, \quad \mathbf{v}_{22} = \mathbf{v}_{12}^* = \begin{bmatrix} -i\\ -i\\ 0\\ -1 \end{bmatrix}$$

Thus

$$P = \begin{bmatrix} 1 & i & 1 & -i \\ 1+i & i & 1-i & -i \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix}, \quad P^{-1} = \frac{1}{2} \begin{bmatrix} i & -i & -1 & 0 \\ -i & 0 & -i & -1 \\ -i & i & -1 & 0 \\ i & 0 & i & -1 \end{bmatrix}$$

and

$$P^{-1}AP = D = \begin{bmatrix} -2+3i & & \\ & -2+3i & \\ & & -2-3i \\ & & & -2-3i \end{bmatrix}$$

* 5.3.2 Invariant Subspaces

Let us take a closer look at the role the eigenspaces \mathbf{K}_i play in diagonalization of a complex matrix: If $\mathbf{v} \in \mathbf{K}_i$ then \mathbf{v} is an eigenvector of A associated with λ_i , and

$$A\mathbf{v} = \lambda_i \mathbf{v} \in \mathbf{K}_i$$

In other words, the eigenspaces \mathbf{K}_i have the property that

$$\mathbf{v} \in \mathbf{K}_i \implies A\mathbf{v} \in \mathbf{K}_i$$

A subspace $\mathbf{V} \subset \mathbf{C}^{n \times 1}$ is said to be **invariant** under A, or A-invariant, if

 $A\mathbf{v} \in \mathbf{V}$ for all $\mathbf{v} \in \mathbf{V}$

Thus each eigenspace \mathbf{K}_i is an A-invariant subspace.

Suppose that we have a direct sum decomposition

$$\mathbf{C}^{n\times 1} = \mathbf{V}_1 \oplus \dots \oplus \mathbf{V}_{\kappa} \tag{5.13}$$

where \mathbf{V}_i are A-invariant subspaces with dim $(\mathbf{V}_i) = \eta_i$, so that $\sum \eta_i = n$. Let columns of the $n \times \eta_i$ matrices

 $T_i = [\mathbf{t}_{i1} \cdots \mathbf{t}_{i\eta_i}]$

form bases for \mathbf{V}_i . Then their union, that is, the columns of the $n \times n$ matrix

$$T = [T_1 \cdots T_{\kappa}]$$

form a basis for $\mathbf{C}^{n \times 1}$, and hence T is nonsingular.

Since $\mathbf{t}_{ij} \in \mathbf{V}_i$ and \mathbf{V}_i is A-invariant, $A\mathbf{t}_{ij} \in \mathbf{V}_i$ and hence $A\mathbf{t}_{ij}$ can be written as a linear combination of the columns of T_i . That is,

 $A\mathbf{t}_{ij} = T_i \mathbf{f}_{ij}, \quad i = 1, \dots, \kappa; \ j = 1, \dots, \eta_i$

for some $\eta_i \times 1$ column vector \mathbf{f}_{ij} . Then

$$AT_i = [A\mathbf{t}_{i1} \cdots A\mathbf{t}_{i\eta_i}] = [T_i\mathbf{f}_{i1} \cdots T_i\mathbf{f}_{i\eta_i}] = T_i[\mathbf{f}_{i1} \cdots \mathbf{f}_{i\eta_i}] = T_iF_i$$

where F_i are $\eta_i \times \eta_i$ matrices, and

$$AT = [AT_1 \cdots AT_{\kappa}] = [T_1F_1 \cdots T_{\kappa}F_{\kappa}] = TF$$

where

$$F = \left[\begin{array}{ccc} F_1 & \cdots & O\\ \vdots & \ddots & \vdots\\ O & \cdots & F_{\kappa} \end{array} \right]$$

Since T is nonsingular we have

$$T^{-1}AT = F = \operatorname{diag}\left[F_1 \cdots F_\kappa\right] \tag{5.14}$$

Thus a direct sum decomposition of $\mathbf{C}^{n \times 1}$ into A-invariant subspaces as in (5.13) results in a block diagonal form of A as in (5.14). The diagonal form in Corollary 5.2.1 is a special case of this result, where $\kappa = k$, $\mathbf{V}_i = \mathbf{K}_i$, $\eta_i = n_i$ and $T_i = P_i$, $i = 1, \ldots, k$. However, there is more in Corollary 5.2.1: Since the columns of

$$P_i = [\mathbf{v}_{i1} \cdots \mathbf{v}_{in_i}]$$

are eigenvectors that satisfy $A\mathbf{v}_{ij} = \lambda_i \mathbf{v}_{ij}, j = 1, \dots, n_i$, they induce a further decomposition of \mathbf{K}_i as

$$\mathbf{K}_i = \mathbf{V}_{i1} \oplus \cdots \oplus \mathbf{V}_{in_i}$$

where each one-dimensional subspace $\mathbf{V}_{ij} = \text{span}(\mathbf{v}_{ij})$ is also A-invariant. That is why we have $F_i = \lambda_i I_{n_i}$, i = 1, ..., k and F = D. In other words, condition (c) of Corollary 5.2.1 provides the finest possible direct sum decomposition

$$\mathbf{C}^{n\times 1} = \bigoplus_{i=1}^{k} \bigoplus_{j=1}^{n_i} \mathbf{V}_{ij}$$

into one-dimensional A-invariant subspaces, which leads to the simplest possible form (the diagonal form) of A.

* 5.3.3 Real Semi-Diagonal Form

We now consider the problem of transforming a real matrix A into a simple form by a suitable choice of a basis for $\mathbf{R}^{n \times 1}$.

We have no difficulty if A has only real eigenvalues: We can choose its eigenvectors to be real and restrict the eigenspaces \mathbf{K}_i to be subspaces of $\mathbf{R}^{n\times 1}$ rather than $\mathbf{C}^{n\times 1}$. Then Corollary 5.2.1 remains valid even with $\mathbf{C}^{n\times 1}$ replaced with $\mathbf{R}^{n\times 1}$, and we can diagonalize A provided the conditions on the dimension of the eigenspaces are satisfied. The difficulty arises if A has one or more complex eigenvalues, in which case the corresponding eigenspaces are no longer subspaces of $\mathbf{R}^{n\times 1}$, and Corollary 5.2.1 becomes void. However, the argument in the preceeding subsection suggests that it may still be possible to decompose $\mathbf{R}^{n\times 1}$ into a direct sum of A-invariant subspaces and choose suitable bases for these subspaces to come up with a simple representation of A.

Example 5.14

Consider the real matrix

$$A = \left[\begin{array}{cc} 0 & 1 \\ -5 & 2 \end{array} \right]$$

in Example 5.2, which has a pair of complex conjugate eigenvalues

 $\lambda_{1,2} = 1 \mp 2i$

Clearly, $\mathbf{R}^{2\times 1}$ cannot be decomposed into smaller A-invariant subspaces, because then A would be similar to a real diagonal matrix whose diagonal elements would have to be the eigenvalues of A. However, we may try to construct a basis for $\mathbf{R}^{2\times 1}$ and transform A into a special form. For this purpose, let us separate the complex conjugate eigenvectors of A into their real and imaginary parts as

$$\mathbf{v}_{1,2} = \begin{bmatrix} 1\\ 1\mp 2i \end{bmatrix} = \begin{bmatrix} 1\\ 1 \end{bmatrix} \mp i \begin{bmatrix} 0\\ 2 \end{bmatrix} = \mathbf{u} \mp i\mathbf{w}$$

Since \mathbf{v}_1 and \mathbf{v}_2 are linearly independent in $\mathbf{C}^{2\times 1}$, \mathbf{u} and \mathbf{w} are linearly independent in $\mathbf{R}^{2\times 1}$. (This statement can be proved by following the same argument in Example 3.15.) Thus

$$P_R = \begin{bmatrix} \mathbf{u} & \mathbf{w} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$$

is nonsingular. The representation of A with respect to the basis (\mathbf{u}, \mathbf{w}) is obtained as

$$P_R^{-1}AP_R = D_R = \begin{bmatrix} 1 & 2\\ -2 & 1 \end{bmatrix}$$

Although D_R is no simpler than A, it has a special structure: Its diagonal elements are the real parts and the off-diagonal elements are the imaginary parts of the complex eigenvalues. Of course, this is not just a coincidence as we explain next.

