6

Eigenvalue Problems of Matrices

Given a square matrix A, to determine the scalars λ and the nonzero column matrix $\mathbf{x}$ which simultaneously satisfy the equation

$$A\mathbf{x} = \lambda\mathbf{x} \tag{6.1}$$

is known as the *eigenvalue* problem (eigen in German means proper). The solution of this problem is intimately connected to the question of whether the matrix can be transformed into a diagonal form.

The eigenvalue problem is of great interests in many engineering applications, such as mechanical vibrations, alternating currents, and rigid body dynamics. It is of crucial importance in modern physics. The whole structure of quantum mechanics is based on the diagonalization of certain type of matrices.

6.1 Eigenvalues and Eigenvectors

6.1.1 Secular Equation

In the *eigenvalue* problem, the value λ is called the *eigenvalue* (*characteristic value*) and the corresponding column matrix $\mathbf{x}$ is called the *eigenvector* (*characteristic vector*). If A is a $n \times n$ matrix, (6.1) is given by

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Since

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda I\mathbf{x},$$

where I is the unit matrix, we can write (6.1) as

$$(A - \lambda I)\,\mathbf{x} = 0. \tag{6.2}$$

This system has nontrivial solutions if and only if the determinant of the coefficient matrix vanishes

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} = 0. \tag{6.3}$$

The expansion of this determinant yields a polynomial of degree n in λ, which is called *characteristic polynomial* $P(\lambda)$. The equation

$$P(\lambda) = |A - \lambda I| = 0 \tag{6.4}$$

is known as the *characteristic equation* (or *secular equation*). Its n roots are the eigenvalues and will be denoted $\lambda_1, \lambda_2, \dots, \lambda_n$. They may be real or complex. When one of the eigenvalues is substituted back into (6.2), the corresponding eigenvector $\mathbf{x}(x_1, x_2, \dots, x_n)$ may be determined. Note that the eigenvectors may be multiplied by any constant and remain a solution of the equation.

We will denote $\mathbf{x}_i$ as the the eigenvector belonging to the eigenvalue λ_i. That is, if

$$P(\lambda_i) = 0,$$

then

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i.$$

If n eigenvalues are all different, we will have n distinct eigenvectors. If two or more eigenvalues are the same, we say that they are degenerate. In some problems, a degenerate eigenvalue may produce only one eigenvector, in other problems a degenerate eigenvalue may produce more than one distinct eigenvectors.

Example 6.1.1. Find the eigenvalues and eigenvectors of A, if

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Solution 6.1.1. The characteristic polynomial of A is

$$P(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 3,$$

and the secular equation is

$$\lambda^2 - 2\lambda - 3 = (\lambda + 1)(\lambda - 3) = 0.$$

Thus the eigenvalues are

$$\lambda_1 = -1, \quad \lambda_2 = 3.$$

Let the eigenvector $\mathbf{x}_1$ corresponding to $\lambda_1 = -1$ be $\begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}$, then $\mathbf{x}_1$ must satisfy

$$\begin{pmatrix} 1-\lambda_1 & 2 \\ 2 & 1-\lambda_1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = 0 \Longrightarrow \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = 0.$$

This reduces to

$$2x_{11} + 2x_{12} = 0.$$

Thus for this eigenvector $x_{11} = -x_{12}$. That is, $x_{11} : x_{12} = -1 : 1$. Therefore the eigenvector can be written as

$$\mathbf{x}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Any constant, positive or negative, times it will also be a solution, but it will not be regarded as another distinct eigenvector. With a similar procedure, we find the eigenvector corresponding to $\lambda_2 = 3$ to be

$$\mathbf{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Example 6.1.2. Find the eigenvalues and eigenvectors of A, if

$$A = \begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix}.$$

Solution 6.1.2. The characteristic polynomial of A is

$$P(\lambda) = \begin{vmatrix} 3-\lambda & -5 \\ 1 & -1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda + 2,$$

so the secular equation is

$$\lambda^2 - 2\lambda + 2 = 0.$$

Thus the eigenvalues are

$$\lambda = 1 \pm \mathrm{i}.$$

Let $\lambda_1 = 1+\mathrm{i}$, and the corresponding eigenvector $\mathbf{x}_1$ be $\begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}$, then $\mathbf{x}_1$ must satisfy

$$\begin{pmatrix} 3-(1+\mathrm{i}) & -5 \\ 1 & -1-(1+\mathrm{i}) \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = 0,$$

which gives

$$(2-\mathrm{i})x_{11} - 5x_{12} = 0,$$

$$x_{11} - (2+\mathrm{i})x_{12} = 0.$$

The first equation gives

$$x_{11} = \frac{5}{2-\mathrm{i}}x_{12} = \frac{5(2+\mathrm{i})}{4+1}x_{12} = \frac{2+\mathrm{i}}{1}x_{12},$$

which is the same result from the second equation, as it should be. Therefore $\mathbf{x}_1$ can be written as

$$\mathbf{x}_1 = \begin{pmatrix} 2+\mathrm{i} \\ 1 \end{pmatrix}.$$

Similarly, for $\lambda = \lambda_2 = 1 - \mathrm{i}$, the corresponding eigenvector is

$$\mathbf{x}_2 = \begin{pmatrix} 2-\mathrm{i} \\ 1 \end{pmatrix}.$$

So we have an example of a real matrix with complex eigenvalues and complex eigenvectors.

Example 6.1.3. Find the eigenvalues and eigenvectors of A, if

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}.$$

Solution 6.1.3. The characteristic polynomial of A is

$$P(\lambda) = \begin{vmatrix} -2-\lambda & 2 & -3 \\ 2 & 1-\lambda & -6 \\ -1 & -2 & -\lambda \end{vmatrix} = -\lambda^3 - \lambda^2 + 21\lambda + 45.$$

The secular equation can be written as

$$\lambda^3 + \lambda^2 - 21\lambda - 45 = (\lambda - 5)(\lambda + 3)^2 = 0.$$

This equation has a single root of 5 and a double root of -3. Let

$$\lambda_1 = 5, \quad \lambda_2 = -3, \quad \lambda_3 = -3.$$

The eigenvector belonging to the eigenvalue of λ_1 must satisfy the equation

$$\begin{pmatrix} -2-5 & 2 & -3 \\ 2 & 1-5 & -6 \\ -1 & -2 & 0-5 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix} = 0.$$

With Gauss' elimination method, this equation can be shown to be equivalent to

$$\begin{pmatrix} -7 & 2 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix} = 0,$$

which means

$$\begin{aligned} -7x_{11} + 2x_{12} - 3x_{13} &= 0, \\ x_{12} + 2x_{13} &= 0. \end{aligned}$$

Assign $x_{13} = 1$, then $x_{12} = -2, x_{11} = -1$. So corresponding to $\lambda_1 = 5$, the eigenvector $\mathbf{x}_1$ can be written as

$$\mathbf{x}_1 = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}.$$

Since the eigenvalue of -3 is twofold degenerate, corresponding to this eigenvalue, we may have one or two eigenvectors. Let us express the eigenvector corresponding to the eigenvalue of -3 as $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. It must satisfy the equation

$$\begin{pmatrix} -2+3 & 2 & -3 \\ 2 & 1+3 & -6 \\ -1 & -2 & 0+3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

With Gauss' elimination method, this equation can be shown to be equivalent to

$$\begin{pmatrix} 1 & 2 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0,$$

which means

$$x_1 + 2x_2 - 3x_3 = 0.$$

We can express x_1 in terms of x_2 and x_3, and there is no restriction on x_2 and x_3. Let $x_2 = c_2$ and $x_3 = c_3$, then $x_1 = -2c_2 + 3c_3$. So we can write

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2c_2 + 3c_3 \\ c_2 \\ c_3 \end{pmatrix} = c_2 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}.$$

Since c_2 and c_3 are arbitrary, we can first assign $c_3 = 0$ and get an eigenvector, and then assign $c_2 = 0$ and get another eigenvector. So corresponding to the degenerate eigenvalue $\lambda = -3$, there are two distinct eigenvectors

$$\mathbf{x}_2 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}.$$

In this example, we have only two distinct eigenvalues, but we still have three distinct eigenvectors.

Example 6.1.4. Find the eigenvalues and eigenvectors of A, if

$$A = \begin{pmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{pmatrix}.$$

Solution 6.1.4. The characteristic polynomial of A is

$$P(\lambda) = \begin{vmatrix} 4-\lambda & 6 & 6 \\ 1 & 3-\lambda & 2 \\ -1 & -5 & -2-\lambda \end{vmatrix} = -\lambda^3 + 5\lambda^2 - 8\lambda + 4.$$

The secular equation can be written as

$$\lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda - 2)^2 = 0.$$

The three eigenvalues are

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_3 = 2.$$

From the equation for the eigenvector $\mathbf{x}_1$ belonging to the eigenvalue of λ_1

$$\begin{pmatrix} 4-1 & 6 & 6 \\ 1 & 3-1 & 2 \\ -1 & -5 & -2-1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix} = 0,$$

we obtain the solution

$$\mathbf{x}_1 = \begin{pmatrix} 4 \\ 1 \\ -3 \end{pmatrix}.$$

The eigenvector $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, corresponding to the twofold degenerate eigenvalue 2, satisfies the equation

$$\begin{pmatrix} 4-2 & 6 & 6 \\ 1 & 3-2 & 2 \\ -1 & -5 & -2-2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

With Gauss' elimination method, this equation can be shown to be equivalent to

$$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0,$$

which means

$$\begin{aligned} x_1 + x_2 + 2x_3 &= 0, \\ 2x_2 + x_3 &= 0. \end{aligned}$$

If we assign $x_3 = -2$, then $x_2 = 1$ and $x_1 = 3$. So

$$\mathbf{x}_2 = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}.$$

The two equations above do not allow any other eigenvector which is not just a constant times $\mathbf{x}_2$. Therefore for this 3×3 matrix, there are only two distinct eigenvectors.

Computer Code

It should be noted that for large systems, the eigenvalues and eigenvectors would usually be found with specialized numerical methods (see, for example, G.H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins University Press, 1983). There are excellent general purpose computer programs for the efficient and accurate determination of eigensystems (see, for example, B.T. Smith, J.M. Boyle, J. Dongarra, B. Garbow, Y. Ikebe, V.C. Klema, and C.B. Moler, *Matrix Eigensystem Routines: EISPACK Guide*, 2nd edn. Springer-Verlag, 1976).

In addition, eigenvalues and eigenvectors can be found with a simple command in computer packages such as, Maple, Mathematica, MathCad, and MuPAD. These packages are known as *Computer Algebraic Systems*.

This book is written with the software "Scientific WorkPlace," which also provides an interface to MuPAD (before version 5, it also came with Maple).

Instead of requiring the user to adhere to a rigid syntax, the user can use natural mathematical notations. For example, to find the eigenvalues and eigenvectors of

$$\begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix},$$

all you have to do is (1) type the expression in the math-mode, (2) click on the "Compute" button, (3) click on the "Matrices" button in the pull-down menu, and (4) click on the "Eigenvectors" button in the submenu. The program will return with

$$\text{eigenvectors}: \quad \left\{ \begin{pmatrix} 1 \\ -\frac{1}{3} \\ 1 \end{pmatrix} \right\} \leftrightarrow 1, \quad \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\} \leftrightarrow 2.$$

You can ask the program to check the results. For example, you can type

$$\begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix},$$

and click on the "Compute" button, and then click on the "Evaluate" button. The program will return with

$$\begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix},$$

which is of course equal to $2\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$, showing $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$ is indeed an eigenvector belonging to eigenvalue 2. The other two eigenvectors can be similarly checked.

Computer Algebraic Systems are wonderful as they are, they must be used with caution. It is not infrequent that the system will return with an answer to a wrong problem without the user knowing it. Therefore answers from these systems should be checked. Computer Algebraic Systems are useful supplements, but they are no substitute for the knowledge of the subject matter.

6.1.2 Properties of Characteristic Polynomial

The characteristic polynomial has several useful properties. To elaborate on them, let us first consider the case of $n = 3$.

$$
\begin{aligned}
P(\lambda) &= \begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} + \begin{vmatrix} -\lambda & a_{12} & a_{13} \\ 0 & a_{22}-\lambda & a_{23} \\ 0 & a_{32} & a_{33}-\lambda \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & a_{13} \\ a_{21} & -\lambda & a_{23} \\ a_{31} & 0 & a_{33}-\lambda \end{vmatrix} \\
&\quad + \begin{vmatrix} -\lambda & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33}-\lambda \end{vmatrix} + \begin{vmatrix} -\lambda & 0 & a_{13} \\ 0 & -\lambda & a_{23} \\ 0 & 0 & a_{33}-\lambda \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & -\lambda \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & a_{13} \\ a_{21} & -\lambda & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & -\lambda & 0 \\ a_{31} & 0 & -\lambda \end{vmatrix} \\
&\quad + \begin{vmatrix} -\lambda & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} -\lambda & a_{12} & 0 \\ 0 & a_{22} & 0 \\ 0 & a_{32} & -\lambda \end{vmatrix} + \begin{vmatrix} -\lambda & 0 & a_{13} \\ 0 & -\lambda & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} \\
&= |A| + \left(\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \right) (-\lambda) \\
&\quad + (a_{11} + a_{22} + a_{33})(-\lambda)^2 + (-\lambda)^3. \qquad (6.5)
\end{aligned}
$$

Now let $\lambda_1, \lambda_2, \lambda_3$ be the eigenvalues, so $P(\lambda_1) = P(\lambda_2) = P(\lambda_3) = 0$. Since $P(\lambda)$ is a polynomial of degree 3, it follows that:

$$P(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)(\lambda_3 - \lambda) = 0.$$

Expanding the characteristic polynomial,

$$P(\lambda) = \lambda_1\lambda_2\lambda_3 + (\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1)(-\lambda) + (\lambda_1 + \lambda_2 + \lambda_3)(-\lambda)^2 + (-\lambda)^3.$$

Comparison with (6.5) shows

$$\lambda_1 + \lambda_2 + \lambda_3 = a_{11} + a_{22} + a_{33} = \mathrm{Tr}\, A.$$

This means that the sum of the eigenvalues is equal to the trace of A. This is a very useful relation to check if the eigenvalues are calculated correctly. Furthermore

$$\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix},$$

which is the sum of principal minors (minors of the diagonal elements), and

$$\lambda_1\lambda_2\lambda_3 = |A|\,.$$

That the product of all eigenvalues is equal to the determinant of A is also a very useful relation. If A is singular $|A| = 0$, at least one of the eigenvalue must be zero. It follows that the inverse of A exists if and only if none of the eigenvalues of A is zero.