Suppose that A has m distinct pairs of complex conjugate eigenvalues

$$\lambda_{1} = \sigma_{1} + i\omega_{1}, \quad \lambda_{2} = \sigma_{1} - i\omega_{1}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\lambda_{2m-1} = \sigma_{m} + i\omega_{m}, \quad \lambda_{2m} = \sigma_{m} - i\omega_{m}$$
(5.15)

with associated complex conjugate pairs of eigenvectors

$$\mathbf{v}_{1} = \mathbf{u}_{1} + i\mathbf{w}_{1}, \quad \mathbf{v}_{2} = \mathbf{u}_{1} - i\mathbf{w}_{1}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\mathbf{v}_{2m-1} = \mathbf{u}_{m} + i\mathbf{w}_{m}, \quad \mathbf{v}_{2m} = \mathbf{u}_{m} - i\mathbf{w}_{m}$$
(5.16)

and n - 2m distinct real eigenvalues

$$\lambda_{2m+1},\ldots,\lambda_n$$

with associated real eigenvectors

 $\mathbf{v}_{2m+1},\ldots,\mathbf{v}_m$

Equating the real and imaginary parts of

$$A(\mathbf{u}_i + i\mathbf{w}_i) = (\sigma_i + i\omega_i)(\mathbf{u}_i + i\mathbf{w}_i), \quad i = 1, \dots, m$$

we obtain

$$A\begin{bmatrix}\mathbf{u}_i & \mathbf{w}_i\end{bmatrix} = \begin{bmatrix}\mathbf{u}_i & \mathbf{w}_i\end{bmatrix} \begin{bmatrix} \sigma_i & \omega_i \\ -\omega_i & \sigma_i \end{bmatrix}, \quad i = 1, \dots, m$$

or in compact form

$$AP_{Ri} = P_{Ri}D_{Ri}, \quad i = 1, \dots, m$$
 (5.17)

where

$$P_{Ri} = \begin{bmatrix} \mathbf{u}_i & \mathbf{w}_i \end{bmatrix}, \quad D_{Ri} = \begin{bmatrix} \sigma_i & \omega_i \\ -\omega_i & \sigma_i \end{bmatrix}$$

Note that the diagonal elements of D_{Ri} are the real parts and the off-diagonal elements are the imaginary parts of the corresponding conjugate pair of complex eigenvalues (as we already observed in Example 5.14). Thus with

$$P_R = [P_{R1} \cdots P_{Rm} \mathbf{v}_{2m+1} \cdots \mathbf{v}_n]$$
(5.18)

we have

$$AP_R = P_R D_R$$

where

$$D_{R} = \begin{bmatrix} D_{R1} & & & & \\ & \ddots & & & & \\ & & D_{Rm} & & & \\ & & & \lambda_{2m+1} & & \\ & & & & \ddots & \\ & & & & & & \lambda_{n} \end{bmatrix}$$
(5.19)

It can be shown that P_R is also nonsingular (see Exercise 5.29), so that

$$P_R^{-1}AP_R = D_R$$

The matrix P_R is called a real modal matrix of A, and D_R the real semi-diagonal form of A.

The structure of P_R in (5.18) implies a decomposition of $\mathbf{R}^{n \times 1}$ as

$$\mathbf{R}^{n\times 1} = \mathbf{V}_1 \oplus \dots \oplus \mathbf{V}_m \oplus \mathbf{V}_{2m+1} \oplus \dots \mathbf{V}_n \tag{5.20}$$

where

$$\mathbf{V}_{i} = \begin{cases} \text{span}(\mathbf{u}_{i}, \mathbf{w}_{i}), & i = 1, \dots, m \\ \text{span}(\mathbf{v}_{i}), & i = 2m + 1, \dots, n \end{cases}$$

(5.17) guarantees that each \mathbf{V}_i , i = 1, ..., m, is a two-dimensional A-invariant subspace of $\mathbf{R}^{n \times 1}$ associated with a pair of complex conjugate eigenvalues. Since the one-dimensional subspaces \mathbf{V}_i , i = 2m + 1, ..., n, associated with real eigenvalues are also A-invariant, this decomposition yields a block diagonal representation of A as in (5.19).

Example 5.15

The matrix

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

has the characteristic polynomial

$$d(s) = s^{3} - s^{2} + 2 = (s+1)(s^{2} - 2s + 2)$$

The eigenvalues are $\lambda_1 = -1$, $\lambda_{2,3} = 1 \mp i$, and the associated eigenvectors can be found as

$$\mathbf{v}_1 = \begin{bmatrix} 1\\ -1\\ 1 \end{bmatrix}, \quad \mathbf{v}_{2,3} = \mathbf{u}_2 \mp i\mathbf{w}_2 = \begin{bmatrix} 1\\ 1\mp i\\ \mp 2i \end{bmatrix}$$

A real modal matrix is

$$P_{R} = \begin{bmatrix} \mathbf{v}_{1} & \mathbf{u}_{2} & \mathbf{w}_{2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$$

which results in the real semidiagonal form of A:

$$P_R^{-1}AP_R = D_R = \begin{bmatrix} -1 & 0 & 0\\ 0 & 1 & 1\\ 0 & -1 & 1 \end{bmatrix}$$

If a real matrix $A \in \mathbf{R}^{n \times n}$ has multiple complex eigenvalues, then as long as $\dim(\mathbf{K}_i) = n_i$ for all the distinct eigenvalues, a real modal matrix can be constructed from the real and imaginary parts of the complex eigenvectors, and A can be semidiagonalized by a similarity transformation in $\mathbf{R}^{n \times 1}$.

Example 5.16

The real matrix

$$A = \begin{bmatrix} -5 & 3 & 0 & 3\\ -3 & 1 & 3 & 3\\ 3 & -3 & -2 & 0\\ -3 & 0 & -3 & -2 \end{bmatrix}$$

has the characteristic polynomial

$$d(s) = s^{4} + 8s^{3} + 42s^{2} + 104s + 169 = (s^{2} + 4s + 13)^{2}$$

and hence the complex conjugate eigenvalues

 $\lambda_{1,2} = -2 \mp 3i \,, \quad n_{1,2} = 2$

Obtaining the reduced row echelon form of $A - \lambda_1 I$ as

$$A = \begin{bmatrix} -3 - 3i & 3 & 0 & 3\\ -3 & 3 - 3i & 3 & 3\\ 3 & -3 & -3i & 0\\ -3 & 0 & -3 & -3i \end{bmatrix} \stackrel{\text{e.r.o}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 & i\\ 0 & 1 & 1 + i & i\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we find two linearly independent eigenvectors associated with λ_1 :

$$\mathbf{v}_1 = \mathbf{u}_1 + i\mathbf{w}_1 = \begin{bmatrix} 1\\1+i\\-1\\0 \end{bmatrix}, \quad \mathbf{v}_2 = \mathbf{u}_2 + i\mathbf{w}_2 = \begin{bmatrix} i\\i\\0\\-1 \end{bmatrix}$$

Constructing a real modal matrix

$$P_{R} = [P_{R1} \ P_{R2}] = [\mathbf{u}_{1} \ \mathbf{w}_{1} | \mathbf{u}_{2} \ \mathbf{w}_{2}] = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

from the real and imaginary parts of \mathbf{v}_1 and \mathbf{v}_2 , we obtain the real semi-diagonal form of A as

$$D_R = P_R^{-1} A P_R = \begin{bmatrix} -2 & 3 & 0 & 0 \\ -3 & -2 & 0 & 0 \\ 0 & 0 & -2 & 3 \\ 0 & 0 & -3 & -2 \end{bmatrix}$$

* 5.4 The Jordan Form

5.4.1 The Complex Jordan Form

Corollary 5.2.1 implies that if A has an eigenvalue λ_i with $n_i > 1$ for which $\nu_i < n_i$, then it can not be diagonalized. However, by a careful choice of the basis vectors of $\mathbf{C}^{n \times 1}$, it can still be transformed into a simple form as stated by the following theorem. The proof of the theorem is beyond the scope of this book, and is omitted.