Similar calculations can generalize these relationships for matrices of higher orders.

Example 6.1.5. Find the eigenvalues and the corresponding eigenvectors of the matrix A if

$$A = \begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix}.$$

Solution 6.1.5.

$$\begin{aligned} P(\lambda) &= \begin{pmatrix} 5-\lambda & 7 & -5 \\ 0 & 4-\lambda & -1 \\ 2 & 8 & -3-\lambda \end{pmatrix} \\ &= \begin{vmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{vmatrix} - \left(\begin{vmatrix} 4 & -1 \\ 8 & -3 \end{vmatrix} + \begin{vmatrix} 5 & -5 \\ 2 & -3 \end{vmatrix} + \begin{vmatrix} 5 & 7 \\ 0 & 4 \end{vmatrix} \right) \lambda + (5+4-3)\,\lambda^2 - \lambda^3 \\ &= 6 - 11\lambda + 6\lambda^2 - \lambda^3 = (1-\lambda)(2-\lambda)\,(3-\lambda) = 0. \end{aligned}$$

Thus the three eigenvalues are

$$\lambda_1 = 1, \quad \lambda_2 = 2, \quad \lambda_3 = 3.$$

As a check, the sum of the eigenvalues

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 + 2 + 3 = 6$$

is indeed equal to the trace of A

$$\operatorname{Tr} A = 5 + 4 - 3 = 6.$$

Furthermore, the product of three eigenvalues

$$\lambda_1\lambda_2\lambda_3 = 6$$

is indeed equal to the determinant

$$\begin{vmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{vmatrix} = 6.$$

Let the eigenvector $\mathbf{x}_1$ corresponding to λ_1 be $\begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix}$, then

$$\begin{pmatrix} 5-\lambda_1 & 7 & -5 \\ 0 & 4-\lambda_1 & -1 \\ 2 & 8 & -3-\lambda_1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix} = \begin{pmatrix} 4 & 7 & -5 \\ 0 & 3 & -1 \\ 2 & 8 & -4 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix} = 0.$$

By Gauss' elimination method, one can readily show that

$$\begin{pmatrix} 4 & 7 & -5 \\ 0 & 3 & -1 \\ 2 & 8 & -4 \end{pmatrix} \Longrightarrow \begin{pmatrix} 4 & 7 & -5 \\ 0 & 3 & -1 \\ 0 & 4.5 & -1.5 \end{pmatrix} \Longrightarrow \begin{pmatrix} 4 & 7 & -5 \\ 0 & 3 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus the set of equations is reduced to

$$\begin{aligned} 4x_{11} + 7x_{12} - 5x_{13} &= 0, \\ 3x_{12} - x_{13} &= 0. \end{aligned}$$

Only one of the three unknowns can be assigned arbitrary. For example, let $x_{13} = 3$, then $x_{12} = 1$ and $x_{11} = 2$. Therefore corresponding to the eigenvalue $\lambda_1 = 1$, the eigenvector can be written as

$$\mathbf{x}_1 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

Similarly, corresponding to $\lambda_2 = 2$ and $\lambda_3 = 3$, the respective eigenvectors are

$$\mathbf{x}_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}.$$

6.1.3 Properties of Eigenvalues

There are other properties related to eigenvalue problems. Taken individually, they are almost self-evident, but collectively they are useful in matrix applications.

- The transpose $\widetilde{A}\left(A^{\mathrm{T}}\right)$ has the same eigenvalues as A.

The eigenvalues of A and A^{T} are, respectively, the solutions of $|A - \lambda I| = 0$ and $\left|A^{\mathrm{T}} - \lambda I\right| = 0$. Since $A^{\mathrm{T}} - \lambda I = (A - \lambda I)^{\mathrm{T}}$ and the determinant of a matrix is equal to the determinant of its transpose

$$|A - \lambda I| = \left|(A - \lambda I)^{\mathrm{T}}\right| = \left|A^{\mathrm{T}} - \lambda I\right|,$$

the secular equations of A and A^{T} are identical. Therefore they have the same set of eigenvalues.

- If A is either upper or lower triangular, then the eigenvalues are the diagonal elements.

  Let $|A - \lambda I| = 0$ be

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ 0 & a_{22}-\lambda & \dots & a_{2n} \\ 0 & 0 & \vdots & \vdots \\ 0 & 0 & 0 & a_{nn}-\lambda \end{vmatrix} = (a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda) = 0.$$

Clearly $\lambda = a_{11}, \lambda = a_{22}, \dots, \lambda = a_{nn}$.

- If $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ are the eigenvalues of A, then the eigenvalues of the inverse A^{-1} are $1/\lambda_1, 1/\lambda_2, 1/\lambda_3, \dots, 1/\lambda_n$.

  Multiplying the equation $A\mathbf{x} = \lambda\mathbf{x}$ from the left by A^{-1}

$$A^{-1}A\mathbf{x} = A^{-1}\lambda\mathbf{x} = \lambda A^{-1}\mathbf{x}$$

and using $A^{-1}A\mathbf{x} = I\mathbf{x} = \mathbf{x}$, we have $\mathbf{x} = \lambda A^{-1}\mathbf{x}$. Thus

$$A^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}.$$

- If $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ are the eigenvalues of A, then the eigenvalues of A^m are $\lambda_1^m, \lambda_2^m, \lambda_3^m, \dots, \lambda_n^m$.

  Since $A\mathbf{x} = \lambda\mathbf{x}$, it follows:

$$A^2\mathbf{x} = A(A\mathbf{x}) = A\lambda\mathbf{x} = \lambda A\mathbf{x} = \lambda^2\mathbf{x}.$$

Similarly

$$A^3\mathbf{x} = \lambda^3\mathbf{x}, \dots, A^m\mathbf{x} = \lambda^m\mathbf{x}.$$

6.2 Some Terminology

As we have seen that for a $n \times n$ square matrix, the eigenvalues may or may not be real numbers. If the eigenvalues are degenerate, we may or may not have n distinct eigenvectors.

However, there is a class of matrices, known as *hermitian matrices*, the eigenvalues of which are always real. A $n \times n$ hermitian matrix will always have n distinct eigenvectors.

To facilitate the discussion of these and other properties of matrices, we will first introduce the following terminology.

6.2.1 Hermitian Conjugation

Complex Conjugation

If $A = (a_{ij})_{m\times n}$ is an arbitrary matrix whose elements may be complex numbers, the complex conjugate matrix denoted by A^* is also a matrix of order $m \times n$ with every element of which is the complex conjugate of the corresponding element of A, i.e.,

$$(A^*)_{ij} = a^*_{ij}.$$

It is clear that

$$(cA)^* = c^* A^*.$$

Hermitian Conjugation

When the two operations of complex conjugation and transposition are carried out one after another on a matrix, the resulting matrix is called the *hermitian conjugate* of the original matrix and is denoted by $A^\dagger$, called A dagger. Mathematicians also refer to $A^\dagger$ as the adjoint matrix. The order of the two operation is immaterial. Thus

$$A^\dagger = (A^*)^{\mathrm{T}} = (\widetilde{A})^*. \tag{6.6}$$

For example, if

$$A = \begin{pmatrix} (6+\mathrm{i}) & (1-6\mathrm{i}) & 1 \\ (3+\mathrm{i}) & 4 & 3\mathrm{i} \end{pmatrix}, \tag{6.7}$$

then

$$A^\dagger = (A^*)^{\mathrm{T}} = \begin{pmatrix} (6-\mathrm{i}) & (1+6\mathrm{i}) & 1 \\ (3-\mathrm{i}) & 4 & -3\mathrm{i} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} (6-\mathrm{i}) & (3-\mathrm{i}) \\ (1+6\mathrm{i}) & 4 \\ 1 & -3\mathrm{i} \end{pmatrix}, \tag{6.8}$$

$$A^\dagger = (\widetilde{A})^* = \begin{pmatrix} (6+\mathrm{i}) & (3+\mathrm{i}) \\ (1-6\mathrm{i}) & 4 \\ 1 & 3\mathrm{i} \end{pmatrix}^* = \begin{pmatrix} (6-\mathrm{i}) & (3-\mathrm{i}) \\ (1+6\mathrm{i}) & 4 \\ 1 & -3\mathrm{i} \end{pmatrix}. \tag{6.9}$$

Hermitian Conjugate of Matrix Products

We have shown in Chap. 5 that the transpose of the product of two matrices is equal to the product of the transposed matrices taken in reverse order. This leads directly to the fact that

$$(AB)^\dagger = B^\dagger A^\dagger,$$

since

$$(AB)^\dagger = (A^* B^*)^{\mathrm{T}} = \widetilde{B}^* \widetilde{A}^* = B^\dagger A^\dagger. \tag{6.10}$$

6.2.2 Orthogonality

Inner Product

If **a** and **b** are two column vectors of the same order n, the inner product (or *scalar product*) is defined as $\mathbf{a}^\dagger \mathbf{b}$. The hermitian conjugate of the column vector is a row vector

$$\mathbf{a}^\dagger = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}^\dagger = \left(a_1^* \; a_2^* \cdots a_n^* \right),$$

therefore the inner product is one number

$$\mathbf{a}^\dagger \mathbf{b} = \left(a_1^* \; a_2^* \cdots a_n^* \right) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \sum_{k=1}^{n} a_k^* b_k.$$

There are two other commonly used notations for the inner product.

The notation most often used in quantum mechanics is the bracket notation of Dirac. The row and column vectors are, respectively, defined as the bra and ket vectors. Thus we may write the column vector

$$\mathbf{b} = |\, \mathbf{b}\rangle$$

as the ket vector, and the row vector

$$\mathbf{a}^\dagger = \langle \mathbf{a}|$$

as the bra vector. The inner product of two vectors is then represented by

$$\langle \mathbf{a}|\mathbf{b}\rangle = \mathbf{a}^\dagger \mathbf{b}.$$

Notice that for any scalar c,

$$\langle \mathbf{a}|c\mathbf{b}\rangle = c \langle \mathbf{a}|\mathbf{b}\rangle,$$

whereas

$$\langle c\mathbf{a}|\mathbf{b}\rangle = c^* \langle \mathbf{a}|\mathbf{b}\rangle.$$

Another notation that is often used is the parenthesis notation:

$$(\mathbf{a}, \mathbf{b}) = \mathbf{a}^\dagger \mathbf{b} = \langle \mathbf{a}|\mathbf{b}\rangle.$$

If A is a matrix, then

$$(\mathbf{a}, A\mathbf{b}) = \left(A^\dagger \mathbf{a}, \mathbf{b} \right)$$

is an identity, since

$$\left(A^{\dagger}\mathbf{a},\mathbf{b}\right) = \left(A^{\dagger}\mathbf{a}\right)^{\dagger}\mathbf{b} = \mathbf{a}^{\dagger}\left(A^{\dagger}\right)^{\dagger}\mathbf{b} = \mathbf{a}^{\dagger}A\mathbf{b} = (\mathbf{a},A\mathbf{b})\,.$$

Thus if

$$(\mathbf{a},A\mathbf{b}) = (A\mathbf{a},\mathbf{b})\,,$$

then A is hermitian. Mathematicians refer to the relation $A^{\dagger} = A$ as *self-adjoint*.

Orthogonality

Two vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be orthogonal if and only if

$$\mathbf{a}^{\dagger}\mathbf{b} = 0.$$

Note that in three-dimensional real space

$$\mathbf{a}^{\dagger}\mathbf{b} = \sum_{k=1}^{n} a_k^* b_k^* = a_1b_1 + a_2b_2 + a_3b_3$$

is just the dot product of $\mathbf{a}$ and $\mathbf{b}$. It is well known in vector analysis that if the dot product of two vectors is equal to zero, then they are perpendicular.

Length of a Complex Vector

If we adopt this definition of the scalar product of two complex vectors, then we have a natural definition of the length of a complex vector in a n-dimensional space. The length $\|\mathbf{x}\|$ of a complex vector $\mathbf{x}$ is taken to be

$$\|\mathbf{x}\|^2 = \mathbf{x}^{\dagger}\mathbf{x} = \sum_{k=1}^{n} a_k^* a_k = \sum_{k=1}^{n} |a_k|^2\,.$$

6.2.3 Gram–Schmidt Process

Linear Independence

The set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is linearly independent if and only it

$$\sum_{i=1}^{n} a_i \mathbf{x}_i = 0$$

implies every $a_i = 0$. Otherwise the set is linearly dependent.

Let us test the three vectors

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

for linear independence. The question is if we can find a set of a_i, not all zero such that

$$\sum_{i=1}^{3} a_i \mathbf{x}_i = a_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + a_3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a_1 + a_3 \\ a_2 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Clearly this requires $a_1 = 0$, $a_2 = 0$, and $a_3 = 0$. Therefore these three vectors are linearly independent.

Note that linear independence or dependence is a property of the set as a whole, not of the individual vectors.

It is obvious that if $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ represent three noncoplanar three-dimensional vectors, they are linearly independent.

Gram–Schmidt Process

Given any n linearly independent vectors, one can construct from their linear combinations a set of n mutually orthogonal unit vectors.

Let the given linearly independent vectors be $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Define

$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

to be the first unit vector. Now define

$$\mathbf{u}_2' = \mathbf{x}_2 - (\mathbf{x}_2, \mathbf{u}_1)\, \mathbf{u}_1.$$

The inner product of $\mathbf{u}_2'$ and $\mathbf{u}_1$ is equal to zero,

$$(\mathbf{u}_2', \mathbf{u}_1) = (\mathbf{x}_2, \mathbf{u}_1) - (\mathbf{x}_2, \mathbf{u}_1)\,(\mathbf{u}_1, \mathbf{u}_1) = 0,$$

since $(\mathbf{u}_1, \mathbf{u}_1) = 1$. This shows $\mathbf{u}_2'$ is orthogonal to $\mathbf{u}_1$.