Theorem 5.3 (The Jordan Form) Let A have the characteristic polynomial in (5.7), where $\lambda_i \neq \lambda_j$ for $i \neq j$, and let $\dim(\mathbf{K}_i) = \nu_i$, i = 1, ..., k. Then there exists a nonsingular matrix P such that

$$P^{-1}AP = J = diag [J_1, \dots, J_k]$$
(5.21)

where

a) each block J_i is an $n_i \times n_i$ matrix that consists of ν_i subblocks

$$J_i = diag \left[J_{i1}, \ldots, J_{i\nu_i} \right], \quad i = 1, \ldots, k$$

b) each subblock J_{ij} is an $n_{ij} \times n_{ij}$ matrix, and

$$J_{ij} = \lambda_i, \quad if \ n_{ij} = 1$$

$$J_{ij} = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}, \quad if \ n_{ij} > 1$$

$$c) \ n_{i1} \ge \cdots \ge n_{i\nu_i} \ and \ \sum_{j=1}^{\nu_i} n_{ij} = n_i, \quad i = 1, \dots, k$$

J is called the Jordan form of A, and is unique up to a reordering of the Jordan blocks J_i and the Jordan subblocks J_{ij} . P is called a modal matrix of A.

The general structure of the Jordan form is illustrated below for a typical case of

 $n_{1} = \nu_{1} = n_{11} = 1$ $n_{2} = \nu_{2} = 2, \quad n_{21} = n_{22} = 1$ $n_{3} = 2, \quad \nu_{3} = 1, \quad n_{31} = 2$ $n_{4} = 3, \quad \nu_{4} = 2, \quad n_{41} = 2, n_{42} = 1$ $\begin{bmatrix} \lambda_{1} \\ & & \\ & \lambda_{2} \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & &$

Note that the algebraic multiplicity n_i of an eigenvalue determines the size of the associated Jordan block J_i , and the geometric multiplicity ν_i determines the number of the subblocks J_{ij} in J_i . If $\nu_i = n_i$ for an eigenvalue λ_i (as for λ_1 and λ_2 above), then the corresponding Jordan block J_i consists of n_i subblocks each of which is a scalar; that is, $J_{ij} = \lambda_i, j = 1, \ldots, n_i$, and $J_i = \lambda_i I_{n_i}$. Thus if $\nu_i = n_i$ for all eigenvalues, then the Jordan form reduces to a diagonal matrix. In other words, the diagonal form of A (if it exists) is a special Jordan form.

We now provide an interpretation of the Jordan form in terms of a direct sum decomposition of $\mathbf{C}^{n\times 1}$. For this purpose let us partition P as

$$P = [P_1 \cdots P_k]$$

where P_i are $n \times n_i$ matrices that are associated with the Jordan blocks P_i . Similarly, if $\nu_i > 1$ then P_i can further be partitioned as

$$P_i = [P_{i1} \cdots P_{i\nu_i}]$$

where P_{ij} are $n \times n_{ij}$ matrices that are associated with the Jordan subblocks J_{ij} . Partitioning of P corresponds to a direct sum decomposition of $\mathbf{C}^{n \times 1}$ as

$$\mathbf{C}^{n\times 1} = \mathbf{V}_1 \oplus \dots \oplus \mathbf{V}_k \tag{5.22}$$

where

$$\mathbf{V}_i = \operatorname{cs}\left(P_i\right), \quad i = 1, \dots, k$$

Similarly, partitioning of P_i corresponds to a direct sum decomposition of \mathbf{V}_i as

$$\mathbf{V}_i = \mathbf{V}_{i1} \oplus \dots \oplus \mathbf{V}_{i\nu_i}, \quad i = 1, \dots, k$$
(5.23)

where

$$\mathbf{V}_{ij} = \operatorname{cs}(P_{ij}), \quad i = 1, \dots, k; \, j = 1, \dots, \nu_i$$

Combining (5.22) and (5.23) we obtain

$$\mathbf{C}^{n\times 1} = \bigoplus_{i=1}^{k} \bigoplus_{j=1}^{\nu_{i}} \mathbf{V}_{ij}$$
(5.24)

Thus the Jordan form is related to a direct sum decomposition of $\mathbf{C}^{n \times 1}$.

Rewriting (5.21) as AP = PJ and performing block multiplication, we get

$$AP_i = P_i J_i , \quad i = 1, \dots, k \tag{5.25}$$

If $\mathbf{v} \in \mathbf{V}_i$ then $\mathbf{v} = P_i \boldsymbol{\alpha}_i$ for some $\boldsymbol{\alpha}_i \in \mathbf{C}^{n_i \times 1}$, and by (5.25) we have

$$A\mathbf{v} = AP_i\boldsymbol{\alpha}_i = P_i(J_i\boldsymbol{\alpha}_i) \in \mathbf{V}_i$$

This shows that each \mathbf{V}_i in (5.22) is an A-invariant subspace. Similarly, from (5.25) we get

$$AP_{ij} = P_{ij}J_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, \nu_i$$
(5.26)

which shows that each \mathbf{V}_{ij} in (5.23) is also an A-invariant subspace. Consequently, (5.24) gives a direct sum decomposition of $\mathbf{C}^{n\times 1}$ into A-invariant subspaces. In fact, it is the finest decomposition of $\mathbf{C}^{n\times 1}$ in the sense that no \mathbf{V}_{ij} can further be decomposed into smaller A-invariant subspaces. In this sense, the Jordan form of A is the simplest matrix that is similar to A. Like the diagonal form of a diagonalizable matrix, it represents the equivalence class of matrices that are similar to A.

Next we investigate the relation of the subspaces \mathbf{V}_{ij} in (5.24) to the eigenvalues of A.

We first consider the special case where $\nu_i = n_i$ for an eigenvalue λ_i . In this case $J_i = \lambda_i I_{n_i}$ as mentioned before, and therefore, $AP_i = P_i J_i = \lambda_i P_i$. That is,

$$P_i = [P_{i1} \cdots P_{in_i}] = [\mathbf{v}_{i1} \cdots \mathbf{v}_{in_i}]$$

where each \mathbf{v}_{ij} is an eigenvector associated with λ_i . In terms of subspaces, we have

$$\mathbf{V}_i = \mathbf{K}_i = \bigoplus_{j=1}^{n_i} \mathbf{V}_{ij}$$

where

$$\mathbf{V}_{ij} = \operatorname{span}\left(\mathbf{v}_{ij}\right)$$

are one-dimensional A-invariant subspaces. In particular, if $\nu_i = n_i = 1$ then the corresponding Jordan block J_i contains only a single subblock $J_i = J_{i1} = \lambda_i$. In this case $\mathbf{V}_i = \mathbf{K}_i = \operatorname{span}(\mathbf{v}_i)$ is a one-dimensional A-invariant eigenspace that cannot be decomposed any further.

Another special case is when $\nu_i = 1 < n_i$ for an eigenvalue λ_i . In this case, J_i consists of a single Jordan subblock J_{i1} of size $n_{i1} = n_i$, and is of the form specified in Theorem 5.3, that is,

$$J_{i} = \begin{bmatrix} \lambda_{i} & 1 & 0 & \cdots & 0 \\ 0 & \lambda_{i} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_{i} \end{bmatrix}_{n_{i} \times n_{i}}$$

Subtracting $\lambda_i P_i$ from both sides of (5.25), we obtain

$$(A - \lambda_i I)P_i = P_i(J_i - \lambda_i I) = P_iQ_i$$
(5.27)

where

$$Q_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{n_i \times n}$$

Let

 $P_i = [\mathbf{v}_{i1} \cdots \mathbf{v}_{in_i}]$

Equating the columns of the products on both sides of (5.27), we get

$$(A - \lambda_i I)\mathbf{v}_{i1} = \mathbf{0}$$

(A - \lambda_i I)\mathbf{v}_{iq} = \mathbf{v}_{i,q-1}, \quad q = 2, ..., n_i (5.28)

A vector $\mathbf{w} \in \mathbf{C}^{n \times 1}$ for which

$$(A - \lambda_i I)^r \mathbf{w} = \mathbf{0}$$
$$(A - \lambda_i I)^{r-1} \mathbf{w} \neq \mathbf{0}$$

for some integer $r \ge 1$ is called a **generalized eigenvector** of rank r associated with λ_i .² A generalized eigenvector of rank one is an ordinary eigenvector. If **w** is a generalized eigenvector of rank r > 1, then $(A - \lambda_i I)$ **w** is a generalized eigenvector of rank r - 1, because

$$(A - \lambda_i I)^{r-1} (A - \lambda_i I) \mathbf{w} = (A - \lambda_i I)^r \mathbf{w} = \mathbf{0}$$

$$(A - \lambda_i I)^{r-2} (A - \lambda_i I) \mathbf{w} = (A - \lambda_i I)^{r-1} \mathbf{w} \neq \mathbf{0}$$

Thus a generalized eigenvector ${\bf w}$ of rank r generates a sequence of generalized eigenvectors

$$\mathbf{v}_1 = (A - \lambda_i I)^{r-1} \mathbf{w}, \, \mathbf{v}_2 = (A - \lambda_i I)^{r-2} \mathbf{w}, \, \dots, \, \mathbf{v}_r = \mathbf{w}$$

of ranks $1, 2, \ldots, r$, such that

$$(A - \lambda_i I)\mathbf{v}_1 = \mathbf{0}$$

(A - \lambda_i I)\mathbf{v}_q = \mathbf{v}_{q-1}, q = 2, \ldots, r

Together with the zero vector, the set of all generalized eigenvectors associated with a multiple eigenvalue λ_i is a subspace, called the **generalized eigenspace** of Aassociated with λ_i , and denoted by \mathbf{L}_i . Like an eigenspace \mathbf{K}_i , a generalized eigenspace \mathbf{L}_i is also A-invariant (see Exercise 5.12).

(5.28) implies that when $\nu_i = 1 < n_i$, the columns $\mathbf{v}_{i1}, \ldots, \mathbf{v}_{in_i}$ of P_i form a sequence of generalized eigenvectors associated with λ_i , generated by $\mathbf{w} = \mathbf{v}_{in_i}$. It can be shown that they form a basis for the generalized eigenspace \mathbf{L}_i . That is, $\mathbf{V}_i = \mathbf{L}_i$, and it can not be further decomposed into smaller A-invariant subspaces.

 $^{^{2}}$ The rank of a generalized eigenvector must not be confused with the rank of a matrix.

Example 5.17

The matrix

$$A = \left[\begin{array}{rrrr} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 2 & -2 & 3 \end{array} \right]$$

has the characteristic polynomial

$$d(s) = \left(s - 1\right)^3$$

Thus $\lambda_1 = 1$ is the only eigenvalue of A with $n_1 = 3$, and the Jordan form of A consists of a single Jordan block J_1 .

Since

$$A - I = \begin{bmatrix} -1 & 1 & 0\\ 1 & -1 & 2\\ 2 & -2 & 2 \end{bmatrix} \xrightarrow{\text{e.r.o.}} \begin{bmatrix} 1 & -1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0 \end{bmatrix}$$

 $\nu_1 = 1$, which means that J_1 consists of a single Jordan subblock J_{11} . Thus the Jordan form of A is

$$J = \left[\begin{array}{rrr} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{array} \right]$$

To find a modal matrix of A all we need to do is to find a generalized eigenvector **w** of rank $n_1 = 3$ associated with $\lambda_1 = 1$. For this puppose we compute $(A - I)^2$ and $(A - I)^3$:

$$(A-I)^{2} = \begin{bmatrix} 2 & -2 & 2 \\ 2 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$
$$(A-I)^{3} = O$$

A vector **w** for which $(A - I)^3 \mathbf{w} = \mathbf{0}$ but $(A - I)^2 \mathbf{w} \neq \mathbf{0}$ can be found as

$$\mathbf{w}_1 = \operatorname{col}[0, 0, 1]$$

which yields a modal matrix

$$P_{1} = [(A - I)^{2} \mathbf{w}_{1} \quad (A - I) \mathbf{w}_{1} \quad \mathbf{w}_{1}] = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$

Computing

$$P_1^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -2 & 2 \end{bmatrix}$$

we can verify that $P_1^{-1}AP_1 = J$.

Note that since A has only one eigenvalue, there is no decomposition of $\mathbf{C}^{n\times 1}$ into A-invariant subspaces, and therefore, the Jordan form of A consists of a single block. However, it has a much simpler form than A because of the appropriate choice of the basis vectors.

A different choice of ${\bf w}$ gives a different modal matrix. For example, the choice

 $\mathbf{w}_2 = \operatorname{col}[1, 0, 0]$

yields

$$P_2 = [(A - I)^2 \mathbf{w}_2 \quad (A - I) \mathbf{w}_2 \quad \mathbf{w}_2] = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 1 & 0 \\ 0 & 2 & 0 \end{bmatrix}$$

that also satisfies $P_2^{-1}AP_2 = J$.

This example also shows that although the Jordan form is unique, the modal matrix is not. The reader may try to find different generalized eigenvectors, and hence different modal matrices, and verify that they all result in the same Jordan form.

A class of matrices for which $\nu_i = 1$ for all eigenvalues (simple or multiple) are those in companion form. A matrix of the form

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$
(5.29)

is said to be in **companion form**. It is left to the reader as an exercise (see Exercise 5.24) to show that if A is of the form in (5.29), then

a) the characteristic polynomial of A is

$$d(s) = s^{n} + a_{1}s^{n-1} + \dots + a_{n-1}s + a_{n}$$

that is, the coefficients of the characteristic polynomial are the negatives of the last row elements of A in reverse order

- b) $\nu_i = 1$, $i = 1, \dots, k$
- c) the vectors

$$\mathbf{v}_{i1} = \mathbf{v}(\lambda_i), \, \mathbf{v}_{i2} = \frac{1}{1!} \, \mathbf{v}'(\lambda_i), \, \dots, \, \mathbf{v}_{in_i} = \frac{1}{(n_i - 1)!} \, \mathbf{v}^{(n_i - 1)}(\lambda_i) \quad (5.30)$$

form a sequence of generalized eigenvectors associated with λ_i , where

$$\mathbf{v}(s) = \operatorname{col}\left[1, s, \dots, s^{n-1}\right]$$

Example 5.18

The matrix

$$A = \left[\begin{array}{rrr} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{array} \right]$$

is in companion form. Its characteristic polynomial can be written from the last row elements as

$$d(s) = s^{3} - 4s^{2} + 5s - 2 = (s - 1)^{2}(s - 2)$$

From

$$\mathbf{v}(s) = \begin{bmatrix} 1\\s\\s^2 \end{bmatrix}, \quad \mathbf{v}'(s) = \begin{bmatrix} 0\\1\\2s \end{bmatrix}$$

we obtain a sequence of generalized eigenvectors associated with $\lambda_1 = 1$ as

$$\mathbf{v}_{11} = \begin{bmatrix} 1\\1\\1 \end{bmatrix}, \quad \mathbf{v}_{12} = \begin{bmatrix} 0\\1\\2 \end{bmatrix}$$

and a simple eigenvector associated with $\lambda_2 = 2$ as

$$\mathbf{v}_2 = \left[\begin{array}{c} 1\\ 2\\ 4 \end{array} \right]$$

Thus

$$P = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} 0 & 2 & -1 \\ -2 & 3 & -1 \\ 1 & -2 & 1 \end{bmatrix}$$

and

$$P^{-1}AP = J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Analysis of the case $1 < \nu_i < n_i$ is more complicated than the cases $\nu_i = n_i$ and $\nu_i = 1$ considered above, and is omitted. A worked example is given Exercise 5.28.