We can normalize $\mathbf{u}_2'$

$$\mathbf{u}_2 = \frac{\mathbf{u}_2'}{\|\mathbf{u}_2'\|}$$

to obtain the second unit vector $\mathbf{u}_2$ which is orthogonal to $\mathbf{u}_1$.

We can continue this process by defining

$$\mathbf{u}_k' = \mathbf{x}_k - \sum_{i=1}^{k-1} (\mathbf{x}_k, \mathbf{u}_i)\, \mathbf{u}_i,$$

and

$$\mathbf{u}_k = \frac{\mathbf{u}'_k}{\|\mathbf{u}'_k\|}.$$

When all $\mathbf{x}_k$ are used up, we will have n unit vectors $u_1, u_2, \ldots, u_k$ orthogonal to each other. They are called an *orthonormal set*. This procedure is known as Gram–Schmidt process.

6.3 Unitary Matrix and Orthogonal Matrix

6.3.1 Unitary Matrix

If the square matrix U satisfies the condition

$$U^\dagger U = I,$$

then U is called unitary. The n columns in a unitary matrix can be considered as n column vectors in an orthonormal set.

In other words, if

$$\mathbf{u}_1 = \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1n} \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} u_{21} \\ u_{22} \\ \vdots \\ u_{2n} \end{pmatrix}, \ldots, \mathbf{u}_n = \begin{pmatrix} u_{n1} \\ u_{n2} \\ \vdots \\ u_{nn} \end{pmatrix},$$

and

$$\mathbf{u}_i^\dagger \mathbf{u}_j = \begin{pmatrix} u_{i1}^* & u_{i2}^* & \cdots & u_{in}^* \end{pmatrix} \begin{pmatrix} u_{j1} \\ u_{j2} \\ \vdots \\ u_{jn} \end{pmatrix} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

then

$$U = \begin{pmatrix} u_{11} & u_{21} & \ldots & u_{n1} \\ u_{12} & u_{22} & \ldots & u_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ u_{1n} & u_{n2} & \ldots & u_{nn} \end{pmatrix}$$

is unitary. This is because

$$U^\dagger = \begin{pmatrix} u_{11}^* & u_{12}^* & \ldots & u_{1n}^* \\ u_{21}^* & u_{22}^* & \ldots & u_{2n}^* \\ \vdots & \vdots & \ldots & \vdots \\ u_{n1}^* & u_{n2}^* & \ldots & u_{nn}^* \end{pmatrix}$$

therefore

$$U^\dagger U = \begin{pmatrix} u_{11}^* & u_{12}^* & \dots & u_{1n}^* \\ u_{21}^* & u_{22}^* & \dots & u_{2n}^* \\ \vdots & \vdots & \vdots & \vdots \\ u_{n1}^* & u_{n2}^* & \dots & u_{nn}^* \end{pmatrix} \begin{pmatrix} u_{11} & u_{21} & \dots & u_{n1} \\ u_{12} & u_{22} & \dots & u_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ u_{1n} & u_{2n} & \dots & u_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Multiply U^{-1} from the right, we have

$$U^\dagger U U^{-1} = I U^{-1}.$$

It follows that hermitian conjugate of a unitary matrix is its inverse, i.e.,

$$U^\dagger = U^{-1}.$$

6.3.2 Properties of Unitary Matrix

- Unitary transformations leave lengths of vectors invariant.

Let

$$\mathbf{a} = U\mathbf{b}, \quad \text{so} \quad \mathbf{a}^\dagger = \mathbf{b}^\dagger U^\dagger,$$

and

$$\|\mathbf{a}\|^2 = \mathbf{a}^\dagger \mathbf{a} = \mathbf{b}^\dagger U^\dagger U \mathbf{b}.$$

Since

$$U^\dagger U = U^{-1} U = I,$$

it follows:

$$\|\mathbf{a}\|^2 = \mathbf{a}^\dagger \mathbf{a} = \mathbf{b}^\dagger \mathbf{b} = \|\mathbf{b}\|^2 .$$

Thus the length of the initial vector is equal to the length of the transformed vector.

- The absolute value of the eigenvalues of an unitary matrix is equal to one.

Let $\mathbf{x}$ be a nontrivial eigenvector of the unitary matrix U belonging to the eigenvalue λ

$$U\mathbf{x} = \lambda \mathbf{x}.$$

Take the hermitian conjugate of both sides

$$\mathbf{x}^\dagger U^\dagger = \lambda^* \mathbf{x}^\dagger.$$

Multiply the last two equations

$$\mathbf{x}^{\dagger}U^{\dagger}U\mathbf{x} = \lambda^{*}\mathbf{x}^{\dagger}\lambda\mathbf{x}.$$

Since $U^{\dagger}U = I$ and $\lambda^{*}\lambda = |\lambda|^{2}$, it follows:

$$\mathbf{x}^{\dagger}\mathbf{x} = |\lambda|^{2}\,\mathbf{x}^{\dagger}\mathbf{x}.$$

Therefore

$$|\lambda|^{2} = 1.$$

In other words, the eigenvalues of a unitary matrix must be on the unit circle in the complex plane centered at the origin.

6.3.3 Orthogonal Matrix

If the elements of an unitary matrix are all real, the matrix is known as an *orthogonal matrix*. Thus the properties of unitary matrices are also properties of orthogonal matrices. In addition,

- The determinant of an orthogonal matrix is equal to either positive one or negative one.

  If A is a real square matrix, then by definition

$$A^{\dagger} = \widetilde{A}^{*} = \widetilde{A}.$$

If, in addition, A is unitary, $A^{\dagger} = A^{-1}$, then

$$\widetilde{A} = A^{-1}.$$

Thus

$$A\widetilde{A} = I. \tag{6.11}$$

Since the determinant of A is equal to the determinant of $\widetilde{A}$, so

$$|A\widetilde{A}| = |A||\widetilde{A}| = |A|^{2}.$$

But

$$|A\widetilde{A}| = |I| = 1,$$

therefore

$$|A|^{2} = 1.$$

Thus, the determinant of an orthogonal matrix is either $+1$ or -1.

Very often (6.11) is used as the definition of an orthogonal matrix. That is, a square real matrix A satisfying the relation expressed in (6.11) is called an orthogonal matrix. This is equivalent to the statement "that the inverse of an orthogonal matrix is equal to its transpose."

Written in terms of its elements, (6.11) is given by

$$\sum_{j=1} a_{ij}\widetilde{a}_{jk} = \sum_{j=1} a_{ij}a_{kj} = \delta_{ik} \tag{6.12}$$

for any i and any j. Similarly, $\widetilde{A}A = I$ can be expressed as

$$\sum_{j=1} \widetilde{a}_{ij}a_{jk} = \sum_{j=1} a_{ji}a_{jk} = \delta_{ik}. \tag{6.13}$$

However, (6.13) is not independent of (6.12), since $A\widetilde{A} = \widetilde{A}A$. If one set of conditions is valid, the other set must also be valid.

Put in words, these conditions mean that the sum of the products of the corresponding elements of two distinct columns (or rows) of an orthogonal matrix is zero, while the sum of the squares of the elements of any column (or row) is equal to unity. If we regard the n columns of the matrix as n real vectors, this means that these n column vectors are orthogonal and normalized. Similarly, all the rows of an orthogonal matrix are orthonormal.

6.3.4 Independent Elements of an Orthogonal Matrix

An nth order square matrix has n^2 elements. For an orthogonal matrix, not all these elements are independent of each other, because there are certain conditions they must satisfy. First, there are n conditions for each column to be normalized. Then there are $n(n-1)/2$ conditions for each column to be orthogonal to any other column. Therefore the number of independent parameters of an orthogonal matrix is

$$n^2 - [n + n(n-1)/2] = n(n-1)/2.$$

In other words, an nth order orthogonal matrix can be fully characterized by $n(n-1)/2$ independent parameters.

For $n = 2$, the number of independent parameters is one. This is illustrated as follows.

Consider an arbitrary orthogonal matrix of order 2

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

The fact that each column is normalized leads to

$$a^2 + b^2 = 1, \tag{6.14}$$

$$c^2 + d^2 = 1. \tag{6.15}$$

Furthermore, the two columns are orthogonal

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = ac + bd = 0. \tag{6.16}$$

The general solution of (6.14) is $a = \cos\theta$, $b = \sin\theta$, where θ is a scalar. Similarly, (6.15) can be satisfied, if we choose $c = \cos\phi$, $d = \sin\phi$, where ϕ is another scalar. On the other hand, (6.16) requires

$$\cos\theta\cos\phi + \sin\theta\sin\phi = \cos(\theta - \phi) = 0,$$

therefore

$$\phi = \theta \pm \frac{\pi}{2}.$$

Thus, the most general orthogonal matrix of order 2 is

$$A_1 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \quad A_2 = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}. \tag{6.17}$$

Every orthogonal matrix of order 2 can be expressed in this form with some value of θ. Clearly the determinant of A_1 is equal to $+1$ and that of A_2, -1.

6.3.5 Orthogonal Transformation and Rotation Matrix

The fact that in real space, orthogonal transformation preserving the length of a vector suggests that the orthogonal matrix is associated with rotation of vectors. In fact the orthogonal matrix is related to two kinds of rotations in space. First it can be thought as an operator which rotates a vector. This is often called *active transformation.* Secondly, it can be thought as the transformation matrix when the coordinate axes of the reference system are rotated. This is also referred as *passive transformation.*

First let us consider the vectors shown in Fig. 6.1a. The x and y components of the vector $\mathbf{r}_1$ are given by $x_1 = r\cos\varphi$ and $y_1 = r\sin\varphi$, where r is the length of the vector. Now let us rotate the vector counterclockwise by an angle θ, so that $x_2 = r\cos(\varphi + \theta)$ and $y_2 = r\sin(\varphi + \theta)$. Using trigonometry, we can write

$$x_2 = r\cos(\varphi + \theta) = r\cos\varphi\cos\theta - r\sin\varphi\sin\theta = x_1\cos\theta - y_1\sin\theta,$$
$$y_2 = r\sin(\varphi + \theta) = r\sin\varphi\cos\theta + r\cos\varphi\sin\theta = y_1\cos\theta + x_1\sin\theta.$$

We can display the set of coefficients in the form of

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

It is seen that the coefficient matrix is the orthogonal matrix A_1 of (6.17). Therefore the orthogonal matrix with determinant equal to $+1$ is also called

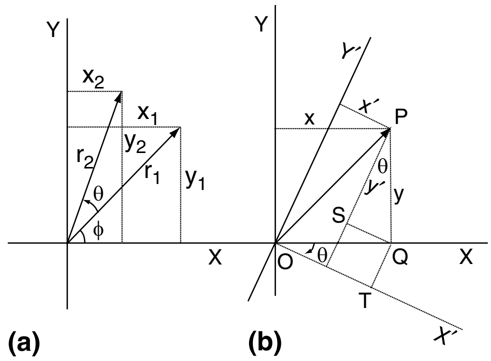

Fig. 6.1. Two interpretations of the orthogonal matrix A_1 whose determinant is equal to $+1$. (**a**) As a operator, it rotates the vector $\mathbf{r}_1$ to $\mathbf{r}_2$ without changing its length. (**b**) As the transformation matrix between the coordinates of the tip of a fixed vector when the coordinate axes are rotated. Note that the rotation in (**b**) is in the opposite direction as in (**a**)

rotation matrix. It rotates the vector from $\mathbf{r}_1$ to $\mathbf{r}_2$ without changing the magnitude.

The second interpretation of rotation matrix is as follows. Let P be the tip of a fixed vector. The coordinates of P are (x, y) in a particular rectangular coordinate system. Now the coordinate axes are rotated clockwise by an angle θ as shown in Fig. 6.1b. The coordinates of P in the rotated system become (x', y'). From the geometry in Fig. 6.1b, it is clear that

$$\begin{aligned} x' &= OT - SQ = OQ\cos\theta - PQ\sin\theta = x\cos\theta - y\sin\theta, \\ y' &= QT + PS = OQ\sin\theta + PQ\cos\theta = x\sin\theta + y\cos\theta, \end{aligned}$$

or

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Note that the matrix involved is again the orthogonal matrix A_1. However, this time A_1 is the transformation matrix between the coordinates of the tip of a fixed vector when the coordinate axes are rotated.

The equivalence between these two interpretations might be expected, since the relative orientation between the vector and coordinate axes is the same whether the vector is rotated counterclockwise by an angle θ, or the coordinate axes are rotated clockwise by the same angle.

Next, let us consider the orthogonal matrix A_2, the determinant of which is equal to -1. The matrix A_2 can be expressed as

$$A_2 = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

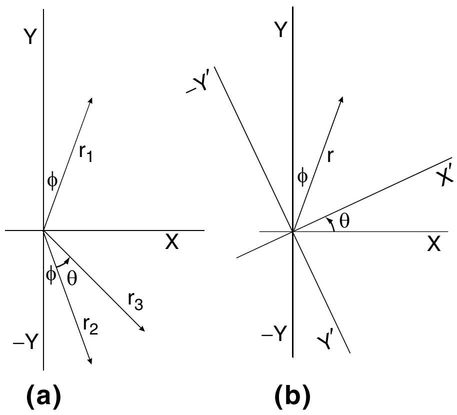

Fig. 6.2. Two interpretations of the orthogonal matrix A_2 whose determinant is -1. (**a**) As an operator, it flips the vector $\mathbf{r}_1$ to $\mathbf{r}_2$ symmetrically with respect to X-axis, and then rotates $\mathbf{r}_2$ to $\mathbf{r}_3$. (**b**) As the transformation matrix between the tip of a fixed vector when the Y-axis is inverted and then the coordinate axes are rotated. Note that the rotation in (**b**) seems to be in the same direction as in (**a**)

The transformation

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

gives

$$x_2 = x_1, \quad y_2 = -y_1.$$

Clearly this corresponds to the reflection of the vector with respect to the X-axis. Therefore A_2 can be considered as an operator which first symmetrically flips the vector $\mathbf{r}_1$ over the X-axis and then rotates it to $\mathbf{r}_3$ as shown in Fig. 6.2a.