5.4.2 The Real Jordan Form

When A is a real matrix with real eigenvalues, we can choose its eigenvectors and generalized eigenvectors also real. We can then repeat the decomposition of the previous section by replacing $\mathbb{C}^{n\times 1}$ with $\mathbb{R}^{n\times 1}$. This leads to a real modal matrix and a real Jordan form. However, if A has complex eigenvalues then its eigenvectors and generalized eigenvectors, and therefore, its Jordan form will also be complex. In such a case, using the real and imaginary parts of the eigenvectors and generalized eigenvectors, we can construct a real modal matrix resulting in a real Jordan form. As in real semi-diagonal form, this corresponds to decomposing $\mathbb{R}^{n\times 1}$ into A-invariant subspaces.

Example 5.19

The matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -4 & 8 & -8 & 4 \end{bmatrix}$$

which is in companion form, has the characteristic polynomial

$$d(s) = s^{4} - 4s^{3} + 8s^{2} - 8s + 4 = (s^{2} - 2s + 2)^{2}$$

Thus A has a pair of complex conjugate eigenvalues $\lambda_{1,2} = 1 \mp i$ with multiplicities $n_{1,2} = 2$. A complex modal matrix can be formed as

$$P = \begin{bmatrix} \mathbf{v}(\lambda_1) & \mathbf{v}'(\lambda_1) & \mathbf{v}(\lambda_2) & \mathbf{v}'(\lambda_2) \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1+i & 1 & 1-i & 1 \\ 2i & 2+2i & -2i & 2-2i \\ -2+2i & 6i & -2-2i & -6i \end{bmatrix}$$

that yields a diagonal Jordan form

$$J = P^{-1}AP = \begin{bmatrix} 1+i & 1 & 0 & 0\\ 0 & 1+i & 0 & 0\\ 0 & 0 & 1-i & 1\\ 0 & 0 & 0 & 1-i \end{bmatrix}$$

On the other hand, by constructing a real modal matrix as

$$P_R = \left[\operatorname{Re} \mathbf{v}(\lambda_1) \operatorname{Im} \mathbf{v}(\lambda_1) \operatorname{Re} \mathbf{v}'(\lambda_1) \operatorname{Im} \mathbf{v}'(\lambda_1) \right]$$
$$= \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 2 & 2 & 2 \\ -2 & 2 & 0 & 6 \end{array} \right]$$

and computing its inverse, we obtain the real Jordan form of A as

$$J_R = P_R^{-1} A P_R = \begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} J_{R1} & I \\ O & J_{R1} \end{bmatrix}$$

5.5 Function of a Matrix

Let $A \in \mathbf{C}^{n \times n}$ have the characteristic polynomial d(s) in (5.7), and let the function $f : \mathbf{C} \to \mathbf{C}$ be analytic³ in an open disc $\mathcal{D} = \{s \in \mathbf{C} : |s| < r\}$ containing the eigenvalues of A. Let p be any polynomial such that

$$p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i), \quad i = 1, \dots, k, \quad j = 0, \dots, n_i - 1$$
 (5.31)

We then define the matrix f(A) to be

$$f(A) = p(A)$$

A polynomial p that satisfies (5.31) is called an **interpolating polynomial** of A for the function f.

There are three issues that must be considered in connection with this definition. The first is the existence of an interpolating polynomial. Let

$$p(s) = p_0 s^{n-1} + \dots + p_{n-2} s + p_{n-1} = \mathbf{v}^t(s)\mathbf{p}$$

 $^{^{3}}$ The reader is referred to a book on complex calculus for a definition and properties of analytic functions. For our purpose, it suffices to know that many functions of interest, such as polynomials, exponential, trigonometric and hyperbolic functions, are analytic.

where

$$\mathbf{v}(s) = \operatorname{col}[1, s, \dots, s^{n-1}], \quad \mathbf{p} = \operatorname{col}[p_{n-1}, p_{n-2}, \dots, p_0]$$

Then (5.31) can be written in matrix form as

$$P^t \mathbf{p} = \mathbf{f} \tag{5.32}$$

where

$$P = [\mathbf{v}(\lambda_1) \cdots \mathbf{v}^{(n_1-1)}(\lambda_1) \cdots \mathbf{v}(\lambda_k) \cdots \mathbf{v}^{(n_k-1)}(\lambda_k)]$$

and

$$\mathbf{f} = \operatorname{col} \left[f(\lambda_1) \cdots f^{(n_1 - 1)}(\lambda_1) \cdots f(\lambda_k) \cdots f^{(n_k - 1)}(\lambda_k) \right]$$

Thus P is the modal matrix (with columns scaled) of a matrix in companion form and having d(s) as its characteristic polynomial, and therefore, it is nonsingular. Hence (5.32) has a unique solution **p** that gives the coefficients of p(s). This proves the existence of an interpolating polynomial.

The second issue concerned with the definition of function of a matrix is the uniqueness of f(A). Obviously, there are infinitely many polynomials that satisfy (5.31). If p_1 and p_2 are any two such polynomials, then we must have $p_1(A) = p_2(A)$ for f(A) to be defined uniquely. This is indeed the case, for if $p = p_1 - p_2$ then

 $p^{(j)}(\lambda_i) = 0, \quad i = 1, \dots, k, \quad j = 0, \dots, n_i - 1$

which means that λ_i is a zero of p of order at least n_i . Then p must be of the form

p(s) = d(s)q(s)

for some polynomial q, and therefore,

$$p(A) = p_1(A) - p_2(A) = d(A)q(A) = O$$

that is, $p_1(A) = p_2(A)$.

The third concern with the definition of f(A) is an issue of consistency. From complex calculus it is known that a function that is analytic in some open disc \mathcal{D} in the complex plane has a power series representation

$$f(s) = \sum_{m=0}^{\infty} c_m s^m$$

which converges for all $s \in \mathcal{D}$. Then we expect that the infinite matrix series

$$\sum_{m=0}^{\infty} c_m A^m$$

also converge to f(A). The proof of the fact that

$$f(A) = \sum_{m=0}^{\infty} c_m A^m \tag{5.33}$$

is worked out in Exercise 5.31.

Example 5.20

Let us find $\sin \frac{\pi}{2} A$ for

$$A = \left[\begin{array}{cc} 0 & 1 \\ -1 & 2 \end{array} \right]$$

Since A is in companion form, we immediately write

$$d(s) = s^{2} - 2s + 1 = (s - 1)^{2}$$

Thus the only eigenvalue is $\lambda_1 = 1$ with $n_1 = 2$. Let $p(s) = p_0 s + p_1$ be an interpolating polynomial for $f(s) = \sin \frac{\pi}{2} s$. Then

$$p(\lambda_1) = p_0 + p_1 = f(\lambda_1) = \sin \frac{\pi}{2} = 1$$

$$p'(\lambda_1) = p_0 = f'(\lambda_1) = \frac{\pi}{2} \cos \frac{\pi}{2} = 0$$

Thus $p_0 = 0, p_1 = 1$, and

$$\sin\frac{\pi}{2}A = p(A) = p_0A + p_1I = I$$

Example 5.21

Let us calculate e^{At} for

$$A = \left[\begin{array}{rrr} -4 & -3 \\ 2 & 1 \end{array} \right]$$

The characteristic polynomial is d(s) = (s+1)(s+2). Let $p(s) = p_0 s + p_1$ be an interpolating polynomial for $f(s) = e^{st}$. Then from

$$p(-1) = -p_0 + p_1 = f(-1) = e^{-t}$$

$$p(-2) = -2p_0 + p_1 = f(-2) = e^{-2t}$$

we get

$$p_0 = e^{-t} - e^{-2t}$$
, $p_1 = 2e^{-t} - e^{-2t}$

Thus

$$e^{At} = p_0 A + p_1 I = \begin{bmatrix} -2e^{-t} + 3e^{-2t} & -3e^{-t} + 3e^{-2t} \\ 2e^{-t} - 2e^{-2t} & 3e^{-t} - 3e^{-2t} \end{bmatrix}$$