In terms of coordinate transformation, one can show that (x', y') in the equation

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

represents the new coordinates of the tip of a fixed vector after the Y-axis is inverted and the whole coordinate axes are rotated by an angle θ, as shown in Fig. 6.2b. In this case one has to be careful about the sign of the angle. The sign convention is that a counterclockwise rotation is positive and a clockwise rotation is negative. However, after the Y-axis is inverted as in Fig. 6.2b, a negative rotation (rotating from the direction of the positive X-axis toward the negative of the Y-axis) appears to be counterclockwise. This is why in Fig. 6.1a, b, the vector and the coordinate axes are rotating in the opposite direction, whereas in Fig. 6.2a, b, they seem to rotate in the same direction.

So far we have used rotations in two dimensions as examples. However, the conclusions that orthogonal matrix whose determinant equals to $+1$ represents pure rotation, and orthogonal matrix whose determinant is equal to -1

represents a reflection followed by a rotations are generally valid in higher-dimensional space. In the chapter on vector transformation, we will have a more detailed discussion.

6.4 Diagonalization

6.4.1 Similarity Transformation

If A is a $n \times n$ matrix and $\mathbf{u}$ is a $n \times 1$ column vector, then $A\mathbf{u}$ is another $n \times 1$ column vector. The equation

$$A\mathbf{u} = \mathbf{v} \tag{6.18}$$

represents a linear transformation. Matrix A acts as a linear operator sending vector $\mathbf{u}$ to vector $\mathbf{v}$. Let

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

where u_i and v_i are, respectively, the ith components of $\mathbf{u}$ and $\mathbf{v}$ in the n-dimensional space. These components are, of course, measured in a certain coordinate system (reference frame). Let the unit vectors, $\mathbf{e}_i$, known as bases, along the coordinate axes of this system be

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

then

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = u_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + u_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{n} u_i \mathbf{e}_i. \tag{6.19}$$

Suppose there is another coordinate system, which we designate as the prime system. Measured in that system, the components of $\mathbf{u}$ and $\mathbf{v}$ become

$$\begin{pmatrix} u_1' \\ u_2' \\ \vdots \\ u_n' \end{pmatrix} = \mathbf{u}', \quad \begin{pmatrix} v_1' \\ v_2' \\ \vdots \\ v_n' \end{pmatrix} = \mathbf{v}'. \tag{6.20}$$

We emphasize that $\mathbf{u}$ and $\mathbf{u}'$ are the same vector except measured in two different coordinate systems. The symbol $\mathbf{u}'$ does not mean a vector different from $\mathbf{u}$, it simply represents the collection of components of $\mathbf{u}$ in the prime system as shown in (6.20). Similarly, $\mathbf{v}$ and $\mathbf{v}'$ are the same vectors. We can find these components if we know the components of $\mathbf{e}_i$ in the prime system.

In (6.19)

$$\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + \cdots + u_n\mathbf{e}_n,$$

the u_i' are just numbers which are independent of the coordinate system. To find the components of $\mathbf{u}$ in the prime system, we only need to express $\mathbf{e}_i$ in the prime system.

Let $\mathbf{e}_i$ measured in the prime system be

$$\mathbf{e}_1 = \begin{pmatrix} s_{11} \\ s_{21} \\ \vdots \\ s_{n1} \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} s_{12} \\ s_{22} \\ \vdots \\ s_{n2} \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} s_{1n} \\ s_{2n} \\ \vdots \\ s_{nn} \end{pmatrix},$$

then the components of $\mathbf{u}$ measured in the prime system can be written as

$$\begin{pmatrix} u_1' \\ u_2' \\ \vdots \\ u_n' \end{pmatrix} = u_1 \begin{pmatrix} s_{11} \\ s_{21} \\ \vdots \\ s_{n1} \end{pmatrix} + u_2 \begin{pmatrix} s_{12} \\ s_{22} \\ \vdots \\ s_{n2} \end{pmatrix} + \cdots + u_n \begin{pmatrix} s_{1n} \\ s_{2n} \\ \vdots \\ s_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} u_1 s_{11} + u_2 s_{12} + \cdots + u_n s_{1n} \\ u_1 s_{21} + u_2 s_{22} + \cdots + u_n s_{2n} \\ \vdots \\ u_1 s_{n1} + u_2 s_{n2} + \cdots + u_n s_{nn} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1n} \\ s_{21} & s_{22} & \ldots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

This equation can be written in the form

$$\mathbf{u}' = T\mathbf{u}, \tag{6.21}$$

where

$$T = \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1n} \\ s_{21} & s_{22} & \ldots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{pmatrix}.$$

It is clear from this analysis that the transformation matrix between the vector components in two coordinate systems is the same for all vectors, since it depends only on the transformation of the basis vectors in the two reference frames. Therefore $\mathbf{v}'$ and $\mathbf{v}$ are also related to the same transformation matrix T,

$$\mathbf{v}' = T\mathbf{v}. \tag{6.22}$$

The operation of sending $\mathbf{u}$ to $\mathbf{v}$, expressed in the original system is given by $A\mathbf{u} = \mathbf{v}$. Let the same operation expressed in the prime system be

$$A'\mathbf{u}' = \mathbf{v}'.$$

Since $\mathbf{u}' = T\mathbf{u}$ and $\mathbf{v}' = T\mathbf{v}$,

$$A'T\mathbf{u} = T\mathbf{v}.$$

Multiply both sides by the inverse of T from the left,

$$T^{-1}A'T\mathbf{u} = T^{-1}T\mathbf{v} = \mathbf{v}.$$

Since $A\mathbf{u} = \mathbf{v}$, if follows that:

$$A = T^{-1}A'T. \tag{6.23}$$

If we multiply this equation by T from the left and by T^{-1} from the right, we have

$$TAT^{-1} = A'.$$

What we have found is that as long as we know the relationship between the coordinate axes of two reference frames, not only we can transform a vector from one reference frame to the other, but we can also transform a matrix representing a linear operator from one reference frame to the other.

In general, if there exits a nonsingular matrix T such that $T^{-1}AT = B$ for any two square matrices A and B of the same order, then A and B are called *similar matrices*, and the transformation from A to B is called *similarity transformation*.

If two matrices are related by a similarity transformation, then they represent the same linear transformation in two different coordinate systems.

If the rectangular coordinate axes in the prime system are generated through a rotation from the original system, then T is an orthogonal matrix as discussed in the Sect. 6.3. In that case $T^{-1} = \widetilde{T}$, and the similarity transformation can be written as $\widetilde{T}AT$. If we are working in the complex space, the transformation matrix is unitary, and the similarity transformation can be written as $T^{\dagger}AT$. Both of these transformations are known as unitary similarity transformation.

A matrix that can be brought to diagonal form by a similarity transformation is said to be *diagonalizable*. Whether a matrix is diagonalizable and how to diagonalize it are very important questions in the theory of linear transformation. Not only because it is much more convenient to work with diagonal matrix, but also because it is of fundamental importance in the structure of quantum mechanics. In the following sections, we will answer these questions.

6.4.2 Diagonalizing a Square Matrix

The eigenvectors of the matrix A can be used to form another matrix S in such a way that $S^{-1}AS$ becomes a diagonal matrix. This process often greatly simplifies a physical problem by a better choice of variables.

If A is a square matrix of order n, the eigenvalues λ_i and eigenvectors $\mathbf{x}_i$ satisfy the equation

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i \tag{6.24}$$

for $i = 1, 2, \ldots, n$. Each eigenvector is a column matrix with n elements

$$\mathbf{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{pmatrix}, \ldots, \mathbf{x}_n = \begin{pmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{nn} \end{pmatrix}.$$

Each of the n equations of (6.24) is of the form

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix} = \begin{pmatrix} \lambda_i x_{i1} \\ \lambda_i x_{i2} \\ \vdots \\ \lambda_i x_{in} \end{pmatrix}. \tag{6.25}$$

Collectively they can be written as

$$\begin{aligned} \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \ldots & x_{nn} \end{pmatrix} &= \begin{pmatrix} \lambda_1 x_{11} & \lambda_2 x_{21} & \ldots & \lambda_n x_{n1} \\ \lambda_1 x_{12} & \lambda_2 x_{22} & \ldots & \lambda_n x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1 x_{1n} & \lambda_2 x_{2n} & \ldots & \lambda_n x_{nn} \end{pmatrix} \\ &= \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \ldots & x_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}. \end{aligned} \tag{6.26}$$

To simplify the writing, let

$$S = \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \ldots & x_{nn} \end{pmatrix}, \tag{6.27}$$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}, \tag{6.28}$$

and write (6.26) as

$$AS = S\Lambda. \tag{6.29}$$

Multiplying both sides of this equation by S^{-1} from the left, we obtain

$$S^{-1}AS = \Lambda. \tag{6.30}$$

Thus, by using the matrix of eigenvectors and its inverse, it is possible to transform a matrix A to a diagonal matrix whose elements are the eigenvalues of A. The transformation expressed by (6.30) is referred to as the *diagonalization* of matrix A.

Example 6.4.1. If $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, find S such that $S^{-1}AS$ is a diagonal matrix. Show that the elements of $S^{-1}AS$ are the eigenvalues of A.

Solution 6.4.1. Since the secular equation is

$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = (\lambda+1)(\lambda-3) = 0,$$

the eigenvalues are $\lambda_1 = -1$, $\lambda_2 = 3$. The eigenvectors are, respectively, $\mathbf{x}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\mathbf{x}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Therefore

$$S = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

It can be readily checked that $S^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and

$$S^{-1}AS = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Note that the diagonalizing matrix S is not necessary unitary. However, if the eigenvectors are orthogonal, then we can normalize the eigenvectors and form an orthonormal set. The matrix with members of this orthonormal set as columns is a unitary matrix. The diagonalization process becomes a unitary similarity transformation which is much more convenient and useful.

The two eigenvectors in the above example are orthogonal, since

$$\begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0.$$

Normalizing them, we get

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The matrix constructed with these two normalized eigenvectors is

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

which is an orthogonal matrix. The transformation

$$\widetilde{U} A U = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}$$

is a unitary similarity transformation.

First we have eliminated the step of finding the inverse of U, since U is an orthogonal matrix, the inverse of U is simply its transpose. More importantly, U represents a rotation as discuss in Sect. 6.3. If we rotate the two original coordinate axes to coincide with $\mathbf{u}_1$ and $\mathbf{u}_2$, then with respect to this rotated axes, A is diagonal.

The coordinate axes of a reference system, in which the matrix is diagonal, are known as *principal axes*. In this example, $\mathbf{u}_1$ and $\mathbf{u}_2$ are the unit vectors along the principal axes. From the components of $\mathbf{u}_1$, we can easily find the orientation of the principal axes. Let θ_1 be the angle between $\mathbf{u}_1$ and the original horizontal axis, then

$$\mathbf{u}_1 = \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

which gives $\theta_1 = -\pi/4$. This means that to get the principal axes, we have to rotate the original coordinate axes 45° clockwise. For consistence check, we can calculate θ_2, the angle between $\mathbf{u}_2$ and the original horizontal axis. Since

$$\mathbf{u}_2 = \begin{pmatrix} \cos\theta_2 \\ \sin\theta_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$\theta_2 = +\pi/4$. Therefore the angle between θ_1 and θ_2 is $\pi/2$. This shows that $\mathbf{u}_1$ and $\mathbf{u}_2$ are perpendicular to each other, as they must.

Since $\theta_2 = \pi/2 + \theta_1$, $\cos\theta_2 = -\sin\theta_1$, and $\sin\theta_2 = \cos\theta_1$, the unitary matrix U can be written as

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \sin\theta_2 \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix},$$

which, as seen in (6.17), is indeed a rotation matrix.

6.4.3 Quadratic Forms

A quadratic form is a homogeneous second degree expression in n variables. For example,

$$Q(x_1, x_2) = 5x_1^2 - 4x_1x_2 + 2x_2^2$$

is a quadratic form in x_1 and x_2. Through a change of variables, this expression can be transformed into a form in which there is no crossproduct term. Such a form is known as *canonical form*. Quadratic forms are important because they occur in a wide variety of applications.

The first step to change it into a canonical form is to separate the crossproduct term into two equal terms, $(4x_1x_2 = 2x_1x_2 + 2x_2x_1)$, so that $Q(x_1, x_2)$ can be written as

$$Q(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{6.31}$$

where the coefficient matrix

$$C = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}$$

is symmetric. As we shall see in Sect. 6.5 that symmetric matrices can always be diagonalized. In this particular case, we can first find the eigenvalues and eigenvectors of C.

$$\begin{vmatrix} 5-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = (\lambda - 1)(\lambda - 6) = 0.$$

Corresponding to $\lambda_1 = 1$ and $\lambda_2 = 6$, the two normalized eigenvectors are found to be, respectively,

$$\mathbf{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Therefore the orthogonal matrix

$$U = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$$

will diagonalize the coefficient matrix

$$\widetilde{U}CU = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

If we make a change of variables

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = U \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

and take the transpose of both sides

$$\left(x_1 \ x_2 \right) = \left(u_1 \ u_2 \right) \widetilde{U},$$

we can write (6.31) as

$$\left(u_1 \ u_2 \right) \widetilde{U} C U \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \left(u_1 \ u_2 \right) \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_1^2 + 6u_2^2, \tag{6.32}$$

which is in a canonical form, i.e., it has no crossterm.

Note that the transformation matrix T defined in (6.21) is equal to $\widetilde{U}$.

Example 6.4.2. Show that the following equation

$$9x^2 - 4xy + 6y^2 - 2\sqrt{5}x - 4\sqrt{6}y = 15$$

describes an ellipse by transforming it into a standard conic section form. Where is the center and what are the lengths of its major and minor axes?

Solution 6.4.2. The quadratic part of the equation can be written as

$$9x^2 - 4xy + 6y^2 = \left(x \ y \right) \begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The eigenvalues of the coefficient matrix are given by

$$\begin{vmatrix} 9-\lambda & -2 \\ -2 & 6-\lambda \end{vmatrix} = (\lambda - 5)(\lambda - 10) = 0.$$

The normalized eigenvectors corresponding to $\lambda = 5$ and $\lambda = 10$ are found to be, respectively,

$$\mathbf{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Therefore the orthogonal matrix

$$U = \left(\mathbf{v}_1 \ \mathbf{v}_2 \right) = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$$

diagonalizes the coefficient matrix

$$\widetilde{U} C U = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}.$$

Let

$$\begin{pmatrix} x \\ y \end{pmatrix} = U \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix},$$

which is equivalent to

$$x = \frac{1}{\sqrt{5}}\left(x' - 2y'\right), \quad y = \frac{1}{\sqrt{5}}\left(2x' + y'\right),$$

then the equation can be written as

$$\begin{pmatrix} x' & y' \end{pmatrix} \widetilde{U}CU \begin{pmatrix} x' \\ y' \end{pmatrix} - 2\sqrt{5}\frac{1}{\sqrt{5}}\left(x' - 2y'\right) - 4\sqrt{5}\frac{1}{\sqrt{5}}\left(2x' + y'\right) = 15,$$

or

$$\begin{gathered} 5x'^2 + 10y'^2 - 10x' = 15, \\ x'^2 + 2y'^2 - 2x' = 3. \end{gathered}$$

Using $\left(x' - 1\right)^2 = x'^2 - 2x' + 1$, the last equation becomes

$$\left(x' - 1\right)^2 + 2y'^2 = 4,$$

or

$$\frac{(x' - 1)^2}{4} + \frac{y'^2}{2} = 1,$$

which is the standard form of an ellipse. The center of the ellipse is at $x = 1/\sqrt{5}$, $y = 2/\sqrt{5}$ (corresponding to $x' = 1$, $y' = 0$). The length of the major axis is $2\sqrt{4} = 4$, that of the minor axis is $2\sqrt{2}$.

To transform the equation into this standard form, we have rotated the coordinate axes. The major axis of the ellipse is along the vector $\mathbf{v}_1$ and the minor axis is along $\mathbf{v}_2$. Since

$$\mathbf{v}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

the major axis of the ellipse makes an angle θ with respect to the horizontal coordinate axis and $\theta = \cos^{-1}(1/\sqrt{5})$.

6.5 Hermitian Matrix and Symmetric Matrix

6.5.1 Definitions

Real Matrix

If $A^* = A$, then $a_{ij} = a_{ij}^*$. Since every element of this matrix is real, it is called a *real matrix*.

Imaginary Matrix

If $A^* = -A$, this implies that $a_{ij} = -a_{ij}^*$. Every element of this matrix is purely imaginary or zero, so it is called a *imaginary matrix*.

Hermitian Matrix

A square matrix is called hermitian if $A^\dagger = A$. It is easy to show that the elements of a hermitian matrix satisfy the relation $a_{ij}^* = a_{ji}$. Hermitian matrix is very important in quantum mechanics.

Symmetric Matrix

If the elements of the matrix are all real, a hermitian matrix is just a symmetric matrix. Symmetric matrix is of great importance in classical physics, hermitian matrix is essential in quantum mechanics.

Antihermitian Matrix and Antisymmetric Matrix

Finally, a matrix is called antihermitian or *skew-hermitian* if

$$A^\dagger = -A, \tag{6.33}$$

which implies $a_{ij}^* = -a_{ji}$.

Again, if the elements of the antihermitian matrix are all real, then the matrix is just an antisymmetric matrix.

6.5.2 Eigenvalues of Hermitian Matrix

- The eigenvalues of a hermitian (or real symmetric) matrix are all real.

Let A be a hermitian matrix, and $\mathbf{x}$ be the nontrivial eigenvector belonging to eigenvalue λ

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{6.34}$$

Take the hermitian conjugate of the equation

$$\mathbf{x}^\dagger A^\dagger = \lambda^*\mathbf{x}^\dagger. \tag{6.35}$$

Note that λ is only a number (real or complex), its hermitian conjugate is just the complex conjugate, it can be multiplied either from left or from the right.

Multiply (6.34) by $\mathbf{x}^\dagger$ from the left

$$\mathbf{x}^\dagger A\mathbf{x} = \lambda\mathbf{x}^\dagger\mathbf{x}.$$

Multiply (6.35) by $\mathbf{x}$ from the right

$$\mathbf{x}^\dagger A^\dagger\mathbf{x} = \lambda^*\mathbf{x}^\dagger\mathbf{x}.$$

Subtract it from the preceding equation

$$(\lambda - \lambda^*)\,\mathbf{x}^\dagger\mathbf{x} = \mathbf{x}^\dagger(A - A^\dagger)\mathbf{x}.$$

But A is hermitian, $A = A^{\dagger}$, so

$$(\lambda - \lambda^{*})\, \mathbf{x}^{\dagger}\mathbf{x} = 0.$$

Since $\mathbf{x}^{\dagger}\mathbf{x} \neq 0$, it follows that $\lambda = \lambda^{*}$. That is, λ is real.

For real symmetric matrices, the proof is identical, since if the matrix is real, a hermitian matrix is a symmetric matrix.

- If two eigenvalues of a hermitian (or a real symmetric) matrix are different, the corresponding eigenvectors are orthogonal.

  Let

$$A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1,$$
$$A\mathbf{x}_2 = \lambda_2 \mathbf{x}_2.$$

Multiply the first equation by $\mathbf{x}_2^{\dagger}$ from the left

$$\mathbf{x}_2^{\dagger} A\mathbf{x}_1 = \lambda_1 \mathbf{x}_2^{\dagger}\mathbf{x}_1.$$

Take the hermitian conjugate of the second equation and multiply by $\mathbf{x}_1$ from the right

$$\mathbf{x}_2^{\dagger} A\mathbf{x}_1 = \lambda_2 \mathbf{x}_2^{\dagger}\mathbf{x}_1,$$

where we have used the facts that $(A\mathbf{x}_2)^{\dagger} = \mathbf{x}_2^{\dagger}A^{\dagger}$, $A^{\dagger} = A$, and $\lambda_2 = \lambda_2^{*}$. Subtracting these two equations, we have

$$(\lambda_1 - \lambda_2)\mathbf{x}_2^{\dagger}\mathbf{x}_1 = 0.$$

Since $\lambda_1 \neq \lambda_2$, it follows:

$$\mathbf{x}_2^{\dagger}\mathbf{x}_1 = 0.$$

Therefore $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal. For real symmetric matrices, the proof is the same.

6.5.3 Diagonalizing a Hermitian Matrix

- A hermitian (or a real symmetric) matrix can be diagonalized by a unitary (or a real orthogonal) matrix.

If the eigenvalues of the matrix are all distinct, the matrix can be diagonalized by a similarity transformation as we discussed before. Here we only need to show that even if the eigenvalues are degenerate, as long as the matrix is hermitian, it can always be diagonalized. We will prove it by actually constructing a unitary matrix that will diagonalize a degenerate hermitian matrix.

Let λ_1 be a repeated eigenvalue of the $n \times n$ hermitian matrix H, let $\mathbf{x}_1$ be a normalized eigenvector corresponding to λ_1. We can take any n linearly independent vectors with the only condition that the first one is $\mathbf{x}_1$ and construct with the Gram–Schmidt process an orthonormal set of n vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, each of them has n elements.

Let U_1 be the matrix with $\mathbf{x}_i$ as its ith column

$$U_1 = \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \ldots & x_{nn} \end{pmatrix},$$

as we have shown this automatically makes U_1 an unitary matrix. The unitary transformation $U_1^\dagger H U_1$ has exactly the same set of eigenvalues as H, since they have the same characteristic polynomial

$$\begin{aligned} \left| U_1^\dagger H U_1 - \lambda I \right| &= \left| U_1^{-1} H U_1 - \lambda U_1^{-1} U_1 \right| = \left| U_1^{-1} (H - \lambda I) U_1 \right| \\ &= \left| U_1^{-1} \right| \left| H - \lambda I \right| \left| U_1 \right| = \left| H - \lambda I \right|. \end{aligned}$$

Furthermore, since H is hermitian, $U_1^\dagger H U_1$ is also hermitian, since

$$\left(U_1^\dagger H U_1 \right)^\dagger = (HU_1)^\dagger \left(U_1^\dagger \right)^\dagger = U_1^\dagger H^\dagger U_1 = U_1^\dagger H U_1.$$

Now

$$\begin{aligned} U_1^\dagger H U_1 &= \begin{pmatrix} x_{11}^* & x_{12}^* & \ldots & x_{1n}^* \\ x_{21}^* & x_{22}^* & \ldots & x_{2n}^* \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}^* & x_{n2}^* & \ldots & x_{nn}^* \end{pmatrix} H \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \ldots & x_{nn} \end{pmatrix} \\ &= \begin{pmatrix} x_{11}^* & x_{12}^* & \ldots & x_{1n}^* \\ x_{21}^* & x_{22}^* & \ldots & x_{2n}^* \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}^* & x_{n2}^* & \ldots & x_{nn}^* \end{pmatrix} \begin{pmatrix} \lambda_1 x_{11} & h_{21} & \ldots & h_{n1} \\ \lambda_1 x_{12} & h_{22} & \ldots & h_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1 x_{1n} & h_{2n} & \ldots & h_{nn} \end{pmatrix}, \end{aligned}$$

where we have used the fact that $\mathbf{x}_1$ is an eigenvector of H belonging to the eigenvalue λ_1

$$H \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix} = \lambda_1 \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix},$$

and have written

$$H \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix} = \begin{pmatrix} h_{i1} \\ h_{i2} \\ \vdots \\ h_{in} \end{pmatrix}$$

for $i \neq 1$. Furthermore

$$U_1^{\dagger} H U_1 = \begin{pmatrix} x_{11}^* & x_{12}^* & \dots & x_{1n}^* \\ x_{21}^* & x_{22}^* & \dots & x_{2n}^* \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}^* & x_{n2}^* & \dots & x_{nn}^* \end{pmatrix} \begin{pmatrix} \lambda_1 x_{11} & h_{21} & \dots & h_{n1} \\ \lambda_1 x_{12} & h_{22} & \dots & h_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1 x_{1n} & h_{2n} & \dots & h_{nn} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_{22} & \alpha_{32} & \dots & \alpha_{n2} \\ 0 & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \alpha_{2n} & \alpha_{3n} & \dots & \alpha_{nn} \end{pmatrix}.$$

The first column is determined by the orthonormal condition

$$\begin{pmatrix} x_{i1}^* & x_{i2}^* & \cdots & x_{in}^* \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix} = \begin{cases} 1 & \text{if } i = 1, \\ 0 & \text{if } i \neq 1. \end{cases}$$

The first row must be the transpose of the first column because $U_1^{\dagger} H U_1$ is hermitian (or real symmetric) and λ_1 is real and the complex conjugate of zero is itself. The crucial fact in this process is that the last $n-1$ elements of the first row are all zero. This is what distinguishes hermitian (or real symmetric) matrices from other square matrices.

If λ_1 is a twofold degenerate eigenvalue of H, then in the characteristic polynomial $p(\lambda) = |H - \lambda I|$, there must be a factor $(\lambda_1 - \lambda)^2$. Since

$$\begin{aligned} p(\lambda) &= |H - \lambda I| = \left| U_1^{\dagger} H U_1 - \lambda I \right| \\ &= \begin{vmatrix} \lambda_1 - \lambda & 0 & 0 & \dots & 0 \\ 0 & \alpha_{22} - \lambda & \alpha_{32} & \dots & \alpha_{n2} \\ 0 & \alpha_{23} & \alpha_{33} - \lambda & \dots & \alpha_{n3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \alpha_{2n} & \alpha_{3n} & \dots & \alpha_{nn} - \lambda \end{vmatrix} \\ &= (\lambda_1 - \lambda) \begin{vmatrix} \alpha_{22} - \lambda & \alpha_{32} & \dots & \alpha_{n2} \\ \alpha_{23} & \alpha_{33} - \lambda & \dots & \alpha_{n3} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{2n} & \alpha_{3n} & \dots & \alpha_{nn} - \lambda \end{vmatrix}, \end{aligned}$$

the part

$$\begin{vmatrix} \alpha_{22} - \lambda & \alpha_{32} & \dots & \alpha_{n2} \\ \alpha_{23} & \alpha_{33} - \lambda & \dots & \alpha_{n3} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{2n} & \alpha_{3n} & \dots & \alpha_{nn} - \lambda \end{vmatrix}$$

must contain another factor of $(\lambda_1 - \lambda)$. In other words, if we define H_1 as the $(n-1) \times (n-1)$ submatrix

$$\begin{pmatrix} \alpha_{22} & \alpha_{32} & \dots & \alpha_{n2} \\ \alpha_{23} & \alpha_{33} & \dots & \alpha_{n3} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{2n} & \alpha_{3n} & \dots & \alpha_{nn} \end{pmatrix} = H_1,$$

then λ_1 must be an eigenvalue of H_1. Thus we can repeat the process and construct an orthonormal set of $n-1$ column vectors with the first one being the eigenvector of H_1 belonging to the eigenvalue λ_1. Let this orthonormal set be

$$\mathbf{y}_1 = \begin{pmatrix} y_{22} \\ y_{23} \\ \vdots \\ y_{2n} \end{pmatrix}, \mathbf{y}_2 = \begin{pmatrix} y_{32} \\ y_{33} \\ \vdots \\ y_{3n} \end{pmatrix}, \dots, \mathbf{y}_{n-1} = \begin{pmatrix} y_{n2} \\ y_{n3} \\ \vdots \\ y_{nn} \end{pmatrix},$$

and let U_2 be another unitary matrix defined as

$$U_2 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & y_{22} & y_{32} & \dots & y_{n2} \\ 0 & y_{23} & y_{33} & \dots & y_{n3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & y_{2n} & y_{3n} & \dots & y_{nn} \end{pmatrix},$$

then the unitary transformation $U_2^\dagger(U_1^\dagger H U_1)U_2$ can be shown as

$$U_2^\dagger(U_1^\dagger H U_1)U_2 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & y_{22}^* & y_{23}^* & \dots & y_{2n}^* \\ 0 & y_{32}^* & y_{33}^* & \dots & y_{3n}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & y_{2n}^* & y_{3n}^* & \dots & y_{nn}^* \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_{22} & \alpha_{32} & \dots & \alpha_{n2} \\ 0 & \alpha_{23} & \alpha_{33} & \dots & \alpha_{n3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \alpha_{2n} & \alpha_{3n} & \dots & \alpha_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & y_{22} & y_{32} & \dots & y_{n2} \\ 0 & y_{23} & y_{33} & \dots & y_{n3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & y_{2n} & y_{3n} & \dots & y_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_1 & 0 & \dots & 0 \\ 0 & 0 & \beta_{33} & \dots & \beta_{n3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \beta_{3n} & \dots & \beta_{nn} \end{pmatrix}.$$

If λ_1 is m-fold degenerate, we repeat this process m times. The rest can be diagonalized by the eigenvectors belonging to different eigenvalues. After the $n \times n$ matrix is so transformed $n-1$ times, it will become diagonal.