The reader can verify that

$$\frac{d}{dt}\,e^{At} = A e^{At}$$

where derivative of e^{At} is obtained by differentiating its elements individually. This is an expected result, which can be derived by differentiating the power series

$$e^{At} = \sum_{m=0}^{\infty} \frac{t^m}{m!} A^m$$

term-by-term as

$$\frac{d}{dt} e^{At} = \sum_{m=1}^{\infty} \frac{t^{m-1}}{(m-1)!} A^m = A \sum_{m=0}^{\infty} \frac{t^m}{m!} A^m = A e^{At}$$

Example 5.22

Let P be a projection matrix so that $P^2 = P$. Since

$$e^{st} = 1 + st + \frac{1}{2!}s^2t^2 + \frac{1}{3!}s^3t^3 + \cdots$$

we have

$$e^{Pt} = I + Pt + \frac{1}{2!}P^2t^2 + \frac{1}{3!}P^3t^3 + \cdots$$
$$= I + (t + \frac{1}{2!}t^2 + \frac{1}{3!}t^3 + \cdots)P$$
$$= I + (e^t - 1)P$$

Let A and B be similar matrices related as

$$A = PBP^{-1}$$

for some nonsingular matrix P. Then

$$A^m = (PBP^{-1})^m = PB^mP^{-1}, \quad m = 0, 1, \dots$$

as can easily be shown by induction on m, so that

$$p(A) = p(PBP^{-1}) = Pp(B)P^{-1}$$

for any polynomial p. Let f be a given function, and let p be an interpolating polynomial of B for f. Since A and B have the same characteristic polynomial, it follows from (5.31) that p is also an interpolating polynomial of A for f. Hence we have

$$f(A) = p(A) = Pp(B)P^{-1} = Pf(B)P^{-1}$$
(5.34)

This property allows us to evaluate f(A) using the Jordan (or diagonal) form of A as we explain below.

Let

$$B = \left[\begin{array}{cc} B_1 & & \\ & \ddots & \\ & & B_k \end{array} \right]$$

Then

$$B^{m} = \begin{bmatrix} B_{1}^{m} & & \\ & \ddots & \\ & & B_{k}^{m} \end{bmatrix}, \quad m = 0, 1, \dots$$

Since any polynomial of B is a linear combination of a finite number of powers of B, it follows that

$$p(B) = \begin{bmatrix} p(B_1) & & \\ & \ddots & \\ & & p(B_k) \end{bmatrix}$$

for any polynomial p. Now let f be a given function and let p be an interpolating polynomial of B for f so that f(B) = p(B). Since any eigenvalue λ of B_1 must be an eigenvalue of B and since the multiplicity of λ in the characteristic polynomial of B_1 can not exceed its multiplicity in the characteristic polynomial of B, it follows from (5.31) that p is also an interpolating polynomial of B_1 for f. The same is also true for all diagonal blocks of B. Hence $p(B_i) = f(B_i), i = 1, \ldots, k$, which implies that

$$f(B) = p(B) = \begin{bmatrix} p(B_1) & & \\ & \ddots & \\ & & p(B_k) \end{bmatrix} = \begin{bmatrix} f(B_1) & & \\ & \ddots & \\ & & f(B_k) \end{bmatrix}$$

Combining the above expression with (5.34) we observe that if A has the Jordan form $J = P^{-1}AP = \text{diag}[J_{ij}]$, then $A = PJP^{-1}$, and therefore

$$f(A) = Pf(J)P^{-1} = P \cdot \text{diag}[f(J_{ij})] \cdot P^{-1}$$
(5.35)

In particular, if A is diagonalizable then $J = D = \text{diag}[d_i]$ and (5.35) reduces to

$$f(A) = P \cdot \operatorname{diag}\left[f(d_i)\right] \cdot P^{-1}$$

If A is not diagonalizable, for (5.35) to be useful in evaluating f(A) we need an expression for the function of a Jordan subblock

$$J = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}_{n \times n}$$
(5.36)

where we omitted the subscripts for simplicity. For a given function f, let

$$p(s) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\lambda)}{k!} (s - \lambda)^k$$

Since

$$p^{(j)}(s) = \sum_{k=j}^{n-1} \frac{f^{(k)}(\lambda)}{(k-j)!} (s-\lambda)^{k-j}, \quad j = 0, 1, \dots, n-1$$

it follows that

$$p^{(j)}(\lambda) = f^{(j)}(\lambda), \quad j = 0, 1, \dots, n-1$$

Thus p is an interpolating polynomial of J for f, and therefore

$$f(J) = p(J) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\lambda)}{k!} (J - \lambda I)^k = \sum_{k=0}^{n-1} \frac{f^{(k)}(\lambda)}{k!} Q^k$$
(5.37)

where

$$Q = J - \lambda I = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{n \times n}$$

Recalling from Exercise 1.18 that the line of 1's above the diagonal of Q shifts one position upwards with each power of Q, we obtain from (5.37)

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!} f''(\lambda) & \cdots & \frac{1}{(n-1)!} f^{(n-1)}(\lambda) \\ 0 & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!} f^{(n-2)}(\lambda) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & f'(\lambda) \\ 0 & 0 & 0 & \cdots & f(\lambda) \end{bmatrix}$$
(5.38)

Example 5.23

Let us calculate e^{At} for the matrix A in Example 5.17.

Since we have already obtained the Jordan form of A, we can immediately write

$$e^{At} = P_1 e^{Jt} P_1^{-1}$$

$$= \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} e^t & te^t & t^2 e^t/2 \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -2 & 2 \end{bmatrix}$$

$$= e^t \begin{bmatrix} t^2 - t + 1 & -t^2 + t & t^2 \\ t^2 + t & -t^2 - t + 1 & t^2 + 2t \\ 2t & -2t & 2t + 1 \end{bmatrix}$$

To check the result, let $p(s) = \alpha s^2 + \beta s + \gamma$ be an interpolating polynomial for $f(s) = e^{st}$. Since A has the only eigenvalue $\lambda_1 = 1$ with $n_1 = 3$, (5.31) becomes

$$\begin{array}{rcl} p(1) & = & \alpha + \beta + \gamma & = & f(1) & = & e^t \\ p'(1) & = & 2\alpha + \beta & = & f'(1) & = & te^t \\ p''(1) & = & 2\alpha & = & f''(1) & = & t^2e^t \end{array}$$

Note that f' denotes derivative with respect to s so that

$$f'(1) = \left[\frac{d}{ds} e^{st}\right]_{s=1} = \left[te^{st}\right]_{s=1} = te^{t}$$

Similarly,

$$f^{\prime\prime}(1) = t^2 e^t$$

Solving for α, β, γ , we get

$$\alpha = \frac{t^2}{2} e^t$$
, $\beta = (t - t^2) e^t$, $\gamma = (1 - t + \frac{t^2}{2}) e^t$

and

 $e^{At} = \alpha A^2 + \beta A + \gamma I$

gives the same result.

We finally note a couple of properties of function of a matrix. The first is that since any function of a matrix is defined through an interpolating polynomial, we have

$$f(A)g(A) = g(A)f(A)$$

for any f and g. A second property, which can be shown using the Jordan form of A, is that if A has the characteristic polynomial in (5.7), then f(A) has the characteristic polynomial

$$\prod_{i=1}^{k} (s - f(\lambda_i))^{n_i}$$

In other words, the eigenvalues of f(A) are $f(\lambda_i)$ with the same multiplicities in the characteristic polynomial. For example, the eigenvalues of A + I are $\lambda_i + 1$, those of A^m are λ_i^m , and if A is nonsingular, the eigenvalues of A^{-1} are λ_i^{-1} .