Let us define

$$U = U_1 U_2 \cdots U_{n-1},$$

then U is unitary because all U_i are unitary. Consequently the hermitian matrix H is diagonalized by the unitary transformation $U^\dagger H U$ and the theorem is established.

This construction leads to the following important corollary.

- Every $n \times n$ hermitian (or real symmetric) matrix has n orthogonal eigenvectors regardless of the degeneracy of its eigenvalues.

This is because $U^{\dagger}HU = \Lambda$, where the elements of the diagonal matrix Λ are the eigenvalues of H. Since $U^{\dagger} = U^{-1}$, it follows from the equation $U(U^{\dagger}HU) = U\Lambda$ that $HU = U\Lambda$, which shows that each column of U is an normalized eigenvector of H.

The following example illustrates this process.

Example 6.5.1. Find an unitary matrix that will diagonalize the hermitian matrix

$$H = \begin{pmatrix} 2 & \mathrm{i} & 1 \\ -\mathrm{i} & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 2 \end{pmatrix}.$$

Solution 6.5.1. The eigenvalues of H are the roots of the characteristic equation

$$p(\lambda) = \begin{vmatrix} 2-\lambda & \mathrm{i} & 1 \\ -\mathrm{i} & 2-\lambda & \mathrm{i} \\ 1 & -\mathrm{i} & 2-\lambda \end{vmatrix} = -\lambda^3 + 6\lambda^2 - 9\lambda = -\lambda(\lambda-3)^2 = 0.$$

Therefore the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ are

$$\lambda_1 = 3, \quad \lambda_2 = 3, \quad \lambda_3 = 0.$$

It is seen that $\lambda = \lambda_1 = \lambda_2 = 3$ is twofold degenerate. Let one of the eigenvectors corresponding to λ_1 be

$$\mathbf{E}_1 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

so

$$\begin{pmatrix} 2-\lambda_1 & \mathrm{i} & 1 \\ -\mathrm{i} & 2-\lambda_1 & \mathrm{i} \\ 1 & -\mathrm{i} & 2-\lambda_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 & \mathrm{i} & 1 \\ -\mathrm{i} & -1 & \mathrm{i} \\ 1 & -\mathrm{i} & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

The three equations

$$\begin{aligned} -x_1 + \mathrm{i}x_2 + x_3 &= 0, \\ -\mathrm{i}x_1 - x_2 + \mathrm{i}x_3 &= 0, \\ x_1 - \mathrm{i}x_2 - x_3 &= 0 \end{aligned}$$

are identical to each other. For example, multiply the middle one by i will change it to the last one. The equation

$$x_1 - \mathrm{i}x_2 - x_3 = 0 \tag{6.36}$$

has an infinite number of solutions. A simple choice is to set $x_2 = 0$, then $x_1 = x_3$. Therefore

$$\mathbf{E}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

is an eigenvector. Certainly

$$\mathbf{E}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{E}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{E}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

are linearly independent. Now let us use the Gram–Schmidt process to find an orthonormal set $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$.

$$\mathbf{x}_1 = \frac{\mathbf{E}_1}{\|\mathbf{E}_1\|} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

$\mathbf{E}_2$ is already normalized and it is orthogonal to $\mathbf{E}_1$, so

$$\mathbf{x}_2 = \mathbf{E}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

$$\begin{aligned} \mathbf{x}_3' &= \mathbf{E}_3 - (\mathbf{E}_3, \mathbf{x}_1)\,\mathbf{x}_1 - (\mathbf{E}_3, \mathbf{x}_2)\,\mathbf{x}_2 \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \left[\begin{pmatrix} 1 \; 0 \; 0 \end{pmatrix} \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right] \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \left[\begin{pmatrix} 1 \; 0 \; 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right] \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \end{aligned}$$

$$\mathbf{x}_3 = \frac{\mathbf{x}_3'}{\|\mathbf{x}_3'\|} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Form a unitary matrix with $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$

$$U_1 = \begin{pmatrix} \mathbf{x}_1 \; \mathbf{x}_2 \; \mathbf{x}_3 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

The unitary similarity transformation of H by U_1 is

$$U_1^\dagger H U_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & \mathrm{i} & 1 \\ -\mathrm{i} & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$
$$= \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & -\sqrt{2}\mathrm{i} \\ 0 & \sqrt{2}\mathrm{i} & 1 \end{pmatrix}.$$

Since H and $U_1^\dagger H U_1$ have the same set of eigenvalues, therefore $\lambda = 3$ and $\lambda = 0$ must be the eigenvalue of the submatrix

$$H_1 = \begin{pmatrix} 2 & -\sqrt{2}\mathrm{i} \\ \sqrt{2}\mathrm{i} & 1 \end{pmatrix}.$$

This can also be shown directly. The two normalized eigenvector of H_1 corresponding to $\lambda = 3$ and $\lambda = 0$ are found, respectively, to be

$$\mathbf{y}_1 = \begin{pmatrix} -\frac{\sqrt{6}}{3}\mathrm{i} \\ \frac{\sqrt{3}}{3} \end{pmatrix}, \quad \mathbf{y}_2 = \begin{pmatrix} \frac{\sqrt{3}}{3}\mathrm{i} \\ \frac{\sqrt{6}}{3} \end{pmatrix}.$$

Therefore

$$U_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{\sqrt{6}}{3}\mathrm{i} & \frac{\sqrt{3}}{3}\mathrm{i} \\ 0 & \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \end{pmatrix},$$

and

$$U = U_1 U_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{\sqrt{6}}{3}\mathrm{i} & \frac{\sqrt{3}}{3}\mathrm{i} \\ 0 & \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & -\frac{\sqrt{6}}{3}\mathrm{i} & \frac{\sqrt{3}}{3}\mathrm{i} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & -\frac{\sqrt{3}}{3} \end{pmatrix}.$$

It can be easily checked that

$$U^\dagger H U = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{3}\mathrm{i} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3}\mathrm{i} & -\frac{\sqrt{3}}{3} \end{pmatrix} \begin{pmatrix} 2 & \mathrm{i} & 1 \\ -\mathrm{i} & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & -\frac{\sqrt{6}}{3}\mathrm{i} & \frac{\sqrt{3}}{3}\mathrm{i} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & -\frac{\sqrt{3}}{3} \end{pmatrix}$$
$$= \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is indeed diagonal and the diagonal elements are the eigenvalues. Furthermore, the three columns of U are indeed three orthogonal eigenvectors of H

$$H\mathbf{u}_1 = \lambda_1 \mathbf{u}_1 : \quad \begin{pmatrix} 2 & \mathrm{i} & 1 \\ -\mathrm{i} & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix} = 3 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix},$$

$$H\mathbf{u}_2 = \lambda_2 \mathbf{u}_2 : \quad \begin{pmatrix} 2 & \mathrm{i} & 1 \\ -\mathrm{i} & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{3}\mathrm{i} \\ -\frac{\sqrt{6}}{6} \end{pmatrix} = 3 \begin{pmatrix} \frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{3}\mathrm{i} \\ -\frac{\sqrt{6}}{6} \end{pmatrix},$$

$$H\mathbf{u}_3 = \lambda_3 \mathbf{u}_3 : \quad \begin{pmatrix} 2 & \mathrm{i} & 1 \\ -\mathrm{i} & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3}\mathrm{i} \\ -\frac{\sqrt{3}}{3} \end{pmatrix} = 0 \begin{pmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3}\mathrm{i} \\ -\frac{\sqrt{3}}{3} \end{pmatrix}.$$

We have followed the steps of the proof in order to illustrate the procedure. Once it is established, we can make use of the theorem and the process of finding the eigenvectors can be simplified considerably.

In this example, one can find the eigenvector for the nondegenerate eigenvalue the usual way. For the degenerate eigenvalue $\lambda = 3$, the components (x_1, x_2, x_3) of the corresponding eigenvectors must satisfy

$$x_1 - \mathrm{i}x_2 - x_3 = 0,$$

as shown in (6.36). This equation can be written as $x_2 = \mathrm{i}(x_3 - x_1)$. So in general

$$\mathbf{u} = \begin{pmatrix} x_1 \\ \mathrm{i}(x_3 - x_1) \\ x_3 \end{pmatrix},$$

where x_1 and x_3 are arbitrary. It is seen that $\mathbf{u}_1$ is just the normalized eigenvector obtained by choosing $x_1 = x_3$

$$\mathbf{u}_1 = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

The other eigenvector must also satisfy the same equation and be orthogonal to $\mathbf{u}_1$. Thus

$$\begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \mathrm{i}(x_3 - x_1) \\ x_3 \end{pmatrix} = 0,$$

which gives $x_1 + x_3 = 0$, or $x_3 = -x_1$. Normalizing the vector $\begin{pmatrix} x_1 \\ -2x_1 \\ -x_1 \end{pmatrix}$, one obtains the other eigenvector belonging to $\lambda = 3$

$$\mathbf{u}_2 = \frac{\sqrt{6}}{6}\begin{pmatrix} 1 \\ -2\mathrm{i} \\ -1 \end{pmatrix}.$$

6.5.4 Simultaneous Diagonalization

If A and B are two hermitian matrices of the same order, the following important question often arises. Can they be simultaneously diagonalized by the same matrix S? That is to say, does there exist a basis in which they are both diagonal? The answer is yes if and only if they commute.

First we will show that if they can be simultaneously diagonalized, then they must commute. That is, if

$$D_1 = S^{-1}AS \quad \text{and} \quad D_2 = S^{-1}BS,$$

where D_1 and D_2 are both diagonal matrices, then $AB = BA$.

This follows from the fact

$$D_1D_2 = S^{-1}ASS^{-1}BS = S^{-1}ABS,$$
$$D_2D_1 = S^{-1}BSS^{-1}AS = S^{-1}BAS.$$

Since diagonal matrices of the same order always commute ($D_1D_2 = D_2D_1$), so we have

$$S^{-1}ABS = S^{-1}BAS.$$

Multiplying S from the left and S^{-1} from the right, we obtain $AB = BA$.

Now we will show that the converse is also true. That is, if they commute, then they can be simultaneously diagonalized. First let A and B be 2×2 matrices. Since hermitian matrix is always diagonalizable, let S be the unitary matrix that diagonalizes A

$$S^{-1}AS = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

where λ_1 and λ_2 are the eigenvalues of A. Let

$$S^{-1}BS = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Now

$$S^{-1}ABS = S^{-1}ASS^{-1}BS = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} b_{11}\lambda_1 & b_{12}\lambda_1 \\ b_{21}\lambda_2 & b_{22}\lambda_2 \end{pmatrix},$$
$$S^{-1}BAS = S^{-1}BSS^{-1}AS = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} b_{11}\lambda_1 & b_{12}\lambda_2 \\ b_{21}\lambda_1 & b_{22}\lambda_2 \end{pmatrix}.$$

Since $AB = BA$, so $S^{-1}ABS = S^{-1}BAS$

$$\begin{pmatrix} b_{11}\lambda_1 & b_{12}\lambda_1 \\ b_{21}\lambda_2 & b_{22}\lambda_2 \end{pmatrix} = \begin{pmatrix} b_{11}\lambda_1 & b_{12}\lambda_2 \\ b_{21}\lambda_1 & b_{22}\lambda_2 \end{pmatrix}.$$

It follows that:

$$b_{21}\lambda_2 = b_{21}\lambda_1, \quad b_{12}\lambda_1 = b_{12}\lambda_2.$$

If $\lambda_2 \neq \lambda_1$, then $b_{12} = 0$ and $b_{21} = 0$. In other words

$$S^{-1}BS = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}.$$

Therefore A and B are simultaneously diagonalized.

If $\lambda_2 = \lambda_1 = \lambda$, we cannot conclude that $S^{-1}BS$ is diagonal. However, in this case

$$S^{-1}AS = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}.$$

Moveover, since B is hermitian, so the unitary similarity transform $S^{-1}BS$ is also hermitian, therefore $S^{-1}BS$ is diagonalizable. Let T be the unitary matrix that diagonalize $S^{-1}BS$

$$T^{-1}\left(S^{-1}BS\right)T = \begin{pmatrix} \lambda_1' & 0 \\ 0 & \lambda_2' \end{pmatrix}.$$

On the other hand,

$$T^{-1}\left(S^{-1}AS\right)T = T^{-1}\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}T = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}T^{-1}T = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}.$$

Thus the product matrix $U = ST$ diagonalizes both A and B. Therefore, with or without degeneracy, as long as A and B commute, they can be simultaneously diagonalized.

Although we have used 2×2 matrices for illustration, the same "proof" can obviously be applied to matrices of higher order.

Example 6.5.2. Let

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

Can A and B be simultaneously diagonalized? If so, find the unitary matrix that diagonalized them.

Solution 6.5.2.

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 8 & 7 \\ 7 & 8 \end{pmatrix},$$

$$BA = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 8 & 7 \\ 7 & 8 \end{pmatrix}.$$

Thus $[A, B] = 0$, therefore they can be simultaneously diagonalized

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (\lambda - 1)(\lambda - 3) = 0.$$

The normalized eigenvectors corresponding to $\lambda = 1, 3$ are, respectively,

$$\mathbf{x}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{x}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Therefore

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad S^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

$$S^{-1}AS = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix},$$
$$S^{-1}BS = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}.$$

Thus they are simultaneously diagonalized. It also shows that 1 and 5 are the eigenvalues of B. This can be easily verified, since

$$\begin{vmatrix} 3-\lambda & 2 \\ 2 & 3-\lambda \end{vmatrix} = (\lambda - 1)(\lambda - 5) = 0.$$

If we diagonalize B first, we will get exactly the same result.

6.6 Normal Matrix

A square matrix is said to be normal if and only if it commutes with its hermitian conjugate. That is, A is normal, if and only if

$$AA^{\dagger} = A^{\dagger}A. \tag{6.37}$$

It can be easily shown that all hermitian (or real symmetric), antihermitian (or real antisymmetric), and unitary (or real orthogonal) matrices are normal. All we have to do is to substitute these matrices into (6.37). By virtue of their definition, it is immediately clear that the two sides of the equation are indeed the same.

So far we have shown that every hermitian matrix is diagonalizable by a unitary similarity transformation. In what follows, we will prove the generalization of this theorem that every normal matrix is diagonalizable.

First, if the square matrix A is given, that means all elements of A are known, so we can take its hermitian conjugate $A^{\dagger}$. Then let

$$B = \frac{1}{2}(A + A^\dagger),$$
$$C = \frac{1}{2\mathrm{i}}(A - A^\dagger).$$

So

$$A = B + \mathrm{i}C. \tag{6.38}$$

Since $(A^\dagger)^\dagger = A$ and $(A+B)^\dagger = A^\dagger + B^\dagger$,

$$B^\dagger = \frac{1}{2}(A + A^\dagger)^\dagger = \frac{1}{2}(A^\dagger + A) = B,$$
$$C^\dagger = \frac{1}{2\mathrm{i}^*}(A - A^\dagger)^\dagger = -\frac{1}{2\mathrm{i}}(A^\dagger - A) = C.$$

Thus, B and C are both hermitian. In other words, a square matrix can always be decomposed into two hermitian matrices as shown in (6.38). Furthermore

$$BC = \frac{1}{4\mathrm{i}}\left(A^2 - AA^\dagger + A^\dagger A - A^{\dagger 2}\right),$$
$$CB = \frac{1}{4\mathrm{i}}\left(A^2 - A^\dagger A + AA^\dagger - A^{\dagger 2}\right).$$

It is clear that if $A^\dagger A = AA^\dagger$, then $BC = CB$. In other words, if A is normal, then B and C commute.

We have shown in Sect. 6.5 that if B and C commute, then they can be simultaneously diagonalized. That is, we can find a unitary matrix S such that $S^{-1}BS$ and $S^{-1}CS$ are both diagonal. Since

$$S^{-1}AS = S^{-1}BS + \mathrm{i}S^{-1}CS,$$

it follows that $S^{-1}AS$ must also be diagonal.

Conversely, if $S^{-1}AS = D$ is diagonal, then

$$\left(S^{-1}AS\right)^\dagger = S^{-1}A^\dagger S = D^\dagger = D^*,$$

since S is unitary and D is diagonal. It follows that:

$$S^{-1}AA^\dagger S = (S^{-1}AS)\left(S^{-1}A^\dagger S\right) = DD^*,$$
$$S^{-1}A^\dagger AS = (S^{-1}A^\dagger S)(S^{-1}AS) = D^*D.$$

Since $DD^* = D^*D$, clearly $AA^\dagger = A^\dagger A$. Therefore we conclude.

- A matrix can be diagonalized by a unitary similarity transformation if and only if it is normal.

Thus both hermitian and unitary matrices are diagonalizable by a unitary similarity transformation.

The eigenvalues of a hermitian matrix are always real. This is the reason why in quantum mechanics observable physical quantities are associated with the eigenvalues of hermitian operators, because the result of any measurement is, of course, a real number. However, the eigenvectors of a hermitian matrix may be complex, therefore the unitary matrix that diagonalizes the hermitian matrix is, in general, complex.

A real symmetric matrix is a hermitian matrix, therefore its eigenvalues must also be real. Since the matrix and the eigenvalues are both real, the eigenvectors can be taken to be real. Therefore, the diagonalizing matrix is a real orthogonal matrix.

Unitary matrices, including real orthogonal matrices, can be diagonalized by a unitary similarity transformation. However, the eigenvalues and eigenvectors of a unitary matrix are, in general, complex. Therefore the diagonalizing matrix is not a real orthogonal matrix, but a complex unitary matrix. For example, the rotation matrix is a real orthogonal matrix, but, in general, it can only be diagonalized by a complex unitary matrix.

6.7 Functions of a Matrix

6.7.1 Polynomial Functions of a Matrix

Any square matrix A can be multiplied by itself. The associative law of matrix multiplication guarantees that the operation of A times itself n times, which is denoted as A^n, is an unambiguous operation. Thus

$$A^m A^n = A^{m+n}.$$

Moreover, we have defined the inverse A^{-1} of a nonsingular matrix in such a way that $AA^{-1} = A^{-1}A = I$. Therefore it is natural to define

$$A^0 = A^{1-1} = AA^{-1} = I, \quad \text{and} \quad A^{-n} = \left(A^{-1}\right)^n.$$

With these definitions, we can define polynomial functions of a square matrix in exactly the same way as scalar polynomials.

For example, if $f(x) = x^2 + 5x + 4$, and $A = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}$, we define $f(A)$, as

$$f(A) = A^2 + 5A + 4.$$

Since

$$A^2 = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 8 & 11 \end{pmatrix},$$

$$f(A) = \begin{pmatrix} 3 & 4 \\ 8 & 11 \end{pmatrix} + 5\begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} + 4\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 12 & 9 \\ 18 & 30 \end{pmatrix}.$$

It is interesting to note that $f(A)$ can also be evaluated by using the factored terms of $f(x)$. For example

$$f(x) = x^2 + 5x + 4 = (x+1)(x+4),$$

so

$$\begin{aligned} f(A) &= (A+I)(A+4I) \\ &= \left[\begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\left[\begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} + 4\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 2 & 1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 5 & 1 \\ 2 & 7 \end{pmatrix} = \begin{pmatrix} 12 & 9 \\ 18 & 30 \end{pmatrix}. \end{aligned}$$

Example 6.7.1. Find $f(A)$, if

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} \quad \text{and} \quad f(x) = \frac{x}{x^2-1}.$$

Solution 6.7.1.

$$f(A) = \frac{A}{A^2 - I} = A\left(A^2 - I\right)^{-1} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 2 & 4 \\ 8 & 10 \end{pmatrix}^{-1} = \frac{1}{6}\begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}.$$

Note that $f(A)$ can also be evaluated by partial fraction. Since

$$f(x) = \frac{x}{x^2-1} = \frac{1}{2}\frac{1}{x-1} + \frac{1}{2}\frac{1}{x+1},$$

$$\begin{aligned} f(A) &= \frac{1}{2}(A-I)^{-1} + \frac{1}{2}(A+I)^{-1} \\ &= \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix}^{-1} + \frac{1}{2}\begin{pmatrix} 2 & 1 \\ 2 & 4 \end{pmatrix}^{-1} = \frac{1}{6}\begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}. \end{aligned}$$

6.7.2 Evaluating Matrix Functions by Diagonalization

When the square matrix A is similar to a diagonal matrix, the evaluation of $f(A)$ can be considerably simplified.

If A is diagonalizable, then

$$S^{-1}AS = D,$$

where D is a diagonal matrix. It follows that:

$$\begin{aligned} D^2 &= S^{-1}ASS^{-1}AS = S^{-1}A^2S, \\ D^k &= S^{-1}A^{k-1}SS^{-1}AS = S^{-1}A^kS. \end{aligned}$$

Thus

$$A^k = SD^kS^{-1},$$
$$A^n + A^m = SD^nS^{-1} + SD^mS^{-1} = S(D^n + D^m)S^{-1}.$$

If $f(A)$ is a polynomial, then

$$f(A) = Sf(D)S^{-1}.$$

Furthermore, since D is diagonal and the elements of D are the eigenvalues of A,

$$D^k = \begin{pmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & \lambda_n^k \end{pmatrix},$$

$$f(D) = \begin{pmatrix} f(\lambda_1) & 0 & \dots & 0 \\ 0 & f(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & f(\lambda_n) \end{pmatrix}.$$

Therefore

$$f(A) = S\begin{pmatrix} f(\lambda_1) & 0 & \dots & 0 \\ 0 & f(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & f(\lambda_n) \end{pmatrix}S^{-1}.$$

Example 6.7.2. Find $f(A)$, if

$$A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix} \quad \text{and} \quad f(x) = x^4 - 4x^3 + 6x^2 - x - 3.$$

Solution 6.7.2. First find the eigenvalues and eigenvectors of A

$$\begin{vmatrix} 0-\lambda & -2 \\ 1 & 3-\lambda \end{vmatrix} = (\lambda - 1)(\lambda - 2) = 0.$$

The eigenvectors corresponding to $\lambda_1 = 1$ and $\lambda_2 = 2$ are, respectively,

$$\mathbf{u}_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Therefore

$$S = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix}, \quad S^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix},$$

and

$$D = S^{-1}AS = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Thus

$$f(A) = Sf(D)\,S^{-1} = S\begin{pmatrix} f(1) & 0 \\ 0 & f(2) \end{pmatrix} S^{-1}.$$

Since

$$f(1) = -1, \quad f(2) = 3,$$

$$f(A) = Sf(D)\,S^{-1} = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} -5 & -8 \\ 4 & 7 \end{pmatrix}.$$

Example 6.7.3. Find the matrix A such that

$$A^2 - 4A + 4I = \begin{pmatrix} 4 & 3 \\ 5 & 6 \end{pmatrix}.$$

Solution 6.7.3. Let us first diagonalize the right-hand side

$$\begin{vmatrix} 4-\lambda & 3 \\ 5 & 6-\lambda \end{vmatrix} = (\lambda - 1)(\lambda - 9) = 0.$$

The eigenvectors corresponding to $\lambda_1 = 1$ and $\lambda_2 = 9$ are, found to be, respectively,

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}.$$

Thus

$$S = \begin{pmatrix} 1 & 3 \\ -1 & 5 \end{pmatrix}, \quad S^{-1} = \frac{1}{8}\begin{pmatrix} 5 & -3 \\ 1 & 1 \end{pmatrix},$$

and

$$D = S^{-1}\begin{pmatrix} 4 & 3 \\ 5 & 6 \end{pmatrix} S = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}.$$

Therefore

$$S^{-1}\left(A^2 - 4A + 4I\right) S = S^{-1}\begin{pmatrix} 4 & 3 \\ 5 & 6 \end{pmatrix} S = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}.$$

The left-hand side must also be diagonal, since the right-hand side is diagonal. Since we have shown that, as long as $S^{-1}AS$ is diagonal, $S^{-1}A^kS$ will be diagonal, we can assume

$$S^{-1}AS = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}.$$

It follows that:

$$S^{-1}\left(A^2-4A+4I\right)S=\begin{pmatrix} x_1^2-4x_1+4 & 0 \\ 0 & x_2^2-4x_2+4 \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix},$$

which gives

$$\begin{aligned} x_1^2-4x_1+4&=1,\\ x_2^2-4x_2+4&=9. \end{aligned}$$

From the first equation we get $x_1=1,3$, and from the second equation we obtain $x_2=5,-1$. Therefore there are four possible combinations for $\begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}$, namely

$$\Lambda_1=\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix},\quad \Lambda_2=\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\quad \Lambda_3=\begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix},\quad \Lambda_4=\begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}.$$

Thus the original equation has four solutions

$$A_1=S\Lambda_1S^{-1}=\begin{pmatrix} 1 & 3 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}\frac{1}{8}\begin{pmatrix} 5 & -3 \\ 1 & 1 \end{pmatrix}=\frac{1}{2}\begin{pmatrix} 5 & 3 \\ 5 & 7 \end{pmatrix},$$

and similarly

$$A_2=\frac{1}{4}\begin{pmatrix} 1 & -3 \\ -5 & -1 \end{pmatrix},\quad A_3=\frac{1}{4}\begin{pmatrix} 15 & 3 \\ 5 & 17 \end{pmatrix},\quad A_4=\frac{1}{2}\begin{pmatrix} 3 & -3 \\ -5 & 1 \end{pmatrix}.$$

For every scalar function that can be expressed as an infinite series, a corresponding matrix function can be defined. For example, with

$$\mathrm{e}^x=1+x+\frac{1}{2}x^2+\frac{1}{3!}x^3+\cdots,$$

we can define

$$\mathrm{e}^A=I+A+\frac{1}{2}A^2+\frac{1}{3!}A^3+\cdots.$$

If A is diagonalizable, then

$$\begin{aligned} S^{-1}AS=D,\quad A^n&=SD^nS^{-1},\\ \mathrm{e}^A&=S\left(I+D+\frac{1}{2}D^2+\frac{1}{3!}D^3+\cdots\right)S^{-1}, \end{aligned}$$

where

$$D=\begin{pmatrix} \lambda_1 & \ldots & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \ldots & \ldots & \ldots & \lambda_n \end{pmatrix}.$$

It follows that:

$$e^A = S \begin{pmatrix} 1+\lambda_1+\frac{1}{2}\lambda_1^2\cdots & \dots & \dots & 0 \\ 0 & 1+\lambda_2+\frac{1}{2}\lambda_2^2\cdots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1+\lambda_n+\frac{1}{2}\lambda_n^2\cdots \end{pmatrix} S^{-1}$$

$$= S \begin{pmatrix} e^{\lambda_1} & \dots & \dots & 0 \\ 0 & e^{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n} \end{pmatrix} S^{-1}.$$

Example 6.7.4. Evaluate e^A if $A = \begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix}$.

Solution 6.7.4. Since A is symmetric, it is diagonalizable.

$$\begin{vmatrix} 1-\lambda & 5 \\ 5 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 24 = 0,$$

which gives $\lambda = 6, -4$. The corresponding eigenvectors are found to be

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus

$$S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Therefore

$$e^A = S \begin{pmatrix} e^6 & 0 \\ 0 & e^{-4} \end{pmatrix} S^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} e^6 & 0 \\ 0 & e^{-4} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix} (e^6+e^{-4}) & (e^6-e^{-4}) \\ (e^6-e^{-4}) & (e^6+e^{-4}) \end{pmatrix}.$$

6.7.3 The Cayley–Hamilton Theorem

The famous Cayley–Hamilton theorem states that every square matrix satisfies its own characteristic equation.