5.6 Exercises

1. Find eigenvalues and eigenvectors of the following matrices

$A = \left[\begin{array}{rrr} 1 & 2\\ 2 & 4 \end{array} \right]$	$B = \left[\begin{array}{cc} 4 & 3\\ -6 & -2 \end{array} \right]$
$C = \left[\begin{array}{cc} 0 & i \\ -i & 0 \end{array} \right]$	$D = \left[\begin{array}{cc} 4i & 3\\ -3 & -4i \end{array} \right]$
$E = \left[\begin{array}{rrrr} 2 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 5 \end{array} \right]$	$F = \left[\begin{array}{rrrr} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{array} \right]$
$G = \left[\begin{array}{rrrr} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{array} \right]$	$H = \left[\begin{array}{rrrr} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{array} \right]$
$K = \left[\begin{array}{rrrrr} -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & -2 & 2 \\ 0 & 1 & -2 & 2 \end{array} \right]$	$M = \begin{bmatrix} -1 & 1 & 0 & 1 \\ 0 & -3 & 0 & -2 \\ -2 & 0 & -1 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix}$

2. (a) Show that det(sI - A) is an *n*th degree polynomial in *s* with a unity leading coefficient. Hint: Let $A(s) = sI - A = [a_{ij}(s)]$ and consider the product terms in

$$\det A(s) = \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1}(s) \cdots a_{nj_n}(s)$$

(b) Show that elements of $B(s) = [b_{ij}(s)] = \text{adj } (sI - A)$ are polynomials in s of degree not exceeding n - 1. Hint: Recall that

$$b_{ij}(s) = (-1)^{i+j} m_{ji}^B(s)$$

3. Verify the Cayley-Hamilton theorem for the matrices in Exercise 5.1.

4. (a) Use the Cayley-Hamilton theorem to calculate A^{100} for

$$A = \left[\begin{array}{cc} 0 & i \\ i & 0 \end{array} \right]$$

(b) Check your result by MATLAB.

5. An $n \times n$ real matrix $A = [a_{ij}]$ whose elements satisfy

$$a_{ij} \ge 0$$
 for all (i, j) , $\sum_{i=1}^{n} a_{ij} = 1$ for all j

is called a **Markov matrix**. Show that s = 1 is an eigenvalue of a Markov matrix.

6. Prove the **Gersgorin's theorem**: The eigenvalues of an $n \times n$ complex matrix $A = [a_{ij}]$ are located in the union of the n discs

$$\mathcal{D}_k: |s - a_{kk}| \le \rho_k = \sum_{j \ne k}^n |a_{kj}|, \quad k = 1, 2, \dots, n$$

Hint: If λ is an eigenvalue of A with an associated eigenvector $\mathbf{x} = \operatorname{col} [x_1, x_2, \dots, x_n]$ then

$$(\lambda - a_{ii})x_i = \sum_{j \neq i}^n a_{ij}x_j, \quad i = 1, 2, \dots, n$$

Suppose

$$\max_{i} |x_i| = |x_k|$$

and consider the absolute values of both sides of the above equality with i = k.

7. Use Gershgorin's theorem to find upper and lower bounds on the magnitudes of the eigenvalues of the matrix

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

8. An $n \times n$ matrix $A = [a_{ij}]$ whose elements satisfy

$$|a_{kk}| < \sum_{j \neq k}^{n} |a_{kj}|, \quad k = 1, 2, \dots, n$$

is said to be **diagonally dominant**.

- (a) Show that a diagonally dominant matrix is nonsingular.
- (b) Show that if the diagonal elements of a diagonally dominant matrix A are real and positive, then all the eigenvalues of A have positive real parts.
- 9. Prove the following properties of the minimum polynomial $\alpha(s)$ of a matrix A having a characteristic polynomial as given in (5.7).
 - (a) Show that $\alpha(s)$ is unique. Hint: Suppose that

$$\alpha(s) = s^m + \alpha_1 s^{m-1} + \dots + \alpha_m$$

and

$$\alpha'(s) = s^m + \alpha'_1 s^{m-1} + \dots + \alpha'_m$$

are two distinct minimum polynomials of A satisfying $\alpha(A) = \alpha'(A) = O$. (Note that α and α' must have the same degree, otherwise one of them can not be a minimum polynomial.) Define $r(s) = \alpha(s) - \alpha'(s)$. Then $r \neq 0$ and r(A) = O. Argue how this fact leads to a contradiction.

(b) Show that every eigenvalue λ_i of A is a zero of $\alpha(s)$. Hint: Divide $\alpha(s)$ with $(s - \lambda_i)$ and write

 $\alpha(s) = q_i(s)(s - \lambda_i) + r_i$

where r_i is a constant remainder. Then

 $q_i(A)(A - \lambda_i I) + r_i I = \alpha(A) = O$

Postmultiply both sides by an eigenvector \mathbf{v}_i associated with λ_i to show that $r_i = 0$.

(c) Show that $\alpha(s)$ divides d(s). Hint: Divide d(s) with $\alpha(s)$ and write

 $d(s) = q(s)\alpha(s) + r(s)$

and show that r(s) = 0.

Properties in (b) and (c) imply that the minimum polynomial must be of the form

$$\alpha(s) = \prod_{i=1}^{n} (s - \lambda_i)^{m_i}, \quad 1 \le m_i \le n_i$$

10. Given that the polynomials

$$p_1(s) = s^7 + 2s^6 - 2s^5 - 4s^4 + 2s^3 + 5s^2 + 2s$$

$$p_2(s) = s^2(s+1)^2(s+2)$$

satisfy $p_1(A) = p_2(A) = O$ for a 10×10 matrix A.

- (a) What can you say about the degree of the minimum polynomial of A?
- (b) What are the remainder polynomials when p_1 and p_2 are divided by the minimum polynomial?
- (c) How many distinct eigenvalues can A have?
- (d) If A has 3 distinct eigenvalues, what are they? What is the minimum polynomial in this case?

11. Show that if

 $\mathbf{C}^{n\times 1} = \mathbf{U}_1 \oplus \mathbf{U}_2$

where \mathbf{U}_1 is A-invariant, then A is similar to a matrix of the form

$$F = \left[\begin{array}{cc} F_{11} & F_{12} \\ O & F_{22} \end{array} \right]$$

Hint: Let

$$P = [P_1 \ P_2]$$

where columns of P_1 and P_2 form bases for \mathbf{U}_1 and \mathbf{U}_2 .

12. Prove that if λ_i is an eigenvalue with multiplicity n_i in d(s) then dim $(\mathbf{K}_i) \leq n_i$. Hint: Let dim $(\mathbf{K}_i) = \eta$, and let \mathbf{V}_i be a complement of \mathbf{K}_i so that

$$\mathbf{C}^{n\times 1} = \mathbf{K}_i \oplus \mathbf{V}_i$$

and use the result of Exercise 5.11 with $\mathbf{U}_1 = \mathbf{K}_i$ and $\mathbf{U}_2 = \mathbf{V}_i$.

13. Prove, without using the Jordan form, that every square matrix $A \in \mathbf{C}^{n \times n}$ is similar to an upper triangular matrix U whose diagonal elements are the eigenvalues of A. Hint: Let $A_1 = A$, and let \mathbf{v}_1 be an eigenvector of A_1 associated with an eigenvalue λ_1 . Choose $\mathbf{v}_2, \ldots, \mathbf{v}_n$ such that

$$P_1 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$$

is nonsingular. Show that

$$P_1^{-1}A_1P_1 = \left[\begin{array}{cc} \lambda_1 & \boldsymbol{\alpha}_1 \\ \mathbf{0} & A_2 \end{array}\right]$$

where α_1 is a row (n-1)-vector and A_2 is of order n-1. Then use induction on n.

14. The matrices

$$Q_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Q_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

that occur in quantum mechanics are called **Pauli spin matrices**. Find eigenvalues and eigenvectors of Pauli spin matrices, and show that they are similar.