That is, if $P(\lambda)$ is the characteristic polynomial of the nth order matrix A

$$P(\lambda) = |A - \lambda I| = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0,$$

then

$$P(A) = c_n A^n + c_{n-1} A^{n-1} + \cdots + c_0 I = 0.$$

To prove this theorem, let $\mathbf{x}_i$ be the eigenvector corresponding to the eigenvalue λ_i. So

$$P(\lambda_i) = 0, \quad A\mathbf{x}_i = \lambda_i \mathbf{x}_i.$$

Now

$$\begin{aligned} P(A)\mathbf{x}_i &= \left(c_n A^n + c_{n-1} A^{n-1} + \cdots + c_0 I\right) \mathbf{x}_i \\ &= \left(c_n \lambda_i^n + c_{n-1} \lambda_i^{n-1} + \cdots + c_0\right) \mathbf{x}_i \\ &= P(\lambda_i)\mathbf{x}_i = 0\mathbf{x}_i. \end{aligned}$$

Since this is true for any eigenvector of A, $P(A)$ must be a zero matrix.

For example, if

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},$$

$$P(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 3.$$

$$\begin{aligned} P(A) &= \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - 2 \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - 3 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} - \begin{pmatrix} 2 & 4 \\ 4 & 2 \end{pmatrix} - \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 5-3-2 & 4-4 \\ 4-4 & 5-3-2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Inverse by Cayley–Hamilton Theorem

The Cayley–Hamilton theorem can be used to find the inverse of a square matrix. Starting with the characteristic equation of A

$$P(\lambda) = |A - \lambda I| = c_n \lambda^n + c_{n-1} \lambda^{n-1} + \cdots + c_1 \lambda + c_0 = 0,$$

we have

$$P(A) = c_n A^n + c_{n-1} A^{n-1} + \cdots + c_1 A + c_0 I = 0.$$

Multiplying this equation from the left by A^{-1}, we obtain

$$A^{-1} P(A) = c_n A^{n-1} + c_{n-1} A^{n-2} + \cdots + c_1 I + c_0 A^{-1} = 0.$$

Thus

$$A^{-1} = -\frac{1}{c_0} \left(c_n A^{n-1} + c_{n-1} A^{n-2} + \cdots + c_1 I\right).$$

Example 6.7.5. Find A^{-1} by Cayley–Hamilton theorem if

$$A = \begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix}.$$

Solution 6.7.5.

$$P(\lambda) = \begin{pmatrix} 5-\lambda & 7 & -5 \\ 0 & 4-\lambda & -1 \\ 2 & 8 & -3-\lambda \end{pmatrix} = 6 - 11\lambda + 6\lambda^2 - \lambda^3,$$

$$\begin{aligned} P(A) &= 6I - 11A + 6A^2 - A^3 = 0, \\ A^{-1}P(A) &= 6A^{-1} - 11I + 6A - A^2 = 0, \\ A^{-1} &= \frac{1}{6}\left(A^2 - 6A + 11I\right), \end{aligned}$$

$$\begin{aligned} A^{-1} &= \frac{1}{6}\left[\begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix}\begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix} - 6\begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix} + 11\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] \\ &= \frac{1}{6}\left[\begin{pmatrix} 15 & 23 & -17 \\ -2 & 8 & -1 \\ 4 & 22 & -9 \end{pmatrix} - \begin{pmatrix} 30 & 42 & -30 \\ 0 & 24 & -6 \\ 12 & 48 & -18 \end{pmatrix} + \begin{pmatrix} 11 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 11 \end{pmatrix}\right] \\ &= \frac{1}{6}\begin{pmatrix} -4 & -19 & 13 \\ -2 & -5 & 5 \\ -8 & -26 & 20 \end{pmatrix}. \end{aligned}$$

It can be readily verified that

$$A^{-1}A = \frac{1}{6}\begin{pmatrix} -4 & -19 & 13 \\ -2 & -5 & 5 \\ -8 & -26 & 20 \end{pmatrix}\begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

High Powers of a Matrix

An important application of the Cayley–Hamilton theorem is in the representation of high powers of a matrix. From the equation $P(A) = 0$, we have

$$A^n = -\frac{1}{c_n}\left(c_{n-1}A^{n-1} + c_{n-2}A^{n-2} + \cdots + c_1A + c_0I\right). \tag{6.39}$$

Multiplying through by A

$$A^{n+1} = -\frac{1}{c_n}(c_{n-1}A^n + c_{n-2}A^{n-1} + \cdots + c_1A^2 + c_0A), \tag{6.40}$$

and substituting A^n from (6.39) into (6.40), we obtain

$$A^{n+1} = \left(\frac{c_{n-1}^2}{c_n^2} - \frac{c_{n-2}}{c_n} \right) A^{n-1} + \cdots + \left(\frac{c_{n-1}c_1}{c_n^2} - \frac{c_0}{c_n} \right) A + \frac{c_{n-1}c_0}{c_n^2} I. \quad (6.41)$$

Clearly this process can be continued. Thus any integer power of a matrix of order n can be reduced to a polynomial of the matrix, the highest degree of which is at most $n-1$. This fact can be used directly to obtain high powers of A.

Example 6.7.6. Find A^{100}, if $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$.

Solution 6.7.6. Since

$$\begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 8 = (\lambda - 4)(\lambda + 2) = 0,$$

the eigenvalues of A are $\lambda_1 = 4$ and $\lambda_2 = -2$. The eigenvalues of A^{100} must be λ_1^{100} and λ_2^{100}, i.e.,

$$A^{100}\mathbf{x}_1 = \lambda_1^{100}\mathbf{x}_1, \quad A^{100}\mathbf{x}_2 = \lambda_2^{100}\mathbf{x}_2.$$

On the other hand, from the Cayley–Hamilton theorem, we know that A^{100} can be expressed as a linear combination of A and I, since A is of second order matrix ($n = 2$).

$$A^{100} = \alpha A + \beta I,$$

thus

$$A^{100}\mathbf{x}_1 = (\alpha A + \beta I)\mathbf{x}_1 = (\alpha\lambda_1 + \beta)\mathbf{x}_1,$$
$$A^{100}\mathbf{x}_2 = (\alpha A + \beta I)\mathbf{x}_2 = (\alpha\lambda_2 + \beta)\mathbf{x}_2.$$

Therefore

$$\lambda_1^{100} = \alpha\lambda_1 + \beta, \quad \lambda_2^{100} = \alpha\lambda_2 + \beta.$$

It follows:

$$\alpha = \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1^{100} - \lambda_2^{100}\right) = \frac{1}{6}(4^{100} - 2^{100}),$$
$$\beta = \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1\lambda_2^{100} - \lambda_2\lambda_1^{100}\right) = \frac{1}{3}(4^{100} + 2^{101}).$$

Hence

$$A^{100} = \frac{1}{6}(4^{100} - 2^{100})A + \frac{1}{3}(4^{100} + 2^{101})I.$$

Exercises

1. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 19 & 10 \\ -30 & -16 \end{pmatrix}.$$

Ans. $\lambda_1 = 4$, $\mathbf{x}_1 = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$; $\lambda_2 = -1$, $\mathbf{x}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

2. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 6-2\mathrm{i} & -1+3\mathrm{i} \\ 9+3\mathrm{i} & -4+3\mathrm{i} \end{pmatrix}.$$

Ans. $\lambda_1 = 2$, $\mathbf{x}_1 = \begin{pmatrix} 1-\mathrm{i} \\ 2 \end{pmatrix}$; $\lambda_2 = \mathrm{i}$, $\mathbf{x}_2 = \begin{pmatrix} 1-\mathrm{i} \\ 3 \end{pmatrix}$.

3. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 2 & -2 & 1 \\ 2 & -4 & 3 \\ 2 & -6 & 5 \end{pmatrix}.$$

Ans. $\lambda_1 = 0$, $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$; $\lambda_2 = 1$, $\mathbf{x}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$; $\lambda_3 = 2$, $\mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$.

4. If $U^{\dagger}U = I$, show that (a) the columns of U form an orthonormal set; (b) $UU^{\dagger} = I$ and the rows of U form an orthonormal set.

5. Show that eigenvalues of antihermitian matrix are either zero or pure imaginary.

6. Find the eigenvalues and eigenvectors of the following matrix:

$$\frac{1}{5}\begin{pmatrix} 7 & 6\mathrm{i} \\ -6\mathrm{i} & -2 \end{pmatrix},$$

and show explicitly that the two eigenvectors are orthogonal.

Ans. $\lambda_1 = 2$, $\mathbf{x}_1 = \begin{pmatrix} 2\mathrm{i} \\ 1 \end{pmatrix}$; $\lambda_2 = -1$, $x_2 = \begin{pmatrix} 1 \\ 2\mathrm{i} \end{pmatrix}$.

7. Show the two eigenvectors in the previous problem are orthogonal. Construct an unitary matrix U with the two normalized eigenvectors, and show that

$$U^{\dagger}U = I.$$

Ans. $U = \frac{1}{\sqrt{5}}\begin{pmatrix} 2\mathrm{i} & 1 \\ 1 & 2\mathrm{i} \end{pmatrix}$.

8. Find the eigenvalues and eigenvectors of the following symmetric matrix:

$$A = \frac{1}{5}\begin{pmatrix} 6 & 12 \\ 12 & -1 \end{pmatrix}.$$

Construct an orthogonal matrix U with the two normalized eigenvectors and show that

$$\widetilde{U}AU = \Lambda,$$

where Λ is a diagonal matrix whose elements are the eigenvalues of A.
Ans. $U = \frac{1}{5}\begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}.$

9. Diagonalize the hermitian matrix

$$A = \begin{pmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & 2 \end{pmatrix},$$

with a unitary similarity transformation

$$U^{\dagger}AU = \Lambda.$$

Find the unitary matrix U and the diagonal matrix Λ.
Ans. $U = \begin{pmatrix} -\frac{1+\mathrm{i}}{\sqrt{3}} & \frac{1+\mathrm{i}}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}.$

10. Diagonalize the symmetric matrix

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

with a similarity transformation

$$U^{\dagger}AU = \Lambda.$$

Find the orthogonal matrix U and the diagonal matrix Λ.
Ans. $U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$

11. Diagonalize the symmetric matrix

$$A = \frac{1}{3}\begin{pmatrix} -7 & 2 & 10 \\ 2 & 2 & -8 \\ 10 & -8 & -4 \end{pmatrix},$$

with a similarity transformation

$$\widetilde{U}AU = \Lambda.$$

Find the orthogonal matrix U and the diagonal matrix Λ.

$$\text{Ans. } U = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{pmatrix}.$$

12. If A is a symmetric matrix (so $\widetilde{A} = A$), S is an orthogonal matrix and $A' = S^{-1}AS$, show that A' is also symmetric.

13. If $\mathbf{u}$ and $\mathbf{v}$ are two column matrices in a two-dimensional space, and they are related by the equation

$$\mathbf{v} = C\mathbf{u},$$

where

$$C = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

find C^{-1} by the following methods:

(a) By Cramer's rule.

(b) Show that C is orthogonal, so $C^{-1} = \widetilde{C}$.

(c) The equation $\mathbf{v} = C\mathbf{u}$ means that C rotates $\mathbf{u}$ to $\mathbf{v}$. Since $\mathbf{u} = C^{-1}\mathbf{v}$, C^{-1} must rotate $\mathbf{v}$ back to $\mathbf{u}$. Therefore C^{-1} must also be a rotation matrix with an opposite direction of rotation.

14. Find the eigenvalues λ_1, λ_2, and the corresponding eigenvectors of the two-dimensional rotation matrix

$$C = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Find the unitary matrix U, such that

$$U^{\dagger}CU = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

$$\text{Ans. } \mathrm{e}^{\mathrm{i}\theta}, \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix}, \mathrm{e}^{-\mathrm{i}\theta}, \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}, U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -\mathrm{i} & \mathrm{i} \end{pmatrix}.$$

15. Show that the canonical form (in which there is no crossproduct terms) of the quadratic expression

$$Q(x_1, x_2) = 7x_1^2 + 48x_1x_2 - 7x_2^2$$

is

$$Q'(x_1', x_2') = 25x_1^2 - 25x_2^2,$$

where

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}.$$

Find the orthogonal matrix S.

Ans. $S = \frac{1}{5}\begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix}$.

16. If $A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$ and $f(x) = x^3 + 3x^2 - 3x - 1$, find $f(A)$.

Ans. $\begin{pmatrix} -13 & -26 \\ 13 & 26 \end{pmatrix}$.

17. If $A = \begin{pmatrix} 1 & 0 \\ 2 & \frac{1}{4} \end{pmatrix}$, find A^n and $\lim_{n\to\infty} A^n$.

Ans. $A^n = \begin{pmatrix} 1 & 0 \\ \frac{8}{3} - \frac{8}{3}\left(\frac{1}{4}\right)^n & \left(\frac{1}{4}\right)^n \end{pmatrix}$, $\quad \lim_{n\to\infty} = \begin{pmatrix} 1 & 0 \\ \frac{8}{3} & 0 \end{pmatrix}$.

18. Solve for X, if $X^3 = \begin{pmatrix} -6 & 14 \\ -7 & 15 \end{pmatrix}$.

Ans. $X = \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix}$.

19. Solve the equation

$$M^2 - 5M + 3I = \begin{pmatrix} 1 & -4 \\ 2 & -5 \end{pmatrix}.$$

Ans. $M_1 = \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix}$, $M_2 = \begin{pmatrix} -1 & 4 \\ -2 & 5 \end{pmatrix}$, $M_3 = \begin{pmatrix} 6 & -4 \\ 2 & 0 \end{pmatrix}$, $M_4 = \begin{pmatrix} 5 & -2 \\ 1 & 2 \end{pmatrix}$.

20. According to Cayley–Hamilton theorem, every square matrix satisfies its own characteristic equation. Verify this theorem for each of the following matrices:

$$\text{(a)} \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} -1 & -2 \\ 3 & 4 \end{pmatrix}.$$

21. Find A^{-1} by Cayley–Hamilton theorem if $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Ans. $A^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.