- 15. Find the Jordan forms of the matrices in Exercise 5.1.
- 16. Use MATLAB command [P,J]=eig(A) to find a modal matrix and the Jordan form of the matrices in Exercise 5.1 and comment on the results.
- 17. Find eigenvalues, eigenvectors, a modal matrix, and the Jordan form of

$$A = \left[\begin{array}{rrrr} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{array} \right]$$

- 18. Use MATLAB command A=rand(n,n) to generate random matrices of different order, and use the command [P,J]=eig(A) to find their Jordan form. This exercise shows that almost all square matrices have distinct eigenvalues, and are therefore diagonalizable.
- 19. Find the Jordan forms of the following matrices.

$$A = \begin{bmatrix} \sigma & & & & \\ & \sigma & 1 & & \\ & & \sigma & & \\ & & & \mu & \\ & & & & \sigma \end{bmatrix} \qquad B = \begin{bmatrix} \sigma & 1 & & \\ & \mu & 1 & \\ & & \sigma & & \\ & & & \mu & \\ & & & & \sigma \end{bmatrix}$$

20. A linear transformation $\mathcal{A}: \mathbf{R}^2 \to \mathbf{R}^2$ is defined as

$$\mathcal{A}(x,y) = (x+y, -x+3y)$$

- (a) Find the matrix representation of \mathcal{A} with respect to the basis $\mathbf{v}_1 = (1, 1), \mathbf{v}_2 = (1, 2)$ in \mathbf{R}^2 .
- (b) Can you find a basis with respect to which A has a diagonal representation? If yes, find it; otherwise, explain why.
- 21. Let $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$ be a basis for a four-dimensional complex vector space \mathbf{X} , and let a linear operator \mathcal{A} on \mathbf{X} be defined as

$$\mathcal{A}(\mathbf{u}_1) = \mathbf{u}_4, \quad \mathcal{A}(\mathbf{u}_2) = \mathbf{u}_1, \quad \mathcal{A}(\mathbf{u}_3) = \mathbf{u}_2, \quad \mathcal{A}(\mathbf{u}_4) = \mathbf{u}_3$$

Find eigenvalues and eigenvectors of \mathcal{A} . Hint: First obtain a matrix representation of \mathcal{A} with respect to the given basis.

22. Let $\mathbf{X} = \mathbf{R}_1[s] = \{ p(s) = as + b \mid a, b \in \mathbf{R} \}$, and let a linear transformation $\mathcal{A} : \mathbf{X} \to \mathbf{X}$ be defined as

 $\mathcal{A}(as+b) = bs+a$

Find a basis for **X** with respect to which the matrix representation of \mathcal{A} is in Jordan form.

- 23. Show that the generalized eigenspace \mathbf{L}_i of an eigenvalue λ_i of a matrix A is A-invariant. Hint: First show that \mathbf{L}_i is a subspace, and then show that if $\mathbf{w} \in \mathbf{L}_i$ is a generalized eigenvector of rank r, then so is $A\mathbf{w}$.
- 24. Let A be an $n \times n$ matrix in companion form as given in (5.29). Define

$$C(s) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ s & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s^{n-1} & s^{n-2} & \cdots & 1 \end{bmatrix}$$

(a) Show that

$$\det (sI - A) = \det [(sI - A)C(s)] = s^{n} + a_{1}s^{n-1} + \dots + a_{n}$$

(b) Show that

$$r(\lambda_i I - A) = r[(\lambda_i I - A)C(\lambda_i)] = n - 1$$

for every eigenvalue of A, so that $\nu_i = 1, i = 1, \dots, k$.

(c) Let $\mathbf{y}(s) = (sI - A)\mathbf{v}(s)$. Calculate $\mathbf{y}(s)$ and show that

$$\mathbf{y}(\lambda_i) = \mathbf{y}'(\lambda_i) = \cdots = \mathbf{y}^{(n_i - 1)}(\lambda_i) = \mathbf{0}$$

Use this result together with

$$\mathbf{y}'(s) = \mathbf{v}(s) + (sI - A)\mathbf{v}'(s)$$
$$\mathbf{y}''(s) = 2\mathbf{v}'(s) + (sI - A)\mathbf{v}''(s)$$

and so on, to show that the vectors in (5.30) form a sequence of generalized eigenvectors associated with λ_i .

25. Find the Jordan form of the following matrix in companion form.

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

- 26. Let $A \in \mathbf{C}^{n \times n}$ have *n* distinct eigenvalues λ_i with the associated eigenvectors \mathbf{v}_i , i = 1, 2, ..., n, and let $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n$. Show that $\mathbf{v}, A\mathbf{v}, \ldots, A^{n-1}\mathbf{v}$ are linearly independent.
- 27. Let $A = \mathbf{bc}^{T}$, where $\mathbf{b}, \mathbf{c} \in \mathbf{R}^{n \times 1}$ and $\mathbf{c}^{T}\mathbf{b} \neq 0$. Obtain the Jordan form of A. Hint: Show that $\lambda_{1} = \mathbf{c}^{T}\mathbf{b}$ is an eigenvalue of A and $\mathbf{v}_{1} = \mathbf{b}$ is an associated eigenvector. What are the other eigenvalues?
- 28. The matrix

$$A = \left[\begin{array}{rrr} 0 & 4 & -4 \\ 1 & 0 & 2 \\ 2 & -4 & 6 \end{array} \right]$$

has the characteristic polynomial

 $d(s) = (s-2)^3$

Thus it has a single eigenvalue $\lambda_1 = 2$ with multiplicity $n_1 = 3$. From

	-2	4	-4		[1]	-2	2]
A - 2I =	1	-2	2	\xrightarrow{R}	0	0	0
	2	-4	4		0	0	0

we find that $1 < \nu_1 = 2 < 3$. Then the Jordan form of A must have a single Jordan block consisting of $\nu_1 = 2$ subblocks of orders two and one respectively (since the sum of the orders must be $n_1 = 3$, this is the only possibility). Find suitable generalized eigenvectors to form a modal matrix P, and show that $P^{-1}AP = J$. Hint: You need a generalized eigenvector of rank two and an ordinary eigenvector that is linearly independent of the vectors of the sequence generated by the generalized eigenvector.

29. Show that the real modal matrix P_R in (5.18) is nonsingular. Hint: Suppose that

$$a_1\mathbf{u}_1 + b_1\mathbf{w}_1 + \dots + a_m\mathbf{u}_m + b_m\mathbf{w}_m + c_{2m+1}\mathbf{v}_{2m+1} + \dots + c_n\mathbf{v}_n = \mathbf{0}$$

Then

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_{2m-1}\mathbf{v}_{2m-1} + c_{2m}\mathbf{v}_{2m} + c_{2m+1}\mathbf{v}_{2m+1} + \dots + c_n\mathbf{v}_n = \mathbf{0}$$

for some $c_i, i = 1, ..., 2m$.

- 30. Find the real Jordan forms of
 - (a) the matrix in Exercise 5.25

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

- 31. (a) Prove (5.33) for a diagonal matrix D.
 - (b) Prove (5.33) for a single Jordan subblock J given in (5.36).
 - (c) Use the results of parts (a) and (b), together with (5.35), to prove (5.33) for an arbitrary matrix A.
- 32. Find a general expression for e^{At} if

$$A = \left[\begin{array}{cc} \sigma & \omega \\ -\omega & \sigma \end{array} \right]$$

Note that A is in real Jordan form with eigenvalues $\lambda_{1,2} = \sigma \mp i\omega$.

33. Calculate
$$e^{At}$$
 for

(a)
$$A = \begin{bmatrix} -2 & 2 \\ 1 & -3 \end{bmatrix}$$
 by diagonalizing A ,
(b) $A = \begin{bmatrix} -1 & 1 \\ -2 & -3 \end{bmatrix}$ by using an interpolating polynomial,
(c) $A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$ by any method.

34. Calculate $\sin(\pi A)$ and $\cos(\pi A)$ for

$$A = \left[\begin{array}{cc} 0 & 1\\ -1/6 & 5/6 \end{array} \right]$$

and verify the equality $\sin^2(\pi A) + \cos^2(\pi A) = I$. Check your results by MATLAB.

- 35. Show that the matrices f(A) and g(A) commute for arbitrary functions f and g.
- 36. Let A be an $n \times n$ real matrix with distinct eigenvalues $\lambda_i, i = 1, 2, ..., n$.
 - (a) Show that $A = \sum_{i=1}^{n} \lambda_i Q_i$ for some matrices Q_i which satisfy

$$Q_i^2 = Q_i$$
, $Q_i Q_j = 0, i \neq j$, and $\sum_{i=1}^n Q_i = I$

Hint: Start with $A = PDP^{-1}$, partition P into columns and P^{-1} into rows.

(b) Show that with Q_i as above,

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) Q_i$$

for any function f for which f(A) is defined.

37. Verify the result of Exercise 5.36 for the matrix A in Exercise 5.33(a) and for the function $f(s) = e^{st}$.